

---

# READ ME — Dynamic cohort simulation model files

## 1 About

This document provides details of the files and code for running the dynamic cohort simulation model used in the Productivity Commission’s report on Wealth Transfers and their Economic Effects. The model builds projections of future wealth and wealth transfers for the primary purpose of understanding the impact of transferred wealth on wealth distributions within and across generations in Australia. Details of the results and methodology are provided in chapter 3 and appendix C of the report.

The model was written in the R programming language. All R scripts used to build the model have been provided. Underlying unit record data, which are required to run scripts for estimating model parameters, are not provided and access should be obtained independently. The model used data from the *Household Income and Labour Dynamics Australia* (HILDA) Survey (Restricted Release 19) and the ABS 2015-16 Household Expenditure Survey.

## 2 Files and folder structure

### Cohort simulation.Rproj

Opening this R Project in RStudio and loading scripts within the R Project will ensure that the working directory is set correctly and relative file paths will work.

### R scripts

The ‘R scripts’ folder contains all R scripts for the model. Table 1 provides a short description of the relevant scripts and the order they should be run, from reading in the data to producing model results.

Outputs from R scripts were saved in the .qs format using the ‘qs’ R package due to its advantages of relatively fast read and write times, small file sizes, and ability to retain useful information such as variable and value labels from HILDA .dta files (Ching 2021; Miller 2020).

**Table 1 R scripts description and step order**

<i>Step</i>	<i>Subfolder</i>	<i>File name</i>	<i>Description</i>
1	1 - Master HILDA data set up	Read and format HILDA.R	Reads in HILDA .dta files with consistent column names, and saves in .qs format.
		Read HILDA indiv wealth.R	Reads in estimated individual wealth data from Stata analysis, and saves in .qs format.
		Read HILDA first estate bequest.R	Reads in estimated first estate bequests from Stata analysis, and saves in .qs format.
		Read HILDA longitudinal weights.R	Reads in HILDA longitudinal weights data and saves in .qs format.
2	Simulation model	2 - Master data and parameters script.R	Runs all R scripts in the “Data and parameters” subfolder in the correct order to produce the dataset for the initial cohort population and estimated parameters required for the base scenario model. Resulting data frames are saved in the ‘Input data’ folder.
3	Simulation model	3 - Cohort simulation model - prep.R	Draws together initial cohort population data and all parameters from step 2. Then creates a list of data frames, one data frame per year, with each containing the relevant parameters for each cohort in that year. Resulting lists for the base scenario and historic returns scenario are saved in the ‘Output data’ folder. These are used as inputs into step 5.
4	Simulation model	4 - Master scenario analysis data and parameters script.R	Runs all R scripts in the “Data and parameters/Scenario adjustments” subfolder to produce new lists of data frames (equivalent to the output of step 3), but with parameters adjusted for each scenario analysed. Resulting lists are saved in the ‘Output data’ folder. These are used as inputs into step 5.
5	Simulation model	5 - Cohort simulation model - run scenarios.R	Draws on ‘Simulation model/Cohort simulation model - source for n and wealth.R’ to generate population, wealth, and wealth transfer sizes at each year, each cohort and for each scenario both with and without the existence of future wealth transfers. Results are saved in the ‘Output data’ folder.
6	Simulation model	6 - Cohort simulation results - base scenario.R	Takes results from the base scenario in step 5 and produces charts and values used in the report. <sup>a</sup>
7	Simulation model	7 - Cohort simulation results - scenario analysis.R	Takes results from select scenarios in step 5 and produces charts and values relating to scenario analysis in the report. <sup>a</sup>

<sup>a</sup> Chart formatting using PCcharts package functions not available outside of the Productivity Commission.

## Input data

The ‘Input data’ folder contains files with estimated parameter values. Files that contain identifying HILDA data are not included but can be created by running the relevant R scripts in step 2 of table 1.

The ‘Intermediate input data’ subfolder contains publicly available data files that were used in estimating model parameters.

---

## Output data

The ‘Output data’ folder is required to save the results of running the scripts in table 1. This folder is empty in the zip folder due to file size and potential confidentiality issues relating to small cohorts in the initial population.

## Charts

The ‘Charts’ folder contains charts produced in steps 6 and 7 of table 1, used in the report.

## 3 R packages used

R packages listed in table 2 need to be installed to run the full set of R scripts. The ‘Master package loading.R’ script loads packages in the listed order to prevent errors due to function name conflicts.

---

**Table 1.2 Attached R packages<sup>a</sup>**

<i>Package</i>	<i>Version</i>	<i>Package</i>	<i>Version</i>
qs	0.23.6	scam	1.2-11
haven	2.3.1	data.table	1.14.0
labelled	2.8.0	tidyverse	1.3.0
readxl	1.3.1	tibble	3.0.6
readabs	0.4.8	tidyr	1.1.2
reldist	1.6-6	readr	1.4.0
lattice	0.20-35	purrr	0.3.4
survival	3.2-7	dplyr	1.0.4
Formula	1.2-4	stringr	1.4.0
ggplot2	3.3.3	forcats	0.5.1
Hmisc	4.5-0	stringi	1.5.3
forecast	8.14	ggalluvial	0.12.3
nlme	3.1-137	PCcharts <sup>b</sup>	3.1.0
mgcv	1.8-24	devEMF	4.0-2

<sup>a</sup> Excludes base R packages. <sup>b</sup> Used to format charts and not available outside of the Productivity Commission. External users can remove these code lines when creating charts.

---

---

## 4 References

- Ching, T. 2021, *Package 'qs'*, <https://cran.r-project.org/web/packages/qs/vignettes/vignette.html> (accessed 3 September 2021).
- Miller, S. 2020, *Comparing the Read/Save Times for RDS, {qs}, and {fst} Formats (or: All Hail {qs})*, Steven V. Miller, <http://svmiller.com/blog/2020/02/comparing-qs-fst-rds-for-bigger-datasets/> (accessed 3 September 2021).