Technical note on FRS 2019-20 reweighting methodology

Context

This technical note provides a description of the methodology used to reweight the Family Resources Survey (FRS) 2019-20 data to obtain an "updated" version of the FRS reflecting the population in 2021-22.[1] All analysis was performed in Stata (StataCorp 2021). Code used to implement the following reweighting methodology is available upon request.

Survey data typically includes unit-level weights (also called grossing factors), which reflect the number of units in the population represented by each unit in the survey sample. For instance, a weight of 100 for a person-level observation in the FRS indicates that that observation represents 100 people in the UK population.

To create an updated version of the FRS for 2021-22, we reweight the 2019-20 FRS to reflect labour market characteristics calculated from the first quarter of Labour Force Survey (LFS) data for 2022 (NISRA, Central Survey Unit, ONS, Social Survey Division, 2022). Despite the availability of 2020-21 FRS data, we choose the 2019-20 data because the limited sample size and possible measurement issues induced by the Covid-19 pandemic make the 2020-21 data less reliable. The goal of the reweighting is to make the weighted 2019-20 data look more similar to the (as yet unknown) 2021-22 population estimates, including labour market characteristics like employment rates among different groups of people. The reweighted data is then used as an input into IPPR tax modelling software to obtain a snapshot of income, income distribution, and poverty in financial year 2022-23.

Reweighting groups

We match labour market characteristics within groups to improve the accuracy of the reweighting. Groups are defined for all adults 16-64 based on combinations of sex, age, and level of highest qualification. Sex and age both have two categories, and level of qualifications three, so that all observations are sorted into one of 12 total groups.[2] Categories of each

---

[1] The 2019-20 FRS data is publicly available through the UK Data Service website (Department for Work and Pensions, Office for National Statistics, NatCen Social Research 2021).

[2] To match 2022 population estimates, two additional groups are included in the reweighting (men and women aged 65+). The reweighting for these groups ensures that the adult weights sum to the estimated population of each group of adults in 2022, rather than matching an employment rate within those groups.

characteristic defining the groups are shown in Table A1. For example, one group contains all women aged 50-64 with a degree or a higher education qualification.

Although the target population is Scotland, the IPPR model uses the FRS sample for all of the UK to calculate tax implications and the poverty line. Therefore, the reweighting is conducted separately for Scotland and rUK within the groups defined in Table A1.

Table A1: Group classifications for reweighting

Sex	Age	Qualification	Group
Women	50+	Degree or higher ed	A
		Highers	B
		Standard grades	C
	16-49	Degree or higher ed	D
		Highers	E
		Standard grades	F
Men	50+	Degree or higher ed	G
		Highers	H
		Standard grades	I
	16-49	Degree or higher ed	J
		Highers	K
		Standard grades	L

Calculation of target population sizes and employment rates

We produce two reweighted versions of the FRS 2019-20 data. The first updates the weights assigned to individuals in the FRS to match estimated population sizes from the first quarter of 2022. This updating is done to reflect any changes in the populations of each group post-pandemic, and acts as a baseline dataset for the IPPR model. The population estimates come from quarterly LFS data for the first quarter of 2022. The second reweighted dataset is updated to reflect both population size changes and changes in the employment rate within each group, and acts as the comparison dataset for the IPPR model. As with the first reweighted dataset, the population estimates come from LFS data, as do employment rate estimates for each group.

The LFS is designed to capture up-to-date information on the labour market at the individual level, while the FRS is recommended as a primary source of household income information (Office for National Statistics 2022). There are significant differences between employment rates obtained from the FRS and those calculated from LFS data. Therefore, instead of matching the absolute employment rate in the 2022Q1 LFS, we calculate the change in the employment rate for each group between the 2020Q1 and 2022Q1 LFS estimates and apply that same percentage point change to the employment rate for each group in the 2019-20 FRS. The resulting employment rate is the target rate for the reweighting, which represents the 2021-22 FRS.

Reweighting of FRS data

For the first reweighted dataset, target population sizes for each group are obtained from LFS 2022Q1 population estimates. We then use the Stata command sreweight to generate new individual-level weights for the FRS 2019-20 data that match the target size for each group (Pacifico 2014). Adult weights are then averaged within each household and applied to all household members.

Producing the second reweighted dataset requires a distinction between target employed and unemployed populations within each group. For each group, we multiply the target employment rate by the population size (obtained from LFS 2022Q1 population estimates) to obtain target employed and unemployed population sizes. The same reweighting procedure described above is applied to the FRS sample, this time to match both population and employment rate changes between 2019-20 and 2022. Groups for men and women in Scotland and rUK are reweighted for population sizes, but not for employment rates.

