

Financial Data Analytics using SAS¹

Table of contents

- Chapter 1: Introduction SAS
- Chapter 2: Data Input
- Chapter 3: Data Manipulation
- Chapter 4: SAS procedure: PROC XYZ
- Chapter 5: Data output
- Chapter 6: Simple macro language
- Chapter 7: Simple graphs and data visualization
- Chapter 8: Simple string manipulations
- Chapter 9: Open data and several financial databases
- Chapter 10: Dealing with big data
- Chapter 11: Text Analysis
- Chapter 12: Parsing SEC filings
- Chapter 13: SAS Matrix manipulation (IML)
- Chapter 14: Advanced SAS macros
- Chapter 15: MySQL, SAS, R, and Python

- Chapter 16: Various distribution
- Chapter 17: T-test, F-test, Normality test
- Chapter 18: Linear models, VaR (Value at Risk)
- Chapter 19: Time series, Dickey–Fuller test
- Chapter 20: Options and Futures
- Chapter 21: Monte Carlo Simulation
- Chapter 22: Portfolio theory
- Chapter 23: Credit Risk Analysis
- Chapter 24: Machine learning
- Chapter 25: Seminar papers and their replications
- Chapter 26: Term projects

¹ © Yuxing Yan, 12/11/2018, yany@canisius.edu , the first draft is expected before 9/1/2019.

Chapter 1: Introduction to SAS

- 1.1 Comparison between R, Python, Matlab and SAS
- 1.2 Data oriented
- 1.3 PC SAS vs. UNIX SAS
- 1.4 Running a PC SAS program
- 1.5 Choosing a meaningful variable name
- 1.6 A few examples
- 1.7 Two types of comments
- 1.8 How to debug?
- 1.9 SAS warning messages
- 1.10 Code alignment
- 1.11 Saving many small programs
- 1.12 SAS online documentation
- 1.13 SAS online sample code
- 1.14 A few SAS books for finance
- 1.15 One-line R code for the book
- 1.16 How to connect to a server?
- 1.17 Several UNIX commands
- 1.18 Vi: a UNIX text editor
- 1.19 Running a UNIX SAS program
- 1.20 Running UNIX commands from SAS
- 1.21 Free SAS University Edition
- 1.22 Running SAS via Anaconda

Exercises

Chapter 2: Data input

- 2.1 Inputting data from keyboard
- 2.2 SAS missing code
- 2.3 Paths for our input files
- 2.4 Inputting data from a CSV file
- 2.5 MISSEVER: missing data items
- 2.6 DSD: multiple commas
- 2.7 Informat: input format
- 2.8 Concept of library
- 2.9 Default library: WORK
- 2.10 SASHEL data sets
- 2.11 Inputting SAS data sets: LIBNAME
- 2.12 Inputting data with fixed widths
- 2.13 PROC CONTENTS
- 2.14 Retrieving a small data set first

- 2.15 DROP=, KEEP=
 - 2.16 SAS date variable
 - 2.17 %LET BEGDATE="01JAN2017"D
 - 2.18 Abbreviation of similar variables
 - 2.19 Random number generators
 - 2.20 PROC IMPORT
 - 2.21 Retrieving data from the internet
 - 2.22 Seeing to believe
 - 2.23 Inputting data from an Excel file
 - 2.24 SAS University Edition: inputting data from shared folders
 - 2.25 Reading from a binary file
 - 2.26 Getting zip files from the internet
 - 2.27 Reading data from a zip file
 - 2.28 Reading data from cloud-based data
- Exercises

Chapter 3: Financial statement analysis

- 3.1 What are the purposes of financial statement analysis
- 3.2 What is Balance-Sheets
- 3.3 What is income statements
- 3.4 What is cash-flow statements
- 3.5 Download financial statements from Yahoo!Finance, Google Finance
- 3.6 Ratio analysis
- 3.7 DuPont identity

Chapter 4: Risk and return

- 4.1 Definitions of returns
- 4.2 Arithmetic mean vs. geometric mean
- 4.3 Time-weighted vs. dollar-weighted
- 4.4 Definitions of risk
- 4.5 Total risk vs. market risk
- 4.6 Excel functions to calculate σ^2 and σ
- 4.7 Annualizing σ^2 and σ
- 4.8 Daily returns to weekly/monthly/annual ones
- 4.9 Tradeoff (I): Sharpe ratio
- 4.10 Tradeoff (II): Treynor ratio
- 4.11 Tradeoff (III): LPSD and Sortino ratio
- 4.12 Tradeoff (IV): a utility function
- 4.13 Concept of an indifference curve

