FINM 25000 **Mark Hendricks**
Summer 2022

Homework #2

Due on Tuesday, June 28, at 6:00pm.

ProShares Hedge Replication ETF [UV6939].
The case is only seven pages, though it contains many exhibits. No need to study/review the exhibits beyond what is useful in addressing the homework questions. But they are interesting if you are inclined.

1 The ProShares ETF Product

- **This section is not graded, and you do not need to submit your answers.**
- But you are expected to consider these issues and be ready to discuss them.
- This section requires no empirical analysis; answer solely based on the material given in the case.

1. "Alternative ETFs"

   Describe the two types of investments referenced by this term.

2. Hedge Funds.

   (a) Using just the information in the case, what are two measures by which hedge funds are an attractive investment?

   (b) What are the main benefits of investing in hedge funds via an ETF instead of directly?

3. The Benchmarks

   (a) Explain as simply as possible how HFRI, MLFM, MLFM-ES, and HDG differ in their construction and purpose.

   (b) How well does the Merrill Lynch Factor Model (MLFM) track the HFRI?

   (c) In which factor does the MLFM have the largest loading? (See a slide in Exhibit 1.)

   (d) What are the main concerns you have for how the MLFM attempts to replicate the HFRI?

4. The HDG Product

   (a) What does ProShares ETF, HDG, attempt to track? Is the tracking error small?

   (b) HDG is, by construction, delivering beta for investors. Isn't the point of hedge funds to generate alpha? Then why would HDG be valuable?

   (c) The fees of a typical hedge-fund are 2% on total assets plus 20% of excess returns if positive. HDG's expense ratio is roughly 1% on total assets. What would their respective net Sharpe Ratios be, assuming both have a gross excess returns of 10% and volatility of 20%?

2 Analyzing the Data

- Use the data found in the github repo, **'proshares_analysis_data.xlsx'**. It has monthly data on financial indexes and ETFs from Aug 2011 through Sep 2021.

- Feel free to use some provided python functions located in the github repo, in the '`cmds`' folder. The TA discussion will review how to import and use these functions. If you use the functions in `cmds/portfolio.py`, I strongly encourage you to review the functions to understand how they work.

- You do not need to use the provided code; feel free to write your own functions to calculate regressions and the statistics in 2.2.

1. For the series in the "hedge_fund_series" tab, report the following summary statistics:[1]

   (a) mean

   (b) volatility

   (c) Sharpe ratio

   Annualize these statistics.

2. For the series in the "hedge_fund_series" tab, , calculate the following statistics related to tail-risk.

   (a) Skewness[2]

   (b) Kurtosis[3]

   (c) the fifth quantile of historic returns, which is also known as the Value-at-Risk (VaR)[4]

   (d) the mean of the returns at or below the fifth quantile, which is also known as the Conditional Value-at-Risk (CVaR)

   (e) Maximum drawdown - include the dates of the max/min/recovery within the max drawdown period.

   There is no need to annualize any of these statistics.

3. For the series in the "hedge_fund_series" tab, run a regression of each against SPY (found in the "merrill_factors" tab.) Include an intercept. Report the following regression-based statistics:

   (a) Market Beta

   (b) Treynor Ratio

   (c) Information ratio

   No need to annualize the market beta. The Treynor ratio is annualized by multiplying by the number of periods in a year (12). The Information ratio is annualized by multiplying by $\sqrt{12}$.

[1] Technically, the Sharpe Ratio is calculated as the mean over volatility of *excess* returns. Here we are using total returns, but we still refer to this mean-volatility ratio as the Sharpe ratio.

[2] In pandas, try `.skew()`

[3] In pandas, try `.kurtosis()`

[4] In pandas, try `.quantile(.05)`

4. Relative Performance

   Discuss the previous statistics, and what they tell us about...

   (a) the differences between SPY and the hedge-fund series?

   (b) which performs better between HDG and QAI.

   (c) whether HDG and the ML series capture the most notable properties of HFRI.

5. Report the correlation matrix for these assets.

   (a) Show the correlations as a heat map.

   (b) Which series have the highest and lowest correlations?

6. Replicate HFRI with the six factors listed on the "merrill_factors" tab. Include a constant, and run the unrestricted regression,[5]

$$r_t^{\text{hfri}} = \alpha^{\text{merr}} + \boldsymbol{x}_t^{\text{merr}} \boldsymbol{\beta}^{\text{merr}} + \epsilon_t^{\text{merr}} \tag{1}$$

$$\hat{r}_t^{\text{hfri}} \equiv \hat{\alpha}^{\text{merr}} + \boldsymbol{x}_t^{\text{merr}} \hat{\boldsymbol{\beta}}^{\text{merr}} \tag{2}$$

   Note that the second equation is just our notation for the fitted replication.[6]

   (a) Report the intercept and betas.

