User guide for code underlying the UK Wealth Tax Commission's revenue modelling and distributional analysis

Code

The code folder provided contains 22 Stata (.do) files underlying the revenue modelling and distributional analysis presented by the UK Wealth Tax Commission. The results of this analysis are published in Advani, Hughson and Tarrant (2021). A brief description of each file is provided below. All code can be run via the master do-file, which is also where user preferences (directory paths, tax rate preferences etc.) should be set by the user.

All analysis results are written to excel files. The directory for saving these is set by the user within the master do-file.

Name of do-file	Description
0.master.do	Sets path directories and tax modelling preferences, runs all other do-files
1.round_6_alt.do	Cleans round 6 of the Wealth and Assets Survey, producing cleaned individual-level and household-level datasets.
2.pareto_adjustment.do	Implements a Pareto adjustment using the Sunday Times Rich List
3.define_sample.do	Excludes unwanted individuals (non-residents) from the analysis sample
4.population_counts.do	Produces population counts by range of total wealth
5.model_flat_tax_annual.do	Revenue calculations for a flat, annual tax
6.model_flat_tax_annual_banded.do	Revenue calculations for a flat, banded, annual tax
7.model_flat_tax_oneoff.do	Revenue calculations for a flat, one-off tax
8.model_flat_tax_oneoff_banded.do	Revenue calculations for a flat, banded, one-off tax
9.model_progressive_tax_annual.do	Revenue calculations for a progressive, annual tax
10.model_progressive_tax_oneoff.do	Revenue calculations for a progressive, one-off tax
11.model_iht_reforms.do	Revenue calculations for an alternative reform to Inheritance Tax
12.council_tax_analysis.do	Revenue calculations for an alternative reform to Council Tax
13.dividend_reform.do	Revenue calculations for an alternative reform to the dividend tax rate
14.liquidity_analysis.do	Liquidity analysis for a range of annual and one-off wealth taxes
15.who_pays.do	Distributional analysis for a range of annual and one-off wealth taxes
16.valuation_cost_analysis	Valuation cost analysis for a range of annual and one-off wealth taxes
17.iht_distriubtional_analysis.do	Distributional analysis for alternative reforms to Inheritance Tax
STRL-1-bucket_rank_level_averages.do	Preparation of Sunday Times Rich List: bucket-level averages for age, residence and nationality
STRL-2-impute_missing_info_byindividual.do	Preparation of Sunday Times Rich List: impute missing values for age and residence

automating_revtarget.do	Computes rates required to generate different revenue targets under a flat annual or one-off wealth tax (can be used to determine rate preferences)
different_bases.do	Revenue calculations under different wealth tax bases
tax_by_wealth_graph.do	Computes tax paid as a function of wealth under a range of annual and one-off wealth taxes

Data

The code requires five datasets as inputs (file names given in brackets)[1]:

- **Raw individual-level WAS: Round 6.** This can be downloaded here: http://doi.org/10.5255/UKDA-SN-7215-13. (was_round_6_person_eul_mar_20.dta)
- **Raw household-level WAS data: Round 6.** This can be downloaded from the same link as above. (was_round_6_hhold_eul_mar_20.dta)
- **Raw individual-level WAS data: Wave 5.** This can be downloaded from the same link as above. (was_wave_5_person_eul_dec_19_final.dta).
- **Individual-level 2020 Sunday Times Rich List:** The latest public version can be found here: https://www.thetimes.co.uk/sunday-times-rich-list. However, replicating the analysis in the paper requires the addition of variables that are not publicly available, in particular the entrants' date of birth (DOB). This information was provided to us by the data owner and cannot be shared by us. Doing so would be a breach of privacy of the individuals since their DOB is not in the public domain, unlike the other information published in STRL. We explain below which variables are required and how the dataset should be structured in order to run the code with or without the information on DOB (strl_2020_clean_byindividual.dta).
- **Raw data from the Survey of Personal Incomes: 2017.** This can be downloaded here: http://doi.org/10.5255/UKDA-SN-8582-1. These data are required for producing revenue estimates under an alternative reform to dividend tax rates. Users who do not wish to replicate this part of the analysis (by running the do-file '13.dividend_reform.do') need not download SPI data. (2017_spi.dta).

