CRSP Stocks and SPGMI Factors Data in FactorAnalytics

February 23, 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". However, the monthly versus the daily versions of these data sets are identified at CRSP only by the separated file names, namely: the file names TFZ_DLY_RF2.* andTFZ_MTH_RF2 for the 4 week and 13 233k risk-free rates, and TFZ_DLY_FT.* and TFZ_MTH_FT.* for the 1 Year Bond. So the CRSP data user who wants to use both daily and monthly versions of the risk-free rates and 1 Year Bonds needs to get those separate files. **For clarity in the CRSP Name Monthly and CRSP Name Daily columns of the Table below, I have added to the purely numeric identifiers the prefixes m4wk and 4wk for the monthly and daily 4 week Bills, the prefixes m13wk and 13wk for the monthly and daily 13 week Bills, and m1yr and 1yr for the monthly and daily 1 year Bonds.**

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". **For clairity in the last row of the Table below, I have added the prefixes mmkt and mkt for the monthly and daily versions of those CRSP purely numeric identifiers.**

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	m4wk2000061	4wk2000061

risk-free	2000062	m13wk2000062	13wk2000062
risk-free	2000003	m1yr2000003	1yr2000003
market index	Tret	mmkt1000200-440	mkt1000200-460

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. when we add a column above 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). But I am not sure, and have queried the CRSP support staff to see if they can confirm the monthly risk-free returns in our .csv file, and if confirmed I will make a .xlsx version for the sake of consistency.

Mapping from CRSP Query Names to NiceNames

CRSP Names Monthly	CRSP Names Daily	NiceNames Monthly	NiceNames Weekly	NiceNames Daily
mcaldt	caldt	Date	Date	Date
mcomnam	comnam	Company	Company	Company
tickerl	tickerl	TickerL	TickerL	TickerL
mticker	ticker	Ticker	Ticker	Ticker
mret	ret	Return	ReturnW	ReturnD
mretx	retx	RetExDiv	RetExDivW	RetExDivD
mprc	prc	Price	PriceW	PriceD
madjprc	adjprc	PrcSplitAdj	PrcSplitAdjW	PrcSplitAdjD
m4wk2000061	4wk2000061	Ret4WkBill	Ret4WkBillW	Ret4WkBillD
m13wk2000062	13wk2000062	Ret13WkBill	Ret13WkBillW	Ret13WkBillD
m1yr2000003	1yr2000003	Ret1YrBond	Ret1YrBondW	Ret1YrBondD
mmkt1000200-440	mkt1000200-440	MktRetCRSP	MktRetCRSPW	MktRetCRSPD

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 chose the names in the FA Sector column of the Table below so that they are short enough to print reasonably nicely as table and plot labels, and still be highly reflect the full Sector names.

Code	GICS Sector	FA Sector
10	Energy	Energy
15	Materials	Materials
20	Industrials	Industrials
25	Consumer Discretionary	ConsumDisc
30	Consumer Stables	ConsumStap
35	Health Care	Healthcare
45	Information Technology	InfoTech
50	Communication Services	ComServices

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 Excel file listing material deleted ***

4 Concluding Comments

xx