Data & Code Downloads for all IDEs (see video)

Note: do not rename the files and save/extract into the SAME folder of your choice

Data Python & R:

DATA (DCR.CSV)

Code for Python:

CODE (DCR.PY, ZIPPED)

Code for R:

CODE (DCR.R, ZIPPED)

Optional: environment for Anaconda Navigator, Book Edition 1 Version 3

ENVIRONMENT (DCR3.YAML, ZIPPED)

Data & Code Download for Python Google Colab users

DATA & CODE & JUPYTER NOTEBOOK

Data Description:

The data set is in panel form and reports origination and performance observations for 5,000 residential U.S. mortgage borrowers over 60 periods. For training purposes we work with a 10% sample. The full sample for 50,000 loans can be downloaded here. The periods have been de-identified. As in the real world, loans may originate before the start of the observation period (this is an issue where loans are transferred between banks and investors as in securitization). The loan observations may thus be censored as the loans mature or borrowers refinance. The data set is a randomized selection of mortgage-loan-level data collected from the portfolios underlying U.S. residential mortgage-backed securities (RMBS) securitization portfolios and provided by International Financial Research (www.internationalfinancialresearch.org).

Key variables include:

- id: borrower id
- time: time stamp of observation
- orig_time: time stamp for origination
- first_time: time stamp for first observation
- mat_time: time stamp for maturity

- LTV_time: loan to value ratio at observation time, in %
- interest_rate_time: interest rate at observation time, in %
- rate_time: risk-free rate
- hpi_time: house price index at observation time, base year=100
- gdp_time: GDP growth at observation time, in %
- uer_time: unemployment rate at observation time, in %
- REtype_CO_orig_time: real estate type condominium: 1, otherwise: 0
- REtype_PU_orig_time: real estate type planned urban developments: 1, otherwise: 0
- REtype_SF_orig_time: single family home: 1, otherwise: 0
- investor_orig_time: investor borrower: 1, otherwise: 0
- balance_orig_time: outstanding balance at origination time
- FICO_orig_time: FICO score at origination time, in %
- LTV_orig_time: loan to value ratio at origination time, in %
- Interest_Rate_orig_time: interest rate at origination time, in %
- state_orig_time: US state in which the property is located
- hpi_orig_time: house price index at observation time, base year=100
- default_time: default observation at observation time
- payoff_time: payoff observation at observation time
- status_time: default (1), payoff (2) and non-default/non-payoff (0) observation at observation time
- lgd_time: LGD assuming no discounting of cash flows
- recovery_res: sum of all cash flows received during resolution period