

Financial Data Analytics using SAS¹

9/29/2019

Table of contents

- Chapter 1: Introduction to SAS
- Chapter 2: UNIX SAS
- Chapter 3: Free SAS (SAS University Edition)
- Chapter 4: Open Data
- Chapter 5: Data Input
- Chapter 6: Data Manipulation
- Chapter 7: SAS Procedures (SAS PROC)
- Chapter 8: Data Output
- Chapter 9: Simple Macro Language
- Chapter 10: Simple Graphs and Data Visualization
- Chapter 11: Simple String Manipulations
- Chapter 12: Dealing with Big Data
- Chapter 13: SAS Matrix Manipulation (IML) and Advanced SAS Macros
- Chapter 14: Text Analysis
- Chapter 15: MySQL, SAS, R, and Python

- Chapter 16: Parsing SEC filings
- Chapter 17: Introduction to Several Financial Databases (CRSP, Compustat and TAQ)
- Chapter 18: How to replicate a published paper?
- Chapter 19: Replicating seminal papers (a): Factor models
- Chapter 20: Replicating seminal papers (b): Liquidity (illiquidity)
- Chapter 21: Replicating seminal papers (c): Accruals
- Chapter 22: Replicating seminal papers (d): Momentum strategy
- Chapter 23: Replicating seminal papers (e): Trading direction
- Chapter 24: Replicating seminal papers (f): Probability of informed trading
- Chapter 25: Replicating seminal papers (g): Transaction costs (spread)
- Chapter 26: Replicating seminal papers (h): Earnings surprises
- Chapter 27: Replicating seminal papers (i): Mutual funds
- Chapter 28: Replicating seminal papers (j): Methodology
- Chapter 29: Replicating seminal papers (k): Dynamic Conditional Correlation
- Chapter 30: Replicating seminal papers (l): Adverse selection
- Chapter 31: Replicating seminal papers (m): Beta and GMM
- Chapter 32: Replicating seminal papers (n): CEO performance: narcissism index
- Chapter 33: Replicating seminal papers (p): Term Structure
- Chapter 34: Replicating seminal papers (q): Credit Risk
- Chapter 35: Replicating seminal papers (r): Impact of delisting returns
- Chapter 36: Replicating seminal papers (s): New ways to construct market indices
- Chapter 37: Replicating seminal papers (t): R² across countries
- Chapter 38: Replicating seminal papers (u): Benford Law (the law of the 1st digit)
- Chapter 39: Replication seminal papers (w): IPO prospects
- Chapter 40: Replication seminal papers (x) : Inside trading
- Chapter 41: Replication seminal papers (w): How to define an industry?
- Chapter 42: Replication seminal papers (x): Volatility spillover
- Chapter 43: Replicating seminal papers (o): Investor sentiment index
- Chapter 44: 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
- Exercises

Chapter 2: UNIX SAS

- 2.1 Advantages and disadvantages of running UNIX SAS
 - 2.2 How to connect to a server?
 - 2.3 Several UNIX commands
 - 2.4 Vi: a UNIX text editor
 - 2.5 Running a UNIX SAS program
 - 2.6 nohup command
 - 2.7 Running UNIX commands from SAS
 - 2.8 Free SAS University Edition
 - 2.9 Running SAS via Anaconda
 - 2.10 Server-less
 - 2.11 SAS installation via Anaconda
- Exercises

Chapter 3: Free SAS (SAS University Edition)

- 3.1 Introduction to SAS University Edition
 - 3.2 Virtual box
 - 3.3 Running SAS via SAS University edition
 - 3.5 Help for SAS University edition
 - 3.7 Several examples
- Exercises

Chapter 4: Open Data

- 4.1 Concept of OSF
- 4.2 Sources of Open Data
- 4.3 Introduction to Yahoo!Finance
- 4.4 Introduction to Google Finance
- 4.5 French's Data Library
- 4.6 SEC Filings
- 4.7 NYSE High-Frequency Data
- 4.8 UCI Machine Learning Repository
- 4.9 Sources of Large Data Sets
- 4.10 Kaggle.com
- 4.11 Efficient Way to Get Data

Exercises

Chapter 5: Data Input

- 5.1 Inputting data from keyboard
- 5.2 SAS missing code
- 5.3 Paths for our input files
- 5.4 Inputting data from a CSV file
- 5.5 MISSOVER: missing data items
- 5.6 DSD: multiple commas
- 5.7 Informat: input format
- 5.8 Concept of library
- 5.9 Default library: WORK
- 5.10 SASHEL data sets
- 5.11 Inputting SAS data sets: LIBNAME
- 5.12 Inputting data with fixed widths
- 5.13 PROC CONTENTS
- 5.14 Retrieving a small data set first
- 5.15 DROP=, KEEP=
- 5.16 SAS date variable
- 5.17 %LET BEGDATE='01JAN2017'D
- 5.18 Abbreviation of similar variables
- 5.19 Random number generators
- 5.20 PROC IMPORT
- 5.21 Retrieving data from the internet
- 5.22 Seeing to believe
- 5.23 Inputting data from an Excel file
- 5.24 SAS University Edition: inputting data from shared folders
- 5.25 Reading from a binary file
- 5.26 Getting zip files from the internet

