Title: A History of Australian Equities

Author: Thomas Mathews

Description

This 'read me' file contains details of the code used to replicate the figures and series for RBA Research Discussion Paper No 2019-04. Publically available plotting data for figures appearing in the RDP can be found in the spreadsheet 'rdp-2019-04-graph-data.xlsx'.

Please note that the underlying data are copyright and cannot be released with this paper.

However, if you have access to them yourself, you can run this code to replicate the results.

The data providers are noted below.

Data sources

Historical share-level data: ASX (originally sourced from Sydney Stock Exchange Limited Gazettes). The ASX do not host these data as they were collected by the RBA: if you acquire permission from the ASX then the RBA should be able to provide the data to you. It is at the discretion of the ASX whether they choose to do this.

Modern time series data: Refinitiv (series codes available on request).

Modern company-level data: Bloomberg (series codes available on request).

Underlying data for Figure 16 (which includes company founding dates and market capitalisation for a range of countries) are from S&P Capital IQ. These are generated using the 'Equity Screening' tool.

Code

The code was written using R 3.5.1, with RStudio v1.1.453.

The following packages were used:

- lubridate_1.7.4
- zoo_1.8-6
- tibble_2.1.3
- readxl_1.3.1
- dplyr_0.8.1
- plyr_1.8.4
- tidyr_0.8.3

How to run the code:

The 'R code' folder includes the following R scripts:

- Master.R - runs all the scripts below in the correct order, and calculates some basic time series from the company-level data.
- additional functions.R - some useful functions for repeated tasks.
- load data.R - loads and formats the data files noted below, for use in the other scripts.
- graphs.R - generates the output csv files behind each graph. Mostly formatting, but in some cases there is some calculation involved.

- data appendix.R - generates the files used in preparing the data appendix - all the time series shown in the graphs, as well as additional ones (e.g. sectoral breakdowns not shown in the paper).
- company age.R - generates the data behind figure 16, which uses additional information the founding year of companies sourced from S&P.

The underlying data need to be in a folder called 'Data' (this is blank in the version provided, due to the copyright restrictions noted above).

Underlying data files (if available) are:

- Sydney Stock Exchange Data - Quarterly.csv - quarterly company-level data for the variables noted in the paper, for the top 100 companies.
- Sydney Stock Exchange Data - Annual.csv - annual market capitalisation for all the companies on the Sydney Stock Exchange.
- Other time series.xlsx - Quarterly time series from other sources which are used for various purposes in the code. This includes bond yields, GDP, and modern equivalents of all the series calculated in the paper from the previous two files (eg price indices, PE ratios).
- Modern company-level data.xlsx - Quarterly company-level data for the S&P/ASX 200 for Bloomberg, from 2000. Used to extend some charts for comparing historical data to modern.

Running the script 'Master.R' (if you have the data) should output the graph data for every graph in a range of numbered csv files, in a subfolder called 'graphs'. If you open the project provided it will be in that folder, otherwise it will be in your working directory.

The output files are provided. The data for the tables (where calculations are required) are in R objects saved into your environment.

More information about what each script does is included in its head.

Additional data

All time series calculated for this paper, including breakdowns by broad sector, are also available in the 'Data appendix' folder. These are presented in two formats: a 'narrow' .csv file, and a 'wide' .xlsx file. Metadata are included in the .xlsx file.

Please feel free to contact the author with any queries (email address is provided on the front of the paper).

Note: Data were re-published 1 August 2019 to fix incorrect dates provided for some modern time series; the author is grateful to Philipp Hofflin for bringing this to his attention. Series calculated from the RBA dataset were unaffected.