

Quantitative Investing: From Theory to Industry

Data Readme

Lingjie Ma, May 19, 2020

The data for this book is used for teaching and learning purposes only. Please do not use and distribute the data for real investing or business purposes.

Below we list the main data sets used in the book by chapter. Each data set has a name with a short description, frequency, and date range. The dates for daily and monthly data are denoted using the formats `yyyymmdd` and `yyyymm`, respectively. There is also a short note specifying the main use of the data in the book. Most of the data sets are in CSV format, which is easy to open in Excel and can be imported into R using the `read.csv` command.

Chapter 2

Chapter2_SP500_dailyIndexPrice.csv

Description: S&P 500 Index daily price movements

Frequency: daily

Date range: 19500104-20190418

Note: The data are used for univariate analysis of the US stock market.

Chapter2_SP500_monthlyIndexPrice.csv

Description: S&P 500 Index monthly price movements

Frequency: monthly

Date range: 187101-201905

Note: The data are used for univariate analysis of the US stock market.

Chapter 3

Chapter3_SSE_dailyIndexPrice.csv

Description: Shanghai Stock Exchange Composite daily price movements

Frequency: daily

Date range: 19901219-20190409

Note: The data are used for analysis of the Chinese stock market.

Chapter3_SP500_CSI300_dailyIndexPrice.csv

Description: S&P 500 and CSI 300 daily price movements, merged data

Frequency: daily

Date range: 20050104-20190412

Note: The data are used for bivariate analysis of the US and Chinese stock markets.

Chapters 4 and 5

Chapter4and5_US_largeCap_signals_rawdata.csv

Description: US large-cap security-level data with IDs and factors

Frequency: monthly

Date range: 199501-200412

Note: The data are used for signal treatment, multi-factor analysis, and stock selection strategies in Chapters 4 and 5.

Chapter 6

Chapter6_Gold_WTI_price.csv

Description: prices of gold and oil (WTI)

Frequency: monthly

Date range: 196804-201906

Note: The data are used for time series analysis, such as unit root and cointegration, in Chapter 6. The prices are the monthly average of daily values. The data are also used in Chapter 8 for asset returns of commodities.

Chapter6_AAL_UAL_price.csv

Description: stock prices of American Airlines and United Airlines

Frequency: daily

Date range: 20060628-20160111

Note: The data are used in a pair trading example in Chapter 6.

Chapter 7

Chapter7and8_returnMatrix.RData,

Chapter7and8_sectorMatrix.RData,

Chapter7and8_mktcapMatrix.RData

Description: The three R data sets contain returns data, industry classifications, and market capitalization data, respectively.

Frequency: monthly

Date range: 199412-201212

Note: The data are used for mean-variance portfolio construction in Chapter 7; note that these data sets are also used in Chapter 8 for the construction of quantile portfolios.

Chapter 8

Chapter8_Russia_MOEX50.csv, Chapter8_Japan_Nikkei225.csv,

Chapter8_USDEUR.csv, Chapter8_USDJPY.csv

Description: Asset returns of a Russian stock index, a Japanese stock index, the euro, and the Japanese Yen

Frequency: daily

Date range: 19970923-, 19650105-, 20000103-, 20000103-, all end on 20190920

Note: they are used for the tail analysis of distribution of asset returns

Chapter8_Gold_example_factors.csv

Description: prices of gold and related factors

Frequency: monthly

Date range: 196804-201012

Note: The data are used for gold price forecasting using quantile regression.

Chapter7and8_returnMatrix.RData,

Chapter7and8_sectorMatrix.RData,

Chapter7and8_mktcapMatrix.RData

Description: The three R data sets contain returns data, industry classifications, and market capitalization data, respectively.

Frequency: monthly

Date range: 199412-201212

Note: The data are used in Chapter 8 to construct quantile portfolios.

Chapter 9

Chapter9_Japan_largeCap_return.csv,

Chapter9_Japan_largeCap_industryStat.csv

Description: Japan large cap returns and industry statistics

Frequency: monthly

Date range: 199212-201206, 199712-201206

Note: The data are used to analyze the Japanese economy and financial market.