Chapter 5: Interest rate, bond and stock evaluation

- 5.1 Introduction
- 5.2 2-step approach for effective rate conversion
- 5.3 Review: present value of perpetuity and present value of growing perpetuity
- 5.4 Review: present value of annuity
- 5.5 How to estimate YTM
- 5.6 How to price bond
- 5.7 Definition of duration
- 5.8 hedging by duration
- 5.9 pricing stock: one-period model
- 5.10 pricing stock: n-period model
- 5.11 Pricing stock: using multiples

Chapter 6: T-test, F-test, tests of equal-means and equal-variances

- 6.1 Introduction
- 6.2 Impact of size, sample vs. population, sample statistics
- 6.3 Review of three concepts
- 6.4 Two important values: 2 for a T-test and 5% for a p-value
- 6.5 T-test, critical value, decision rule
- 6.6 F-test for equal variance, decision rule and critical value
- 6.7 Chi square distribution, normality test, test whether following a special distribution
- 6.8 How to run a single factor linear regression?
- 6.9 Event study using Excel

Chapter 7: CAPM (a single-factor linear model)

- 7.1 Introduction
- 7.2 How to run a linear regression using Excel
- 7.3 download data from Yahoo!Finance, Google Finance
- 7.4 Download data by using `.getdata`
- 7.5 introduction to market indices, such as S&P500
- 7.6 value-weighted, equal-weighted, price-weighted indices
- 7.7 How to estimate market risk for IBM
- 7.8 How to adjusted beta
- 7.9 Beta adjustment, Dimson (1972)
- 7.10 beta adjustment: Sholes and William (1973)
- 7.11 Beta of a portfolio

Chapter 8: Multifactor models: FF3, FFC4, FF5

- 8.1 Fama-French 3-factor model
- 8.2 Prof. French's Data Library
- 8.3 Download data from French's Data Library
- 8.4 Alternative way to get data: `.getdata`
- 8.5 Running Fama-French 3-factor model
- 8.6 Running Fama-French-Carhart 4-factor model
- 8.7 Running Fama-French 5-factor model

Chapter 9: Various distributions (uniform, normal distribution)

- 9.1 Introduction
- 9.2 Density distribution vs. cumulative density distribution
- 9.3 standard normal vs. cumulative standard normal distribution
- 9.4 Normality tests
- 9.5 Uniform distribution
- 9.6 χ^2 distribution
- 9.7 Poisson distribution

Chapter 10: Black-Scholes-Merton options model

- 10.1 Introduction
- 10.2 Payoff and profit/loss functions
- 10.3 Hedge, speculation and arbitrage
- 10.4 Various trading strategies involving options
- 10.5 Density distribution vs. cumulative density distribution
- 10.6 Review: standard normal vs. cumulative standard normal distribution
- 10.7 An easy way to remember the logic behind Black-Scholes-Merton option model
- 10.8 Black-Scholes-Merton option model
- 10.9 VBA for call and put

- 10.10 Implied volatility
- 10.11 Option Greeks
- 10.12 Put-call parity

Chapter 11: Monte Carlo Simulation

- 11.1 Review of several basic concepts
- 11.2 Concept of seed
- 11.3 How to generate a set of random numbers using Excel
- 11.4 Simulating stock price
- 11.5 Simulating Black-Scholes-Merton call
- 11.6 Simulating Asian options
- 11.7 QQ plot
- 11.8 Review: Normality Tests
- 11.9 Generating 2 sets of correlated random sequences
- 11.10 Generating n sets of correlated random sequences

Chapter 12: VaR (Value at Risk)

- 12.1 Definition of VaR
- 12.2 Review of a normal distribution
- 12.3 Generating n-day returns
- 12.4 Estimating VaR based on the normality assumption
- 12.5 Conversion of σ between frequencies
- 12.6 Estimating VaR based on sorted returns
- 12.7 Modified VaR
- 12.8 Portfolio VaR

Chapter 13: Portfolio Theory

- 13.1 Review of basic concepts
- 13.2 Markowitz mean-variance efficiency
- 13.3 Converting monthly returns to annual ones
- 13.4 How to generate a return matrix
- 13.5 Estimating portfolio returns
- 13.6 Estimating portfolio risk for a 2-stock portfolio
- 13.7 Estimating variance-covariance/correlation matrices
- 13.8 Portfolio variance for an n-stock portfolio
- 13.9 Finding an optimal portfolio for 2 stocks
- 13.10 Constructing an efficient frontier with 2 stocks
- 13.11 Portfolio insurance
- 13.12 Portfolio performance measure
- 13.13 Simulating future portfolio distribution