   (b) Are the betas realistic position sizes, or do they require huge long-short positions?

   (c) Report the R-squared.

   (d) Report the volatility of ϵ^{merr}, (the tracking error.)

[5] ML restricts their regression to ensure the position sizes stay within certain bounds. We leave it unrestricted.

[6] This is just the fitted value of your regression, which in classic regression notation would be $\hat{y}$, and based on your regression fit for y.

3 Extensions

This section is not graded, and you do not need to submit your answers. This is only provided for those interested in further study.

1. Let's examine the replication out-of-sample.

   Starting with $t = 61$ month of the sample, do the following:

   - Use the previous 60 months of data to estimate the regression equation, (1). This gives time-t estimates of the regression parameters, $\widetilde{\alpha}_t^{\text{merr}}$ and $\widetilde{\boldsymbol{\beta}}_t^{\text{merr}}$
   - Use the estimated regression parameters, along with the time-t regressor values, $\boldsymbol{x}_t^{\text{merr}}$, to calculate the time-$t$ replication value[7] that is, with respect to the regression estimate, built "out-of-sample" (OOS).

   $$\widetilde{r}_t^{\text{hfri}} \equiv \widetilde{\alpha}_t^{\text{merr}} + (\boldsymbol{x}_t^{\text{merr}})' \widetilde{\boldsymbol{\beta}}_t^{\text{merr}}$$

   - Step forward to $t = 62$, and now use $t = 2$ through $t = 61$ for the estimation. Re-run the steps above, and continue this process throughout the data series. Thus, we are running a rolling, 60-month regression for each point-in-time.

   How well does the out-of-sample replication perform with respect to the target?

2. We estimated the replications using an intercept. Try the full-sample estimation, but this time without an intercept.

   $$r_t^{\text{hfri}} = \alpha^{\text{merr}} + \boldsymbol{x}_t^{\text{merr}} \boldsymbol{\beta}^{\text{merr}} + \epsilon_t^{\text{merr}}$$
   $$\check{r}_t^{\text{hfri}} \equiv \check{\alpha}^{\text{merr}} + \boldsymbol{x}_t^{\text{merr}} \check{\boldsymbol{\beta}}^{\text{merr}}$$

   Report

   (a) the regression beta. How does it compare to the estimated beta with an intercept, $\hat{\boldsymbol{\beta}}^{\text{merr}}$?

   (b) the mean of the fitted value, $\check{r}_t^{\text{hfri}}$. How does it compare to the mean of the HFRI?

   (c) the correlations of the fitted values, $\check{r}_t^{\text{hfri}}$ to the HFRI. How does the correlation compare to that of the fitted values with an intercept, $\hat{r}_t^{\text{hfri}}$

   Do you think Merrill and ProShares fit their replicators with an intercept or not?

3. Merrill constrains the weights of each asset in its replication regression of HFRI. Try constraining your weights by re-doing 2.6.

   (a) Use Non-Negative Least Squares (NNLS) instead of OLS.[8]

   (b) Go further by using a Generalized Linear Model to put separate interval constraints on each beta, rather than simply constraining them to be non-negative.[9]

4. Let's decompose a few other targets to see if they behave as their name suggests.

[7] This is just a single number!

[8] Try using LinearRegression in scikit-learn with the parameter "positive=True".

[9] Try using GLM in statsmodels.

(a) Regress HEFA on the same style factors used to decompose HFRI. Does HEFA appear to be a currency-hedged version of EFA?

(b) Decompose TRVCI with the same style factors used to decompose HFRI. The TRVCI Index tracks venture capital funds–in terms of our styles, what best describes venture capital?

(c) TAIL is an ETF that tracks SPY, but that also buys put options to protect against market downturns. Calculate the statistics in questions 2.1-2.3 for TAIL. Does it seem to behave as indicated by this description? That is, does it have high correlation to SPY while delivering lower tail risk?

5. The ProShares case introduces Levered ETFs. ProShares made much of its name originally through levered, or "geared" ETFs.

(a) Explain conceptually why Levered ETFs may track their index well for a given day but diverge over time. How is this exacerbated in volatile periods like 2008?

(b) Analyze SPXU and UPRO relative to SPY.

- SPXU is ProShares -3x SPX ETF.
- UPRO is ProShres +3x SPX ETF.

i. Analyze them with the statistics from 2.1-2.3. Do these two ETFs seem to live up to their names?

ii. Plot the cumulative returns of both these ETFs along with SPY.

What do you conclude about levered ETFs?