The IPPR model also calls on data from the Households Below Average Income 2019-20 data (Department for Work and Pensions 2022), which uses most of the FRS household sample but omits some households. These data are used to produce measures of inequality and poverty. The process described above to obtain two reweighted datasets, one adjusting for population changes between 2019-20 and 2022 and the other adjusting for both population and employment changes, is applied to HBAI 2019-20 data as well. These form another input into the IPPR model to obtain estimates of income, poverty, and inequality in 2022-23.[3]

Verification of reweighting process

We verify that the reweighting has been successful using several checks. First, we check that the sum of the individual weights by group match the targets for both reweighted datasets.

Second, we compare each group's weekly hours worked and weekly pay in the reweighted FRS to the 2022Q1 LFS data to make sure that these characteristics are comparable. Because FRS and LFS generally yield different estimates, we test that the differences in mean hours and pay between the reweighted FRS (representing 2021-22) and 2022Q1 LFS are not significantly different than the differences in means between the original 2019-20 FRS and 2020Q1 LFS. Confidence intervals for these differences between versions of the FRS and LFS are shown in Figures 1-2. Groups are indexed in Table A1 for reference.

Since the purpose of this project is to study the impact of the pandemic on income, poverty, and inequality, the most important measure to match is weekly pay. Based on Figure A1, the differences between the reweighted FRS and the LFS 2022Q1 estimates are close to the same differences for the original FRS and the LFS 2020Q1, although for all but a few groups we

---

[3] HBAI data captures the same sample as FRS, excluding only households containing absent partners. HBAI also uses the HMRC Survey of Personal Incomes (SPI) to adjust very high incomes. As a result, the weights in HBAI data are slightly different to those in FRS data, and are reweighted separately in our analysis.

cannot conclude that the reweighted FRS is strictly within the usual margin of difference from the LFS 2022 data. The mean difference in the difference of means between surveys is about £8.46, with a maximum difference in differences of £77.81 per week (measured in 2019Q1 £) for women over 50 with Standard Grade qualifications.[4]
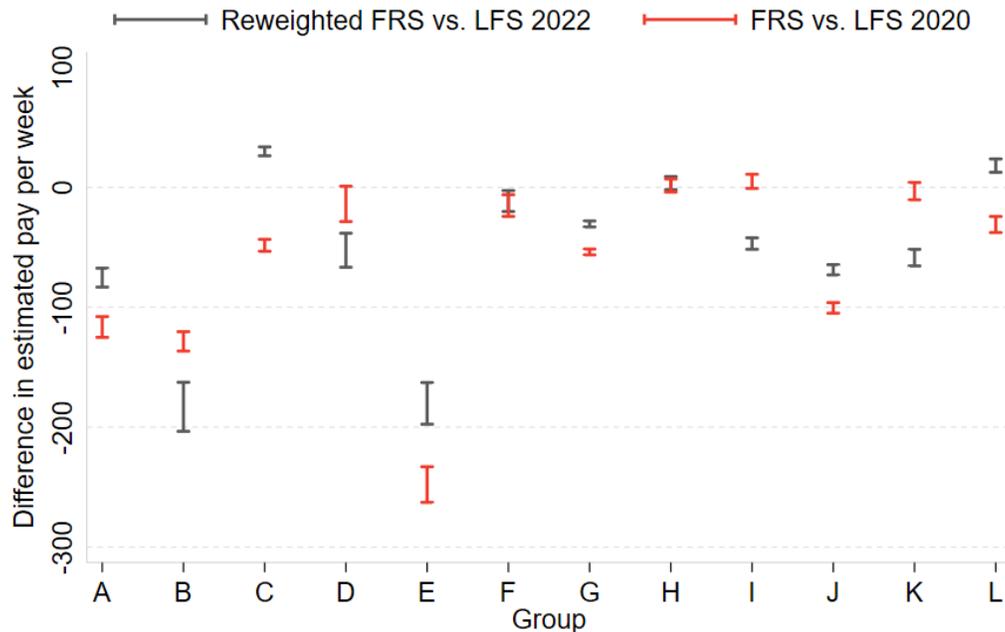

Figure A1: Differences between FRS and LFS measures of weekly pay
Source: Author calculations from quarterly LFS data 2019-2022 (ONS 2022), and annual FRS data 2019-20
(Department for Work and Pensions, Office for National Statistics, NatCen Social Research 2021)
Notes: Each range shows the 95% confidence interval for the difference between weekly pay for each group as measured in the FRS and the first quarter of LFS data for the given year. Differences are calculated as the mean value in the FRS less the mean value in the LFS. For comparison purposes, all pay is expressed in 2019Q1 £, although pay measures are expressed in nominal terms for input into the IPPR model.