Chapter 14: Liquidity measure and high-frequency data

- 14.1 Why liquidity matters and how to measure it
- 14.2 Aminud (2002) illiquidity measure
- 14.3 Is liquidity a risk factor?
- 14.4 Pastor and Stambaugh (2003) liquidity measure
- 14.5 Getting High-frequency data
- 14.6 Roll spread (1984)
- 14.7 Introduction to high-frequency data
- 14.8 CT (Consolidated Trade) data
- 14.9 CQ (Consolidated Quote) data
- 14.10 How to merge CT and CQ?
- 14.11 Who initiated trade? (Lee and Ready methodology, 1991)

Chapter 15: Credit Risk Analysis

- 15.1 Introduction
- 15.2 default spread
- 15.3 Credit rating
- 15.4 Rating migration matrix
- 15.5 Credit rating vs. default probability
- 15.6 Bond pricing with default probability
- 15.7 Hedging credit risk

Part II

Chapter 16: R basics

- 16.1 R installation
- 16.2 How to launch R
- 16.3 How to quit R
- 16.4 R is case sensitive
- 16.5 One line R code for this course
- 16.6 Three ways to assign a value
- 16.7 Choosing a meaningful variable name
- 16.8 `ls()` function to list variables and functions
- 16.9 `rm()` function to remove unnecessary variables/functions
- 16.10 using R as a calculator
- 16.11 Using up-arrow/down-arrow keys
- 16.12 Some commonly used functions
- 16.13 How to list all three-letter functions
- 16.14 How to find help about a specific function
- 16.15 How to write a one-line R function
- 16.16 How to add a comment

16.17 Using `source()` function

Chapter 17: Excel basics (II)

- 17.1 `average()`, `geomean()`, `min()` and `max()`
- 17.2 `power()` function and \wedge
- 17.3 string manipulation: `upper()`, `lower()`, `len()` and `substitute()`
- 17.4 `round()`, `roundup()` and `rounddown()`
- 17.5 `count()`, and `counta()`
- 17.6 `stdev()`, `stdev.s()`, `stdev.p()`, `var.s()`, and `var.p()`
- 17.7 `if()` function
- 17.8 `pv()` and `fv()`
- 17.9 Excel sign convention
- 17.10 `rate()`, `effect()`, and `nper()`
- 17.11 `year()`, `month()`, and `day()`
- 17.12 `npv()` and `IRR()` functions
- 17.13 `isodd()`, `iseven()` and the like
- 17.14 `index()` and `match()`
- 17.15 auto fill

Chapter 18: Sources of open data

- 18.1 Open data for economics
- 18.2 Open data for finance
- 18.3 Open data for accounting

Chapter 19: Utility or supporting functions

- 19.1 Menu for all utility functions, `.uu`
- 19.2 Menu for retrieving data, `.getdata`
- 19.3 Menu for in-class exercises, `.ice`
- 19.4 Menu for showing all functions, `.showFormula`
- 19.5 Menu for a free financial calculator, `.finca1`
- 19.6 Menu for mimic Excel `.mimicExcel`
- 19.7 Menu for data cases, `.dataCases`
- 19.8 Menu for term projects, `.termProjects`
- 19.9 Menu for interactive mode, `.interactive`
- 19.10 Menu for useful link, `.usefulLinks`
- 19.11 Menu for YouTube, `.youtube`
- 19.12 Menu for visual finance. `.vf`

Chapter 20: Issues with Excel

- 20.1 Excel sign convention

- 20.2 meaning of a blank cell
- 20.3 Excel NPV() function actually is a PV() function
- 20.4 inconsistency [stdev.s(), stdev()] vs. [covariance.s, covar()]
- 20.5 No formula for present (future) value for a growing annuity
- 20.6 No explanation when we have #NUM
- 20.7 no formula for converting effective rates

Chapter 21: vlookup and solver

- 21.1 if Excel solver is not available on the menu bar
- 21.2 what is the usage of solver
- 21.3 examples of using solver
- 21.4 Why should we care about vlookup?
- 21.5 The first example for using Excel vlookup()
- 21.6 vlookup() meaning of range_lookup
- 21.7 Return the left column using vlookup
- 21.8 vlookup() for text
- 21.9 vlookup() for text: current ratio

Chapter 22: Data input

- 22.1 copy and paste
- 22.2 input from a text file
- 22.3 input from a csv (comma separated value)
- 22.4 input data with a fixed width
- 22.5 text to columns
- 22.6 retrieve different data sets into the same excel file
- 22.7 retrieve data from web
- 22.8 generate a true date variable

Chapter 23: Data manipulation

- 23.1 If 'data analysis' is not available on your menu bar
- 23.2 sort data
- 23.3 Remove blank lines
- 23.4 add a date variable, year(), month(), day() functions
- 23.5 find out the last row of our data set
- 23.6 name our return matrix
- 23.7 log return vs. percentage return
- 23.8 Converting daily returns to weekly, 10-day, monthly, quarterly and annual ones
- 23.9 Generating T1, T2, T50000
- 23.10 Conditional formatting
- 23.11 Merge different data set by date, year and month

