

Problem Set 1: Market Portfolio

Professor Bernard Herskovic
bernardherskovic@anderson.ucla.edu

Due on **Sunday April 13**. This is an **individual assignment**, but you can discuss it with your classmates. If you discuss with other classmates, indicate their names in your write-up. Please submit Python code (.py file only) as well as a separate write-up. Explain the procedure and your answers clearly in the write-up (such that someone unfamiliar with the problem could solve it). Code must be formatted as instructed in order to receive a grade. Use Bruin Learn to submit your answers.

You should submit two files:

- .py **PS1_YourStudentID** (for example, PS1_012345678.py), with **all** code used in answering the questions written below
- .pdf **PS1_YourStudentID** (for example, PS1_012345678.pdf), with discussion on how you answered the questions written below, as well as responses to any particular questions asked

1. Construct the value-weighted market return using CRSP data,¹ replicating the market return time series available in Kenneth French website.² Also calculate the equal-weighted market return, and the lagged total market capitalization. Your output should be from January 1926 to December 2024, at a monthly frequency.

- Suggested function: **PS1_Q1**
 - Input: dataframe **CRSP_Stocks**

Variable Name	Variable type
PERMNO	integer
date	datetime
SHRCD	integer
EXCHCD	integer
RET	float
DLRET	float
PRC	float
SHROUT	integer

- This should be the data as pulled from WRDS, with one exception. Format the date column as a datetime. This should be the full dataset available on WRDS; do not pre-filter by SHRCD, EXCHCD, or date.

¹Available at WRDS: <https://wrds-www.wharton.upenn.edu/>. You will have to go to CRSP > Annual Update > Stock / Security Files > CRSP Monthly Stock. Download the full dataset (do not limit by SHRCD, EXCHCD, or date).

²http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

- Output: dataframe with each row corresponding to a unique year and month, with columns

Variable Name	Variable type	Variable description
Year	Integer	Year
Month	Integer	Month
Stock_lag_MV	float	Total market value the previous month (in millions)
Stock_Ew_Ret	float	Equal-weighted returns
Stock_Vw_Ret	float	Value-weighted returns

- Note: Returns should be formatted in decimal proportion (not percent).

- Hints:

- Do not forget to include delisting returns (DLRET).
- Use cum-dividend returns (RET and DLRET).
- Do not forget to restrict your sample by exchange type (EXCHCD) and share codes (SHRCD).
- Construct market equity by multiplying prices (PRC) by number of shares outstanding (SHROUT).
- Check the CRSP manuals available. Do not forget to fix missing values, especially for returns and prices.

2. Using the risk-free rate of return from French's website³, report the following moments of the market excess returns for both time series (2 decimal digits): annualized return, annualized volatility, annualized Sharpe ratio, skewness, and excess kurtosis. You should be comparing between July 1926 to December 2024, at a monthly frequency.

- Suggested function: **PS1_Q2**

- Inputs

- * dataframe **Monthly_CRSP_Stocks**, the output of **PS1_Q1**
- * dataframe **FF_mkt**, with columns

Variable Name	Variable type
Year	Integer
Month	Integer
Market_minus_Rf	float
SMB	float
HML	float
Rf	float

- Note: Returns should be formatted in decimal proportion (not percent).

- Output

- * 5×2 numeric matrix (dataframe). Rows: Annualized Mean, Annualized Standard Deviation, Annualized Sharpe Ratio, Skewness, and Excess Kurtosis. Columns: Estimated FF Market Excess Return, Actual FF Market Excess Return.

³In the Fama/French 3 Factors dataset from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

3. Report (up to 8 decimal digits) the correlation between your time series and French's time series, and the maximum absolute difference between the two time series. It is zero? If not, justify whether the difference is economically negligible or not. What are the reasons a nonzero difference? You should be comparing between July 1926 to December 2024, at a monthly frequency.

- Suggested function: **PS1_Q3**
 - Inputs
 - * dataframe **Monthly_CRSP_Stocks**, the output of **PS1_Q1**
 - * dataframe **FF_mkt** as described in **PS1_Q2**
 - Output
 - * Vector of length two. Correlation between time series and maximum absolute difference between two time series.
- Hint: the correlation should be close to one, and the moments reported for both time series should be very very similar (maybe the same?).

Additional Material

As a reference to guide your replication exercise, here is the replication table obtained by the professor:

(Question 2) Summary statistics:

	French Data	Replication
Excess Return	8.24	8.24
Volatility	18.45	18.45
Sharpe Ratio	0.45	0.45
Skewness	0.15	0.15
Excess Kurtosis	7.44	7.44

Sample from 07/31/1926 to 12/31/2024

(Question 3) Correlation between French data and our replication: 0.99999957