Similarly, differences in weekly hours worked are generally close in the reweighted data to those in the actual data, but significantly different in statistical terms (Figure A2). For instance, the largest difference between the reweighted and original data is for women 50+ with Higher level qualifications or equivalent. For this group, the reweighted FRS underestimates hours worked by about an hour per week compared to the LFS 2022Q1 data, whereas the FRS 2019-20 data overestimated hours worked by over five hours. The mean difference in the differences between surveys is about two hours of work per week across groups. As hours of work is not in and of

---

[4] Interestingly, the poorest "fit" in terms of matching the usual difference in pay and hours between the FRS and the LFS survey data tends to be for women over 50 (see groups A, B, and C in Figures 1-2). This is possibly an indication that the labour market behaviour of this group changed during the pandemic in ways that are not well-reflected by this method of updating the FRS data. Further research on incomes and the labour market following the Covid-19 pandemic should pay particular attention to women over 50 and how their behaviour has changed since before the pandemic.

itself a determinant of benefits or other income, mismatch between the reweighted data and the unknown actual values are not expected to substantially affect the results from the IPPR model.
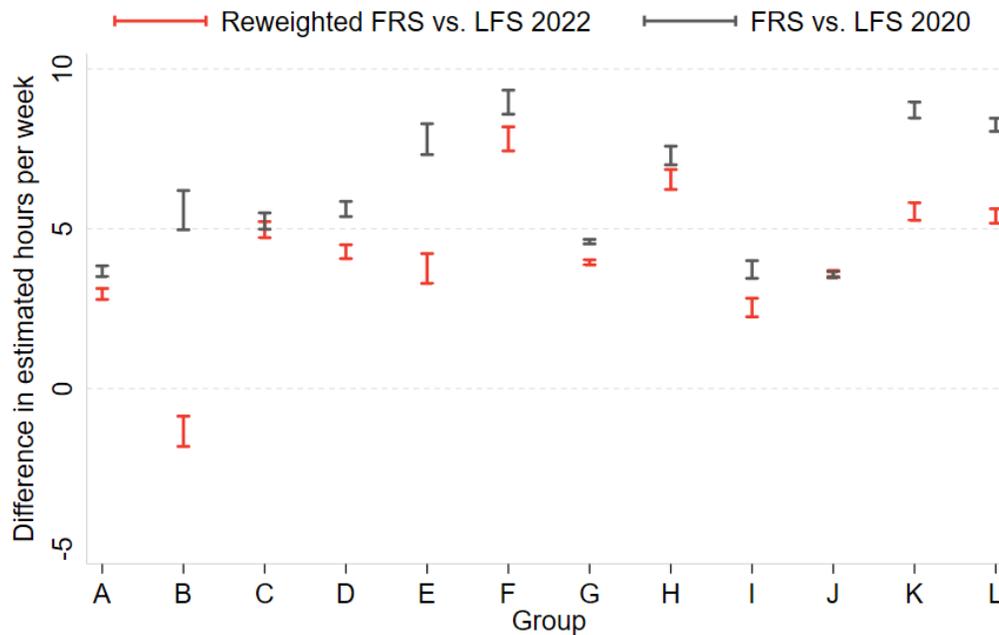

Figure A2: Differences between FRS and LFS measures of weekly hours worked
Source: Author calculations from quarterly LFS data 2019-2022 (ONS 2022), and annual FRS data 2019-20
(Department for Work and Pensions, Office for National Statistics, NatCen Social Research 2021)
Notes: Each range shows the 95% confidence interval for the difference between weekly hours worked for each
group as measured in the FRS and the first quarter of LFS data for the given year. Differences are calculated as the
mean value in the FRS less the mean value in the LFS.

Transitioning from individual-level analysis of the labour market to household-level analysis of income introduces two issues. Both arise because household weights are an average of all adults in the household, and households are often comprised of adults who were in different groups in the reweighting process. The first is that household weights may not change much if reweighting at the individual level "cancels out" when averaged. The second (and more impactful) issue is that, if household type is systematically related to the groups used for reweighting, then reweighting may change how well the data reflect the actual shares of different household types in the population.

Our third check on the accuracy of the reweighting process is to look at how the proportions of different types of households change between versions of the reweighted dataset. Table A2 shows the distribution of household types in three versions of the FRS 2019-20 data (for Scotland only): the original dataset, the dataset reweighted for population sizes in 2022Q1, and the dataset reweighted for population sizes and employment rate changes in 2022Q1.