23.12 Randomly select 10 stocks from 200 stocks

Chapter 24: Data output

- 24.1 some types of output formats
- 24.2 why we care about the output formats?
- 24.3 Difference between ASCII text and Unicode text
- 24.4 Save as a file with an extension of (*.xlsx)
- 24.5 Save as an Macro-Enabled Workbook (*.xlsm)
- 24.6 Save as a csv file (*.csv)
- 24.7 Save as a text file (*.txt)
- 24.8 Copy and paste
- 24.9 Save as an Excel Binary Workbook (*.xlsb)
- 24.10 Save as Excel 97-2003 Workbook (*.xls)
- 24.11 Save as an XML Data format (*.xml)
- 24.12 Save as a pdf file (*.pdf)
- 24.13 Save as an Excel Add-in (*.xlam)
- 24.14 Save as an XPS Document (*.xps)
- 24.15 Save as a 97-2003 Add-in (*.xla)
- 24.16 Save as a Unicode Text (*.txt)
- 24.17 csv file when we have comma
- 24.18 output an image using snipper
- 24.19 output an image using Ctrl-Print
- 24.20 use paint software after hitting Ctrl-Print

Chapter 25: Simple graph

- 25.1 A simple x, y graph
- 25.2 Column
- 25.3 Line
- 25.4 Pie and Bar
- 25.5 histogram
- 25.6 Area
- 25.7 (x,y) scatter
- 25.8 stock chart
- 25.9 Present your data in a surface chart
- 25.10 Doughnut
- 25.11 Present your data in a bubble chart
- 25.12 Present your data in a radar chart
- 25.13 Other charts
- 25.14 Win/loss
- 25.15 one graph with two sets of lines (x,y) and (a,b)

- 25.16 Macro 4 pie
- 25.17 Macro 4 x,y
- 25.18 Macro 4 more.....
- 25.19 Two ways to output our images

Chapter 26: Matrix manipulation

- 26.1 What is a matrix?
- 26.2 Row, column vs. matrix
- 26.3 Dimension requirement for matrix manipulation
- 26.4 hit ctrl-shift-Enter 3 keys simultaneously
- 26.5 Convert a row (row)to a column (column)
- 26.6 Transpose an $n*m$ matrix to an $m*n$ matrix
- 26.7 for a large data set, how to find its last row?
- 26.8 for a large data set, how to highlight it?
- 26.9 Explanation of matrix multiplication
- 26.10 Excel `mmult()` function
- 26.11 Using excel `mmult()` function for the 1st time
- 26.12 one row times a column will be one number
- 26.13 one column ($n * 1$) times a row ($1 * m$) will be a matrix ($n * m$)
- 26.14 one row times the transpose of itself
- 26.15 one row times the transpose of another row
- 26.16 one column times the transpose of itself
- 26.17 one column times the transpose of another column
- 26.18 If 'Developer' is not available on your menu bar
- 26.19 VBA for estimating variance-covariance matrix
- 26.20 VBA for estimating correlation matrix

Chapter 27: Simple Macro: record Macros

- 27.1 If “Developer” is not available
- 27.2 Recording our first macro
- 27.3 How to view/edit our recorded Macro
- 27.4 Copy and paste others' Macros
- 27.5 Three ways to run our Macro
- 27.6 Linking a button to a Macro
- 27.8 Assigning a short-cut key an existing macro
- 27.9 Ctrl-a, Ctrl-b, Ctrl-c and Ctrl-d
- 27.10 Generating a message box
- 27.11 Calling a Macro(s) from another Macro
- 27.12 A set of one-line codes
- 27.13 A list of 2 dozen macros

Chapter 28: Simple VBA: copy-and-paste

- 28.1 The simplest VBA
- 28.2 Data Analysis is not available
- 28.3 correct extension: x.xlsm
- 28.4 How to copy-and-paste others' VBA
- 28.5 A list of VBAs

Chapter 29: Pivotal table

- 29.1 What is pivotal table
- 29.2 The simplest example
- 29.3 Converting daily returns into monthly and annual ones
- 29.4 More examples

Chapter 30: Term projects

- 30.1 How important of a term project
- 30.2 A list of 40 term projects
- 30.3 Which party, Republican or Democratic Party, could manage economy better
- 30.4 What is the market risk for IBM, C and WMT
- 30.5 What is the VaR for three stocks based on two methods?
- 30.6 Which model is the best, CAPM, FF3, FFC4 or FF4?
- 30.7 Benford law and accounting fraud
- 30.8 Test of January effect and weekday effect
- 30.9 Event study using Excel
- 30.10 Replicating a slot machine by using simulation