UK Statistics Authority Inclusive Data Consultation – Submission

Arun Advani

26/03/2021

## About me

Arun Advani is Assistant Professor of Economics and Impact Director of the CAGE Research Centre at the University of Warwick. He is also a Research Fellow at the Institute for Fiscal Studies, and a Visiting Fellow at the LSE International Inequalities Institute. He studies issues of tax compliance and tax design, with a particular focus on those with high incomes or wealth.

## What is your main area of interest or your broad area of research, in terms of topic and groups that you're interested in?

Inequality, in income, wealth, educational attainment and university outcomes. Inequality here incorporates both differences in outcomes within the whole population, and differences between particular groups, whether defined by age, sex, ethnicity, region, socioeconomic status, or something else.

## Are there any questions you are currently unable to answer because of a lack of data or evidence? If there are any, please tell us what they are.

There are many such questions. A (non-exhaustive) list of questions I am currently interested in trying to answer, and don't currently have the data to study includes:

1. The distribution of wealth at the very top (see Advani, Bangham and Leslie, 2020). There are problems of both undercoverage and insufficient granularity at the top of the wealth distribution – the very wealthiest households are undersampled, and the relatively small number of observations at the top (just 24 households/39 individuals representing the likely 22,000 individuals with wealth in excess of £10m) means relatively little can be learned about variation in this group.
2. The distribution of wealth by ethnicity (see Advani, Bangham and Leslie, 2020; Runnymede Trust, 2020). Although differences in average wealth by ethnicity (Runnymede Trust, 2020) and share of households by ethnicity above £500k in wealth (Advani, Bangham and Leslie, 2020) can be calculated using the Wealth and Assets Survey (WAS), differences in share of households above £1m or higher thresholds cannot be calculated in the data because there are too few households headed by individuals of most ethnic minority groups. This means, for example, that the distributional effect of the proposals by the Wealth Tax Commission (Advani, Chamberlain and Summers, 2020) couldn't be studied by ethnicity.
3. The distribution of income by ethnicity when focused particularly on top incomes. Survey data (sometimes) include ethnicity, but top incomes are poorly measured here. Hence these are typically supplemented with administrative data to do a 'top income adjustment' (various versions of this exist, from ONS, DWP and other organisations). These corrections no longer allow the distribution of income by ethnicity to be obtained. Also, recent work highlighting that much of the issue in top income

adjustments is the coverage of capital income means that capital incomes are hard to accurately measure by ethnicity (since again, they are well measured only in administrative data that exclude ethnicity) (Advani, Ooms and Summers, 2021).

4. The distribution of incomes by household. As above, when there is an interest in correctly measuring top incomes, administrative data are needed either directly or combined with survey data. While I have extensively used the Survey of Personal Incomes (based on admin data from HMRC) to study the distribution of individual incomes, there are no links in either the public use tapes version of the SPI nor the secure Datalab version of the underlying administrative data, that allow households to be linked up. This makes it impossible to implement a top incomes adjustment to household survey data to give an accurate picture of the distribution of household incomes.
5. Educational choices and outcomes by gender as well as sex. There is good evidence that education choices and outcomes vary substantially by sex (Advani, Sen and Warwick, 2020,2021; Advani, Drayton, van der Erve and Warwick, 2021), but despite recent interest in gender identity issues, gender is typically not measured.
6. Educational choices and outcomes by socioeconomic status. There is good evidence that education choices and outcomes vary substantially by crude socioeconomic proxies such as Free School Meals receipt (Advani, Drayton, van der Erve and Warwick, 2021), and other socioeconomic proxies are available based on local area information, but household income information is not included in the relevant datasets.

## Please tell us the reasons why you are unable to answer these questions. (Please provide further details below)

See comments alongside each point above.

## Are you currently able to access the data you need for your purposes?

Yes, some of it

## If you are not able to access all the data or evidence you need for your purposes, what data are you unable to access and what are the barriers to you accessing this data?

As noted earlier, some of the data I would like do not appear to currently exist.

Other data that do exist – for example the Longitudinal Education Outcomes (LEO) data – are not easily available, and I currently do not have access.

Other data to which I do have access – for example HMRC data via the Datalab – are only available at a single physical location, historically for relatively short opening hours, and with limited computational resource available (constraining the number of people who can be there as well as the analysis that can be done).

## Are there any issues with how the data or evidence that you currently rely on are presented? If so, please provide details.

Not for my purposes, since I use the microdata rather than the summaries.