Using information on residence and nationality from Companies House

In our main analysis we only include STRL individuals who are UK resident. To obtain information on residence (and nationality), we link individuals named in the STRL to Companies House records contained in the 'Persons with Significant Control' and 'Officers' datasets. These datasets are publicly available,[2] however the linking variables we use (full name and DOB) are not. Users who do not have access to this information may replicate our analysis without excluding non-resident individuals (i.e. using the full STRL). We set out below the requirements for each option. The code is designed to run using either option, provided the data are set up as described below.

1. Using the STRL with Companies House data

The cleaned STRL dataset 'strl_2020_clean_byindividual.dta' is created as follows:

---

[1] We use the 13th Edition of the WAS released in September 2020. File names may differ in alternative download versions.

[2] PSC data can downloaded here: http://download.companieshouse.gov.uk/en_pscdata.html. Officers data can be searched online or using the Companies House API: https://developer-specs.company-information.service.gov.uk/companies-house-public-data-api/reference/officers.

1. For each entry in the list, the forenames and surnames of each named individual are extracted and linked to the individual's DOB. A string variable containing each named individual's full name and DOB is created.
2. This string variable is used to link individuals to records in Companies House 'Persons with Significant Control' and 'Officers' datasets.
3. For each matching Companies House record, we extract the individual's nationality and country of residence.
4. For each record, we create two indicator variables: one indicating if the individual is a UK resident according to Companies House, and one indicating whether the individual is a British national.
5. The dataset is collapsed to one row per individual. Where there are conflicts in the residence and nationality indicators for a given individual, we take the modal value.

The result is an individual-level dataset in which we define the following variables:

- rank = the rank of the STRL entry in which the individual appears
- item_id_richlist = rank – 1
- worth_rank = the published worth (£) held by the individual(s) in the STRL entry in which the individual appears
- num_named = the number of named individuals in the STRL entry in which the individual appears
- strl_flag = 1 (this is used to identify STRL observations after appending to WAS data)
- ukresident = 1 if the individual is UK resident, 0 otherwise
- britishnational = 1 if the individual is British national, 0 otherwise
- male = 1 if the individual is male, 0 otherwise (this information is added manually)
- age = individual's age (based on year of birth)
- over80 = 1 if individual is over the age of 80, 0 otherwise

If using this option, the global 'using_companies_house' should be set to 1 in the master code.

2. Using the STRL without Companies House data
Without information on each individual's DOB, it is not possible to link the STRL to Companies House records. In this case, users may choose to run the analysis using the full STRL. To do so, the dataset 'strl_2020_clean_by_individual.dta' should be created as follows.

1. Starting from the raw STRL, a variable indicating the number of individuals in each entry should be created (this can be done manually).
2. Duplicated entries should be created for each entry such that the number of occurrences of a given entry matches the number of named individuals, i.e. there should be a row for each named individual. These rows need not be separately identifiable.

The following variables should be created in the final dataset:

- rank = the rank of the STRL entry in which the individual appears
- item_id_richlist = rank – 1
- worth_rank = the published worth (£) held by the individual(s) in the STRL entry in which the individual appears
- num_named = the number of named individuals in the STRL entry in which the individual appears
- strl_flag = 1 (this is used to identify STRL observations after appending to WAS data)

- male = 1 if the individual is male, 0 otherwise (this information is added manually)

Note that as there is no information on age, STRL individuals will be excluded from distributional analysis by age. Likewise, if the variable 'male' is not created.

If using this option, the global 'using_companies_house' should be set to 0 in the master code.

<h1 style="color:red">References</h1>

Advani, A., Hughson, H. and Tarrant, H. (2021) "Revenue and distributional modelling for a UK wealth tax". *Fiscal Studies,* 42.

HM Revenue and Customs, KAI Data, Policy and Co-Ordination. (2019). Survey of Personal Incomes, 2016-2017: Public Use Tape. [data collection]. UK Data Service. SN: 8582, http://doi.org/10.5255/UKDA-SN-8582-1

Office for National Statistics, Social Survey Division (2020), Wealth and Assets Survey, Waves 1–5 and Rounds 5–6, 2006–18. *13th Edition.* UK Data Service SN: 7215, http://doi.org/10.5255/UKDA-SN-7215-13

For more information on the UK Wealth Tax Commission, see https://www.ukwealth.tax/