Table A2: Distribution of household types in Scotland

Household type	% of original FRS households	% of FRS households, reweighted for population 2022Q1	% of FRS households, reweighted for population and employment 2022Q1	# sample households
Pensioner only	24.99	25.03	25.00	864
Mixed pensioner and working-age adult (50+)	4.13	4.11	4.09	127
Mixed pensioner and working-age adult (16-50)	1.39	1.48	1.44	32
Single working-age female	13.30	13.64	13.59	422
Single working-age male	10.18	10.29	10.43	254
Multiple working-age adults	46.01	45.45	45.46	1022

Source: Author calculations using FRS 2019-20 data (ONS 2022)

We conduct t-tests of the difference in population proportions between the two reweighted versions of the FRS dataset (the third and fourth columns of Table A2). No differences are significant at the 95% level of confidence. Nevertheless, small differences in proportions of household types represented in the data may still contribute to the baseline differences between the two datasets.

An alternative approach to data updating

Reweighting observations in the FRS dataset to reflect 2022Q1 population sizes and changes in employment rates is one approach to "updating" the data. Another approach to matching employment and wage data in the first quarter of 2022 would be to reassign employment randomly within groups, possibly with some further adjustments to match average wage rates within groups.

Random reassignment of employment status by group would explicitly match a target employment rate, just like the reweighting process. However, it would randomly choose people within each group to either move into or move out of employment to match the target rate. In a simple implementation, this process is completely random, and does not make the assignment based on who in each group was most likely to move in or out of employment through the pandemic period. A more complicated version of this random assignment would be to estimate a propensity to be in work in 2022Q1 contingent on group and other individual and/or household characteristics, and move individual observations in or out of work based on that propensity. This process would introduce another estimation step (and additional uncertainty) into the

updating process. Reweighting partially sidesteps these issues, although it is also agnostic to other characteristics beyond group when up- or down-weighting given individuals.

Regardless of how the random process is informed, there is still the issue of assigning a pay rate to those who are moved into work. The group average could be assigned, but as with a completely random reassignment process, that decision ignores other personal characteristics that determine pay. A more complex assignment of a pay rate could estimate pay contingent on group and additional individual or household characteristics, but again, this process would introduce more uncertainty. Further adjustments could be made to match average pay rates within groups, including multiplying each individual's pay by the same factor or using data from the LFS in 2022Q1 to estimate who is most likely within each group to have higher or lower wages than the group average. Reweighting avoids these issues by preserving the wage information of employed individuals and up- or down-weighting people within each group as needed to meet the target group average employment rate.

A further challenge of the employment reassignment and pay adjustment approach is mirroring changes in population by group between 2019-20 and 2022Q1. Beyond creating entirely new or duplicate observations, the only way to make these adjustments is by reweighting. Updating the baseline FRS to match 2022Q1 population sizes would therefore require reweighting regardless of the approach used to match employment data. If the population adjustment is not done, then any differences in the updated data compared to the original FRS 2019-20 would not take into account how the population structure changed during the pandemic. Since reweighting can be easily adjusted to accommodate matching both population changes and target employment rates, that approach is preferred over direct reassignment and adjustment of key variables.

References

Department for Work and Pensions. 2022. *Households Below Average Income, 1994/95-2020/21*. 16th ed. UK Data Service. https://www.doi.org/10.5255/UKDA-SN-5828-14.

Department for Work and Pensions, Office for National Statistics, NatCen Social Research. 2021. 'Family Resources Survey, 2019-2020'. SN: 8802. Family Resources Survey. UK Data Service. https://www.doi.org/10.5255/UKDA-SN-8802-1.

Kumar, Ashwin. 2022. 'IPPR Tax-Benefit Microsimulation Model'. Manchester, UK: Manchester Metropolitan University.

Northern Ireland Statistics and Research Agency, Central Survey Unit, Office for National Statistics, Social Survey Division. 2022. 'Quarterly Labour Force Survey, 2010-2022'. Dataset. Labour Force Survey. UK Data Service. https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6548#!/details.

Office for National Statistics. 2022. 'Income and Earnings Statistics Guide'. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/aguidetosourcesofdataonearningsandincome.

Pacifico, Daniele. 2014. 'Sreweight: A Stata Command to Reweight Survey Data to External Totals'. *The Stata Journal* 14 (1): 4–21. https://doi.org/10.1177/1536867X1401400103.

StataCorp. 2021. 'Stata Statistical Software: Release 17'. College Station, TX: StataCorp LLC.