CRSP Stocks and SPGMI Factors Data in FactorAnalytics

January 28, 2022

Doug Martin

1 The stocksCRSP Data (310 securities)

We have 300 of the CRSP stocks in FactorAnalytics, and have 10 "spares" that are discussed further in Section 1.1.

In the Table that follows: The CRSP Spreadsheet column contains the abbreviated .xlsx spreadsheet data file names as obtained from CRSP; the Spreadsheet Names column contains the data component names used in those .xlsx spreadsheets; the CRSP Monthly and CRSP Daily columns contain the CRSP names for the monthly and daily data, respectively. Those names were obtained from the definitions in the "CRSP data descriptions guide 02 02 21.pdf" and "CRSP U.S. treasury database guide 01 26 21.pdf" documents.

The abbreviated entries in CRSP Spreadsheet column of the Table refers to the following specific original files sent to me by CRSP:

The "stocksCRSP310" entries refer to one of the following files:

- stocksCRSP310 - 202010 - output monthly.xlsx

- stocksCRSP310 - 202010 - output weekly.xlsx (as as constructed for us by CRSP using a CRSP Utility)

- stocksCRSP310 - 202010 - daily output1.xlsx

- stocksCRSP310 - 202010 - daily output2.xlsx

The "risk-free" entries refer to one of the following files:

- risk-free montly.csv (.xlsx files not available?)

- risk-free weekly.csv (.xlsx files not available? Constructed for us by CRSP using a CRSP Utility)

- risk-free 202010 output.xlsx (daily data)

Note 1: In each of the above files the IDNO's are 2000061, 200062, 2000003, with names RISK FREE 4-WEEK, RISK FREE 13-WEEK, 1 YEAR BONDS. See "CRSP U.S. treasury database guide 01 26 21.pdf".

The "market index" entries refer to one of the following files:

- market index 202010 - monthly.xlsx

- market index 202010 - weekly.xlsx (as constructed for us by CRSP using a CRSP Utility)

- market index 202010 - daily.xlsx

Note 2: This maket index is the CRSP NYSE/NYSE MKT/NASDAQ/Arca Value-Weighted Market Index, which is abbreviated to CRSP Market Index in the Table that follows. Its IDNO is 1000200, with dailly SetId 460 and monthly SetId 440. See "CRSP data descriptions guide 02 02 21.pdf".

CRSP Spreadsheet	Spreadsheet Name	CRSP Name Monthly	CRSP Name Daily
stocksCRSP310	DATE	mcaldt	caldt
stocksCRSP310	Last C... Name	mcomnam	comnam
stocksCRSP310	Tickerl	tickerl	tickerl
stocksCRSP310	Ticker	mticker	ticker
stocksCRSP310	Ret	mret	ret
stocksCRSP310	Retx	mretx	retx
stocksCRSP310	Prc	mprc	prc
stocksCRSP310	Adjprc	madjprc	adjprc
risk-free	2000061	RISK FREE 4-WEEK 2000061	RISK FREE 4-WEEK 2000061
risk-free	2000062	RISK FREE 13-WEEK 2000062	RISK FREE 13-WEEK 2000062
risk-free	2000003	1 YEAR BONDS 2000003	1 YEAR BONDS 2000003
market index	Tret	CRSP Market Index 1000200-440	CRSP Market Index 1000200-460

Note 3: For the 1 YEAR BONDS (TREASNOX 2000003), the annualized TDYTM=100*TDYLD*365.0 was transformed back to TDYLD (daily yield) in decimal form for me by CRSP.

Note 4: The weekly stocksCRSP310 returns were computed from the daily returns by CRSP, using a CRSP

utiltity, as a one-off. So I suggest that we use the names wcaldt, wcomnam, etc. for this weekly data.

Note 5: I am guessing that the monthly and weekly risk-free returns were both computed from the daily risk-free rates, hence just having .csv files for those (no .xlsx files for monthly and weekly risk-free returns).

Mapping from CRSP Query Names to NiceNames

CRSP Names Monthly	NiceNames Monthly
mcaldt	Date
mcomnam	Company
tickerl	Ticker
mticker	TickerHistory
mret	Return
mretx	RetExDiv
mprc	Price
madjprc	PrcSplitAdj
RISK FREE 4-WEEK	Ret4WkBill
RISK FREE 13-WEEK	Ret13WkBill
1 YEAR BONDS	Ret1YrBond
CRSP Market Index	MktRetCRSP

I will write the function for mapping from CRSP names to NiceNames.

1.1   GICS Codes, Sectors and CapGroups

The GICS codes are the 6 digits codes that were supplied by SPGMI (GICS is a joint project of SPGMI and MSCI), and describe fully in the document "GICS Map Book.pdf".

The Sectors are determined by the first two digits of the GICS code that appears in the Code column of the two Tables below. As I explain below, in the final version of FactorAnalytics CRSP and SPGMI data for the initial CRAN release, we will not have any stocks in the Financials, Utilities and Real Esetate Sectors, so we only use 8 of the 11 Sectors. I want to have FA Sector names that are short enough to print reasonably nicely as table and plot labels, and the two possible choices in the Table came to mind in discussion with Tom. At the moment, I prefer Choice 2.[1]

---

[1]I note that this is a different situation than the need to have CRSP and SPGMI original data item names. First of all the short names will be reasonably clear to most people, and furthermore the final table with two columns will be in the Vignette and man page, so anyone in doubt can find the full names easily from our documentation (as well as from the readily available "GICS Map Book.pdf" document)

Code	GICS Sector	FA Sector Choice 1	FA Sector Choice 2
10	Energy	Energy	Energy
15	Materials	Materials	Materials
20	Industrials	Industrials	Industrials
25	Consumer Discretionary	ConsumerDisc	ConsDisc
30	Consumer Stables	ConsumerStap	ConsStaples
35	Health Care	Healthcare	Healthcare
45	Information Technology	InfoTech	InfoTech
50	Communication Services	CommServices	CommSvcs

Code	Unused GICS Sectors
40	Financials
55	Utilities
60	Real Estate

Note 5: We chose the 310 stocks to have none in the Financials Sector. Then it turned out that we have only 3 in the Utilities Sector and 1 in the Real Estate Sector, the latter of which breaks risk decomposition code. So we will delete the one stock in the Real Estate Sector, and delete the 3 in the Utilities Sector, and replace these 4 stocks that I will designate 4 from the surplus set of 10.

2 The factorsSPGMI Data

The following names of the factors in the data set `factorsSPGMI` are the official SPGMI names of the factors (also know as "scores" and "alpha factors" internally at SPGMI):

- AnnVol12M

- Beta60M

- BP

- EP

- LogMktCap

- PM12M1M

- AccrualRatioCF

- AstAdjChg1YOCF

- CFROIC

- Chg1YAstTo

- EBITDAEV

- FCFP

- PM1M

- SEV

The descriptions of the factors are provided in the man page for the `factorsSPGMI` data set.

3 Loading, Merging and Selecting Subsets

*** To be added for Vignette version of this document, with original file names material deleted ***

4 Concluding Comments

xx