

## Data description.

There are three datasets associated with this project:

1. *Stock\_data\_part1.xlsx*
  - Contains **daily** data on US stocks and ETFs, over the period 20/08/2019 - 20/08/2020
  - Variables:

Variable	Definition	Units of measurement
PERMNO	PERMNO is a unique permanent identifier assigned by CRSP to all companies with issues on a CRSP file. This number is permanent for all securities issued by this company regardless of name changes. The mnemonics PERMNO and COMPNO are interchangeable. If the value in this field is less than 20,000, CRSP has used the Nasdaq-assigned Company Number, for when the company had an issue trading on The Nasdaq Stock Market (SM). A value over 20,000 indicates that the number was assigned by CRSP.	
Date	date the variable is observed	
TICKER	ticker (security identifier)	
COMNAM	Company Name	
PERMCO	CRSP Permanent Company Number	
BIDLO	Bid or Low Price. Daily: Bid or Low Price is the lowest trading price during the day, or the closing bid price on days when the closing price is not available. The field is set to zero if no Bid or Low Price is available.	USD
ASKHI	Ask or High Price is the highest trading price during the day, or the closing ask price on days when the closing price is not available. The field is set to zero if no Ask or High Price is available. If the Price or Bid/Ask Average is negative, this field contains the closing ask. If positive, the field contains the highest trade.	USD
PRC	<p>Prc is the closing price or the negative bid/ask average for a trading day. If the closing price is not available on any given trading day, the number in the price field has a negative sign to indicate that it is a bid/ask average and not an actual closing price. Please note that in this field the negative sign is a symbol and that the value of the bid/ask average is not negative.</p> <p>If neither closing price nor bid/ask average is available on a date, prc is set to zero. In a monthly database, prc is the price on the last trading date of the month. The price series begins the first month-end after the security begins trading and ends the last complete month of trading.</p>	USD
VOL	VOL is the total number of shares of a stock sold on day I. It is expressed in units of one share, for daily data, and on hundred shares for monthly data. Our data source for NYSE/AMEX reports the number rounded to the nearest hundred. For example, 12,345 shares traded will be reported on the Nasdaq Stock Exchange as 12,345 and on the NYSE or AMEX exchanges as 12,300. Volume is set to -99 if the value is missing. A volume of zero usually indicates that there were no trades during the time period and is usually paired with bid/ask quotes in price fields.	Number of shares (for daily data)

RET	<p>A return is the change in the total value of an investment in a common stock over some period of time per dollar of initial investment. RET(I) is the return for a sale on day I. It is based on a purchase on the most recent time previous to I when the security had a valid price. Usually, this time is I - 1.</p> <p>Missing Return Codes</p> <p>RET(t) Reason For Missing Return</p> <p>-66.0 more than 10 periods between time t and the time of the preceding price t?</p> <p>-77.0 not trading on the current exchange at time t</p> <p>-88.0 no return, array index t not within range of BEGRET and ENDRET</p> <p>-99.0 missing return due to missing price at time t</p>	fraction
BID	Bid is available both daily and monthly for all securities on the three major exchanges: NYSE, AMEX and NASDAQ.	USD
ASK	Ask is available both daily and monthly for all securities on the three major exchanges: NYSE, AMEX and NASDAQ.	USD
SHROUT	SHROUT is the number of publicly held shares, recorded in thousands	thousands
OPENPRC	Daily open prices are available for securities traded on NYSE, AMEX, and NASDAQ exchanges beginning June 15, 1992. They represent the first trade after market opens. For NYSE, additional daily open prices are available between December 1925 and June 1962.	USD
NUMTRD	Daily: Number of Trades, Nasdaq contains the number of trades made on the Nasdaq Stock Market each date for a security. Trades on all exchanges are connected to Nasdaq's composite pricing network and all paper trades are included in the count. If the number of trades is unavailable, the field is set to 99.	number
sprtrn	SPRTRN is the return on the Standard & Poor's Composite Index	fraction

## 2. Stock\_data\_part2.xlsx

- Contains **monthly** data on valuation and profitability ratios for US stocks and ETFs, over the period January 2010 – December 2022.
- Variables:

Variable	Definition	Units of measurement
public_date	Date the data was released	
CAPEI	Shillers Cyclically Adjusted P/E Ratio, Multiple of Market Value of Equity to 5-year moving average of Net Income	number
Bm	Book / Market ratio, Book Value of Equity as a fraction of Market Value of Equity	number
Evm	Enterprise value multiple, Multiple of Enterprise Value to EBITDA	number
Pe_exi	Price-to-Earnings, excl. Extraordinary Items (diluted)	number
dpr	Dividend payout ratio, Dividends as a fraction of Income Before Extraordinary Items	number
npm	Net Income as a fraction of Sales	number
roa	Return on assets, Operating Income Before Depreciation as a fraction of average Total Assets based on most recent two periods	number

roe	Return on equity, Net Income as a fraction of average Book Equity based on most recent two periods, where Book Equity is defined as the sum of Total Parent Stockholders' Equity and Deferred Taxes and Investment Tax Credit	number
roce	Return on capital employed, Earnings Before Interest and Taxes as a fraction of average Capital Employed based on most recent two periods, where Capital Employed is the sum of Debt in Long-term and Current Liabilities and Common/Ordinary Equity	number
ptb	Price to book, Multiple of Market Value of Equity to Book Value of Equity	number
divyield	Dividend Yield, indicated Dividend Rate as a fraction of Price	number
ticker	Ticker (security identifier)	

