

READ ME FILE

Title: The Determinants of Mortgage Defaults in Australia – Evidence for the Double-trigger Hypothesis

Author: Michelle Bergmann

Description

This 'read me' file contains details of the data and code used to generate the results reported in RDP 2020-03.

If you make use of any of these files you should clearly attribute the author in any derivative work.

Data

The following data sources were used:

- Loan-level data:
 - Obtained from the Securitisation Dataset – not available for release. For information on permitted data users, see <https://www.rba.gov.au/securingisations/reporting-guidelines/index.html>. For more information on the dataset, see K Fernandes and D Jones (2018), 'The Reserve Bank's Securitisation Dataset', RBA *Bulletin*, December, available at <https://www.rba.gov.au/publications/bulletin/2018/dec/the-reserve-banks-securitisation-dataset.html>.
- Regional data:
 - SA3-level housing price indices and turnover ratios: obtained from CoreLogic – not available for release.
 - SA3-level unemployment rate: calculated using Australian Bureau of Statistics (ABS) Census of Population and Housing data, obtained through TableBuilder, 2016 – not available for release. The data can be accessed by logging in to TableBuilder Basic on the ABS website.
 - SA3-level employment by industry: obtained from the ABS Census of Population and Housing – DataPacks – General Community Profile, 2016 (SA3_mining.csv).
 - Postcode-level SEIFA indices: obtained from ABS Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), ABS Cat No 2033.0.55.001, 2016 (SEIFA_IRSAD.csv).
 - State-level average weekly earnings: obtained from Average Weekly Earnings, ABS Cat No 6302.0, May 2019 (AWE.csv).
 - Operating mine locations: obtained from GeoScience Australia at the Australian Atlas of Minerals Resources, Mines, and Processing Centres downloads, February 2015 (operating_mines.csv).

Data for figures are not publically available due to confidentiality reasons.

Code

The results reported in this RDP were generated using R 3.5.1 (64 bit), RStudio v1.1.453 and Stata 13.0.

The code is run in two parts:

- Part 1 analyses entries to 90+ day arrears over the period 2015:M7–2019:M6 for loans originated since 2013. This includes the stage 1 Cox model.

- Part 2 analyses transitions of loans from 90+ day arrears over the period 2015:M7–2019:M6. This includes the stage 2 Cox model.

The code used for data cleaning and transformations are not included in this archive due to confidentiality reasons.

The code used to estimate the first- and second-stage Cox models, as well as the multinomial logit models as robustness checks, are included in this archive and are as follows:

- first_stage_hazard_models.R
- first_stage_baseline_hazard.R
- first_stage_MNL.do
- second_stage_hazard_models.R
- second_stage_MNL.do

22 July 2020