- 5.27 Reading data from a zip file
- 5.28 Reading data from cloud-based data
- Exercises

Chapter 6: Data Manipulation

- 6.1 PROC SORT
- 6.2 SAS Counting Variable
- 6.3 Number of Observations
- 6.4 Unique Values
- 6.5 FIRST.VAR and LAST.VAR
- 6.6 Generating Lags
- 6.7 PROC MEANS
- 6.8 Daily Returns to Monthly Returns
- 6.9 Estimating EW or VW Portfolios
- 6.10 PROC CONTENTS
- 6.11 Subset
- 6.12 Conversions between Character and Numerical
- 6.13 Merging Data Sets
- 6.14 Matching
- 6.15 Defining an Array
- 6.16 PROC FREQ
- 6.17 Retain Functionality
- 6.18 _all_
- 6.19 PROC DATASETS
- 6.20 Forward Values: PROC EXPAND
- 6.21 Loop in a Data Step
- 6.22 Loop Defined by a Macro
- 6.23 Double and Single Quotation Marks
- 6.24 Renaming Variables
- 6.25 Deleting Unnecessary Data Sets
- 6.26 Last Data Set and Stopping a Running Program
- 6.27 Interval Functions INTNX and INTCK
- 6.28 Label
- 6.29 Example of Using an Array
- 6.30 PROC SQL
- Exercises

Chapter 7: SAS Procedures (SAS PROC)

- 7.1 Introduction
- 7.2 SAS Procedures by Name
- 7.3 PROC PRINT
- 7.4 PROC CONTENT

- 7.5 PROC MEANS
- 7.6 PROC SORT
- 7.7 PROC APPEND
- 7.8 PROC CORR
- 7.9 PROC TRANSPOSE
- 7.10 Critical values for T-test, F-test
- 7.11 PROC TTEST
- 7.12 PROC REG: run a linear regression
- 7.13 PROC MODEL
- 7.14 PROC GLM
- 7.15 PROC GENMOD
- 7.16 PROC LOGITSTIC
- 7.17 Optimization: PROC OPTMODEL
- 7.18 PROC TABULATE
- 7.19 PROC SUMMARY
- 7.20 PROC RANK
- 7.21 More Complex Examples
- 7.22 PROC SQL
- 7.23 Concept of Pipe
- 7.24 Lists of All SAS Procedures

Exercises

Chapter 8: Data Output

Chapter 9: Simple Macro Language

Chapter 10: Simple Graphs and Data Visualization

Chapter 11: Simple String Manipulations

Chapter 12: Dealing with Big Data

Chapter 13: SAS Matrix Manipulation (IML) and Advanced SAS Macros

Chapter 14: Text Analysis

Chapter 15: MySQL, SAS, R, and Python

Chapter 16: Parsing SEC filings

Chapter 17: Introduction to Several Financial Databases (CRSP, Compustat and TAQ)

Chapter 18: How to replicate a published paper?

Chapter 19: Replicating seminal papers (a): Factor models

Chapter 20: Replicating seminal papers (b): Liquidity (illiquidity)

Chapter 21: Replicating seminal papers (c): Accruals

Chapter 22: Replicating seminal papers (d): Momentum strategy

Chapter 23: Replicating seminal papers (e): Trading direction

Chapter 24: Replicating seminal papers (f): Probability of informed trading

Chapter 25: Replicating seminal papers (g): Transaction costs (spread)

Chapter 26: Replicating seminal papers (h): Earnings surprises

Chapter 27: Replicating seminal papers (i): Mutual funds

Chapter 28: Replicating seminal papers (j): Methodology

Chapter 29: Replicating seminal papers (k): Dynamic Conditional Correlation

Chapter 30: Replicating seminal papers (l): Adverse selection

Chapter 31: Replicating seminal papers (m): Beta and GMM

Chapter 32: Replicating seminal papers (n): CEO performance: narcissism index

Chapter 33: Replicating seminal papers (p): Term Structure

Chapter 34: Replicating seminal papers (q): Credit Risk

Chapter 35: Replicating seminal papers (r): Impact of delisting returns
Chapter 36: Replicating seminal papers (s): New ways to construct market indices
Chapter 37: Replicating seminal papers (t): R² across countries
Chapter 38: Replicating seminal papers (u): Benford Law (the law of the 1st digit)
Chapter 39: Replication seminal papers (w): IPO prospects
Chapter 40: Replication seminal papers (x) : Inside trading
Chapter 41: Replication seminal papers (w): How to define an industry?
Chapter 42: Replication seminal papers (x): Volatility spillover
Chapter 43: Replicating seminal papers (o): Investor sentiment index
Chapter 44: Term projects