3. *Stock\_data\_part3.xlsx*

- Contains **monthly** data on US stocks and ETFs, over the period January 2010 – December 2022.
- Variables:

Variable	Definition	Units of measurement																				
PERMNO	PERMCO is a unique permanent identifier assigned by CRSP to all companies with issues on a CRSP file. This number is permanent for all securities issued by this company regardless of name changes. The mnemonics PERMCO and COMPNO are interchangeable. If the value in this field is less than 20,000, CRSP has used the Nasdaq-assigned Company Number, for when the company had an issue trading on The Nasdaq Stock Market (SM). A value over 20,000 indicates that the number was assigned by CRSP.																					
Date	date the variable is observed																					
SHRCD	<p>SHRCD is a two-digit code describing the type of shares traded. The first digit describes the type of security traded.</p> <p>First Digit - Share Code - Security Type</p> <table><thead><tr><th>Code</th><th>Definition</th></tr></thead><tbody><tr><td>1</td><td>Ordinary Common Shares</td></tr><tr><td>2</td><td>Certificates</td></tr><tr><td>3</td><td>ADRs (American Depositary Receipts)</td></tr><tr><td>4</td><td>SBIs (Shares of Beneficial Interest)</td></tr><tr><td>7</td><td>Units (Depository Units, Units of Beneficial Interest, Units of Limited Partnership Interest, Depository Receipts, etc.)</td></tr></tbody></table> <p>Note: "Units" (code 7) does not represent combinations of common stock and anything else, such as warrants. The second digit gives more detailed information about the type of security traded.</p> <p>Second Digit - Share Code - Security Type</p> <table><thead><tr><th>Code</th><th>Definition</th></tr></thead><tbody><tr><td>0</td><td>Securities which have not been further defined.</td></tr><tr><td>1</td><td>Securities which need not be further defined.</td></tr><tr><td>2</td><td>Companies incorporated outside the US</td></tr></tbody></table>	Code	Definition	1	Ordinary Common Shares	2	Certificates	3	ADRs (American Depositary Receipts)	4	SBIs (Shares of Beneficial Interest)	7	Units (Depository Units, Units of Beneficial Interest, Units of Limited Partnership Interest, Depository Receipts, etc.)	Code	Definition	0	Securities which have not been further defined.	1	Securities which need not be further defined.	2	Companies incorporated outside the US	
Code	Definition																					
1	Ordinary Common Shares																					
2	Certificates																					
3	ADRs (American Depositary Receipts)																					
4	SBIs (Shares of Beneficial Interest)																					
7	Units (Depository Units, Units of Beneficial Interest, Units of Limited Partnership Interest, Depository Receipts, etc.)																					
Code	Definition																					
0	Securities which have not been further defined.																					
1	Securities which need not be further defined.																					
2	Companies incorporated outside the US																					

	3 Americus Trust Components (Primes and Scores). 4 Closed-end funds. 5 Closed-end fund companies incorporated outside the US 8 REIT's (Real Estate Investment Trusts).  For example: A SHRCD of 14 would represent ordinary common shares of a closed-end fund.	
TICKER	ticker (security identifier)	
PERMCO	CRSP Permanent Company Number	
CUSIP	CUSIP is the latest eight-character CUSIP identifier for the security through the end of the file.	
HSICIG	Header Standard Industrial Classification Industry Group. The first three digits of HSICCD-- this refers to an industry group.	
PRC	Prc is the closing price or the negative bid/ask average for a trading day. If the closing price is not available on any given trading day, the number in the price field has a negative sign to indicate that it is a bid/ask average and not an actual closing price. Please note that in this field the negative sign is a symbol and that the value of the bid/ask average is not negative.  If neither closing price nor bid/ask average is available on a date, prc is set to zero. In a monthly database, prc is the price on the last trading date of the month. The price series begins the first month-end after the security begins trading and ends the last complete month of trading.	number
VOL	In monthly files, VOL is the sum of the trading volumes during that month. It is expressed in hundreds of shares in monthly files.	hundreds
RET	Holding period return. A return is the change in the total value of an investment in a common stock over some period of time per dollar of initial investment. RET(I) is the return for a sale on day I. It is based on a purchase on the most recent time previous to I when the security had a valid price. Usually, this time is I – 1.  Missing Return Codes  RET(t) Reason For Missing Return -66.0 more than 10 periods between time t and the time of the preceding price t? -77.0 not trading on the current exchange at time t -88.0 no return, array index t not within range of BEGRET and ENDRET -99.0 missing return due to missing price at time t	fraction
SHROUT	SHROUT is the number of publicly held shares, recorded in thousands	thousands
SPREAD	Monthly: Spread Between Bid and Ask is the difference between the closing bid and ask quotes for a security. It's available only when Ask or High Price and Bid or Low Price are available and Closing Price or Bid/Ask Average is a bid/ask average.  If Closing Price of Bid/Ask Average is zero and Spread between Bid and Ask is negative, the spread represents a Bid or Low Price.	USD

	<p>If Closing Price or Bid/Ask Average is zero and Spread between Bid and Ask is positive, Spread Between Bid and Ask represents an Ask or High Price.</p> <p>It is set to zero if unavailable.</p>	
--	---	--