Quantifying Hedge Fund Alpha

Part A - Data Wrangling and Visualization

Please code either in python entirely or in R entirely for Part A. If you have trouble completing Part A, you can start with Part B and Part C. However, you cannot use `data.csv` for Part A.

[*a*] The file, `Barclay_Hedge_Fund_Indexhistorical_data.xls`, is downloaded from

`https://portal.barclayhedge.com/cgi-bin/indices/displayHfIndex.cgi?indexCat=Barclay-Hedge-Fund-Indices&indexName=Barclay-Hedge-Fund-Index`

It contains the Barclay Hedge Fund Index performance data.

Note that the monthly returns in the file are in decimals, expressed as percent. For example, the return for Jan 1997 is 3.620% = 0.03620

Do not manually alter or process the file.

[*b*] The file, `F-F_Research_Data_Factors.CSV`, is downloaded from

`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/F-F_Research_Data_Factors_CSV.zip`

It contains the monthly returns of Fama-French Factor portfolios. Note that the monthly returns in the file are in percentage. These returns need to be converted to decimals, when used along with hedge fund returns. For example, the market excess return (Mkt-RF) for July 1926 is 2.96 percent = 0.0296.

- `Mkt-RF` is the excess return of the (proxy) market portfolio
- `SMB` is the return of a self-financing portfolio which is long small stocks and short large stocks
- `HML` is the return of a self-financing portfolio which is long high book-to-market stocks and short low book-to-market stocks.
- `RF` is the return of the (proxy) risk-free asset.

Do not manually alter or process the file.

[*c*] Code to produce the following performance statistics over the period Jan 1997 (inclusive) to Sep 2022(inclusive). We adopt the typical assumption that monthly returns of a portfolio are independent and come from the same distribution. And Sharpe Ratio = $\dfrac{E(R) - RF}{\sigma(R)}$

We see that over the past 25 years, hedge funds in aggregate have achieved a much higher excess returns per unit of risk exposure.

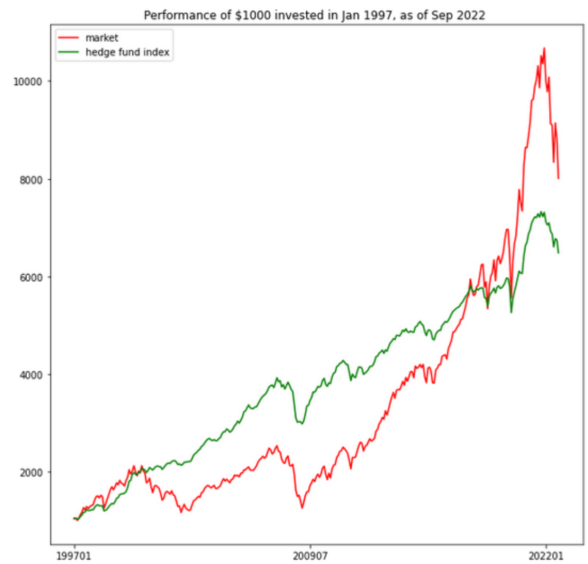
Note that you can use `pandas.read_excel()` to read in the xls file. Install the `xlrd` package to handle xls files if necessary.

	Market	Hedge Fund
Expected Excess Return (% per month)	0.630518	0.473269
Return Volatility (% per month)	4.675097	2.071968
Sharpe Ratio (annualized)	0.467194	0.791253

[*d*] Code to produce the following visual presentation of the growth of $1000 invested in either the Market Portfolio or the Hedge Fund Index.

Note that the monthly return of the (proxy) Market Portfolio is `Mkt-RF + RF`.

It is obvious that the aggregate hedge fund performance has been much less volatile, which is the main driver of the higher Sharpe ratio.



Please submit your work as hw13A.ipnb and hw13A.html to `Canvas`.

Part B -Quantifying Alpha by the traditional 3-factor model
Please code in python entirely for Part B.
If you have trouble completing Part A, you can start with Part B and Part C.

The file `data.csv` contains the necessary inputs - see Part A for a description of the data. Note that the monthly returns in `data.csv` are in decimals and expressed as decimals. For example, the excess return of the Hedge Fund Index for Jan 1997 is 0.0317 (which is 3.17%)

[*a*] Run the following linear regression, report the intercept and R^2.

$$E(HFI - RF) = \alpha + \beta_1(Mkt - RF) + \beta_2 SMB + \beta_3 HML$$

The intercept α, if statistically significant, represents the average excess return that is unaccounted for by the 3-factor model. Thus α can be interpreted as a measure of asset managers'ability to generate risk-adjusted return, with respect to this model. α, if present and adequate, is what justifies your high fees.

[*b*] Use the fitted model in [*a*] to predict and report the hedge fund index excess return for a month where the 3 factors take on their respective **median** historical value (over the period Jan 1997 (inclusive) to Sep 2022(inclusive)).

Please submit your work as hw13B.ipnb and hw13B.html to `Canvas`.

Part C -Quantifying Alpha by a modified Principal Component Regression
Please code in python entirely for Part C.
If you have trouble completing Part A, you can start with Part B and Part C.

The file `data.csv` contains the necessary inputs - see Part A for a description of the data. Note that the monthly returns in `data.csv` are in decimals and expressed as decimals. For example, the excess return of the Hedge Fund Index for Jan 1997 is 0.0317 (which is 3.17%)

[*a*] Let Z be the first principal component (score) of `(SMB, HML)`. Note that we do not consider `Mkt-RF` when constructing Z. This is the direction that explains the most variance of `(SMB, HML)`.

Run the following linear regression, report the intercept and R^2.

$$E(HFI - RF) = \alpha + \beta_1(Mkt - RF) + \beta_2 Z$$

This is a more parsimonious model than that in part B. This model is as competitive (similar R^2) but of lower dimension and likely lower prediction variance.

The intercept α, if statistically significant, represents the average excess return that is unaccounted for by the 3-factor model. Thus α can be interpreted as a measure of asset managers'ability to generate risk-adjusted return, with respect to this model. α, if present and adequate, is what justifies your high fees.

[*b*] Use the fitted model in [*a*] to predict and report the hedge fund index's excess return for a month where each of the 3 factors takes on its **median** historical value (over the period Jan 1997 (inclusive) to Sep 2022(inclusive)).

Please submit your work as hw13C.ipnb and hw13C.html to `Canvas`.