However, although I don't rely on it, I do have concerns about how some of the wealth-ethnicity work by the ONS is presented (see here: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/articles/householdwealthbyethnicitygreatbritain/april2016tomarch2018) The results here in Fig 1 show differences in wealth after controlling for lots of things, including tenure and socioeconomic status. These items are outcomes, rather than demographic characteristics, and conditioning on these can give a rather strange and misleading impression to the casual reader. For example, by controlling for tenure Indian ethnicity individuals appear at first glance not to be doing that well in terms of wealth accumulation. However, as can be seen in Figure 3, this is not the case. Instead what one should learn from Figure 1 is that Indian ethnicity individuals invest a relatively large share of their wealth into owner-occupied housing, so that conditional on being homeowners (and other characteristics) they have apparently less wealth ownership relative to White British individuals. This is not something that a lay reader is likely to correctly interpret.

## Thinking of all the issues you may have experienced with the data and evidence, which of the following improvements would you like to see? Please provide details.

### *Fill gaps in the current data?*

As above, data collection on wealth at the very top of the distribution, as well as both top incomes and top wealth by ethnicity would be useful improvements.

Improvements in the collection of data on the wealthiest households could be achieved through the implementation of a 'booster' sample to WAS, which targets high wealth households. These households could be sampled using information on beneficial owners of registered companies, which can be obtained from Companies House or Bureau van Dijk's Orbis dataset. Alternatively, or additionally, the ONS could collaborate with HMRC's Wealthy unit, to construct a sampling frame from high wealth households that are known to the tax authority.

Improvements in the measurement of top wealth by ethnicity could be achieved by oversampling ethnic minority headed households using the existing WAS survey instrument. These households could be sampled by oversampling otherwise equivalent households from areas that have higher ethnic minority populations, using census data to identify at a hyper-local level where these areas are, or by directly using the census to determine which households are additionally sampled if this is permitted.

### *Address problems with measurement?*

Ultimately improving coverage at the top of the wealth distribution can best be done through additional surveys. However, such data cannot as easily be collected for the past. Hence, in the absence of collecting historical data to align with previous waves/rounds of WAS, creating improved weights to better account for missing top wealth would help give a better measure of aggregate wealth using existing waves/rounds.

### *Improve the level of detail available?*

As above, additional questions to learn about gender and socioeconomic background in education data would be valuable.

### *Enable greater access to data?*

As above, access to administrative incomes data linking individuals within households would be incredibly useful.

More broadly, making access to existing administrative tax data possible without physical presence at a single location would make it much easier to do research using those data, for example via ONS Secure Lab. Such access is common in many other countries.

Relatedly, making LEO data more easily available would allow more research into long run educational outcomes, including understanding what happens once individuals leave the education system.

### *Improve presentation of the evidence?*

See comment above on presentation of evidence on wealth differences by ethnicity.

## Please tell us about how important it is for your purposes that data or evidence are comparable across different geographies, for example, across the 4 countries of the UK, internationally or at a more local level? Please give details of what geographies you would like to be able to compare across.

My research is mainly in the UK, so ideally the same data would be across the four UK nations, or there would at least be some reasonable way to harmonise data to make them comparable after collection. In some cases, the latter is clearly the only option e.g. in education the exams taken by most Scottish school students differ from those taken by most English students. But in others, such as ethnicity, there is in principle no reason the same categorisations could not be used.

International comparisons are often useful, but my experience is that even in something like National Accounts, which in principle have a common methodology internationally, there are still some differences. I therefore wouldn't expect other data to be collected in a way which is internationally comparable, not least because other countries are not comparable to each other, so it is not clear what the benchmark should be. My view is that unless there is an agreed international standard, it would be much better to focus on the best possible approach to measurement for the UK context. Within the UK I would like to be able to compare across geographies, both across the four nations, but also often at more local (local authority, parliamentary constituency, LSOA) level.

## What change to the current data or evidence would you most like to see to be able to answer the questions that are most relevant to you?

If I can only have one, then I would go for better measurement of top wealth, both to better capture the extremely wealthy and to provide more detail on variation in wealth within ethnicity among those with high wealth.

## References

Advani, A., Bangham, G. and Leslie, J. (2020). "The UK's wealth distribution and characteristics of high-wealth households." *Wealth and Policy Working Paper 101*.

Advani, A., Chamberlain, E. and Summers, A. (2020). "A wealth tax for the UK". *Wealth Tax Commission Final Report*.

Advani, A., Ooms, T. and Summers, A. (2021). "Missing incomes in the UK: New evidence and policy implications". *CAGE Working Paper 543*.

Advani, A., Sen, S. and Warwick, R. (2020) "Ethnic diversity in UK economies". *IFS Briefing Note BN307*.

Advani, A., Sen, S. and Warwick, R. (2021) "A level Economics is a gateway to the economics profession". *IFS Observation*.

Runnymede (2020). "The colour of money: how racial inequalities obstruct a fair and resilient economy". *Runnymede report*.