FINM 36700 **Mark Hendricks**
Autumn 2021

Homework #5

Due on Monday, November 1, at 6:00pm.

Smart Beta Exchange-Traded-Funds and Factor Investing.
This case is a good introduction to important pricing factors. It also gives useful introduction and context to ETFs, passive vs active investing, and so-called "smart beta" funds.

1 The Case

This section will not be graded, but it will be discussed in class.

1. Describe how each of the factors (other than MKT) is measured.[1] That is, each factor is a portfolio of stocks–which stocks are included in the factor portfolio?
2. Is the factor portfolio...
   - long-only
   - long-short
   - value-weighted
   - equally-weighted
3. What steps are taken in the factor construction to try to reduce the correlation between the factors?
4. What is the point of figures 1-6?
5. How is a "smart beta" ETF different from a traditional ETF?
6. Is it possible for all investors to have exposure to the "value" factor?
7. How does factor investing differ from traditional diversification?

[1] If you need more info in how these factor portfolios are created, see Ken French's website, and the following details: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_mom_factor.html

2 The Factors

Use the data found in **'factor_pricing_data.xlsx'**.

- FACTORS: Monthly **excess** return data for the overall equity market, $\tilde{r}^{\text{MKT}}$. The sheet also contains data on five additional factors. All factor data is already provided as **excess returns**.[2]

1. Analyze the factors, similar to how you analyzed the three Fama-French factors in Homework 4. You now have three additional factors, so let's compare there univariate statistics.

   - mean
   - volatility
   - Sharpe

2. Based on the factor statistics above, answer the following.

   (a) Does each factor have a positive risk premium (positive expected excess return)?

   (b) How have the factors performed since the time of the case, (2015-present)?

3. Report the correlation matrix across the six factors.

   (a) Does the construction method succeed in keeping correlations small?

   (b) Fama and French say that HML is somewhat redundant in their 5-factor model. Does this seem to be the case?

4. Report the tangency weights for a portfolio of these 6 factors.

   (a) Which factors seem most important? And Least?

   (b) Are the factors with low mean returns still useful?

   (c) Re-do the tangency portfolio, but this time only include MKT, SMB, HML, and UMD. Which factors get high/low tangency weights now?

   What do you conclude about the importance or unimportance of these styles?

[2] The column header to the market factor is "MKT" rather than "MKT-RF", but it is indeed already in excess return form.

3 Testing Modern LPMs

Consider the following factor models:

- CAPM: MKT
- Fama-French 3F: MKT, SMB, HML
- Fama-French 5F: MKT, SMB, HML, RMW, CMA
- AQR: MKT, HML, RMW, UMD

For instance, for the AQR model[3]

$$\mathbb{E}\left[\tilde{r}^i\right] = \beta^{i,\text{MKT}}\mathbb{E}\left[\tilde{f}^{\text{MKT}}\right] + \beta^{i,\text{HML}}\mathbb{E}\left[\tilde{f}^{\text{HML}}\right] + \beta^{i,\text{RMW}}\mathbb{E}\left[\tilde{f}^{\text{RMW}}\right] + \beta^{i,\text{UMD}}\mathbb{E}\left[\tilde{f}^{\text{UMD}}\right] \tag{1}$$

We will test these models with the time-series regressions. Namely, for each asset i, estimate the following regression to test the AQR model:

$$\tilde{r}^i_t = \alpha^i + \beta^{i,\text{MKT}}\tilde{f}^{\text{MKT}}_t + \beta^{i,\text{HML}}\tilde{f}^{\text{HML}}_t + \beta^{i,\text{RMW}}\tilde{f}^{\text{RMW}}_t + \beta^{i,\text{UMD}}\tilde{f}^{\text{UMD}}_t + \epsilon_t \tag{2}$$

So you are running that regression n times, once for each security, $\tilde{r}^i$.
Data

- PORTFOLIOS: Monthly **excess** return data on 49 equity portfolios sorted by their industry. Denote these as $\tilde{r}^i$, for $n = 1, \ldots, 49$.

- You do NOT need the risk-free rate data. It is provided only for completeness. The other two tabs are already in terms of excess returns.

1. Test the AQR 4-Factor Model using the time-series test. (We are not doing the cross-sectional regression tests.)

   (a) For each regression, report the estimated α and r-squared.

   (b) Calculate the mean-absolute-error of the estimated alphas, (one for each security, $\tilde{r}^i$.)

   $$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\hat{\alpha}^i|$$

   If the pricing model worked, should these alpha estimates be large or small? Why? Based on your MAE stat, does this seem to support the pricing model or not?

2. Test the CAPM, FF 3-Factor Model and the the FF 5-Factor Model. Report the MAE statistic for each of these models and compare it with the AQR Model MAE. Which model fits best?

3. Does any particular factor seem especially important or unimportant for pricing? Do you think Fama and French should use the Momentum Factor?

4. This does not matter for pricing, but report the average (across n estimations) of the time-series regression r-squared statistics. Do this for each of the three models you tested. Do these models lead to high time-series r-squared stats? That is, would these factors be good in a Linear Factor Decomposition of the assets?

[3] We are not saying this is "the" AQR model, but it is a good illustration of their most publicized factors: value, momentum, and more recently, profitability.

4 Extensions

1. In Problem 2.1 we tested three models using the time-series tests (focusing on the time-series alphas.) Re-test these models, but this time use the cross-sectional test, as we did in Homework 4. Unlike in Homework 4, do NOT include an intercept in the cross-sectional regression.

   (a) Report the time-series premia of the factors (just their sample averages,) and compare to the cross-sectionally estimated premia of the factors. Do they differ substantially?[4]

   (b) Report the MAE of the cross-sectional regression residuals for each of the four models, (the υ^i.) How do they compare to the MAE of the time-series alphas?

2. Using the LPM.

   Let's use the AQR model in (1) for forecasting excess returns. We will do this at each point in time to build a point-in-time series of forecasts. We will then see how well they perform.

   - The model does not give us any info about forecasting the factors themselves. Accordingly, calculate the "expanding" mean of the four factors. We will use these as our point-in-time factor premia.[5]

   - For each of the n securities, estimate (2) over a window of 60 months. Make sure to estimate these rolling regressions WITH an intercept.[6] But we only need to save the beta estimates.[7]

   - For every security, i, and at every month, t (after the first 60), calculate (1) using the point-in-time factor premia and betas calculated in the prior two steps. This is your forecast made at the end of period t, for $\tilde{r}^i_{t+1}$. You are using end-of-time t info in the estimation, so it is a forecast for $t+1$. In order to better align it with our data, shift it ahead a time period.[8] So the dataframe of forecasts has been pushed one month later. (The Feb value is now a March value.) This is your forecasted table for $\tilde{r}^i_t$. Denote these as $\widehat{\tilde{r}}^i_t$.

   - In order to decide if these forecasts are good, we need a comparison. Use the point-in-time mean estimates of $\tilde{r}^i_t$. So calculate the expanded mean, and once again, be sure to shift them one period into the future. This gives us the benchmark forecast: $\overline{\tilde{r}}^i_t$

   - Compare our LFP forecasts with the naive forecasts using Out-of-Sample (OOS) R-squared.

   $$\text{OOS r-squared} = 1 - \frac{\text{MSE}_{\text{forecast}}}{\text{MSE}_{\text{baseline}}}$$

   where MSE stands for Mean Squared Error. Thus,

   $$\begin{aligned} \text{MSE}_{\text{forecast}} \equiv &\quad \text{sample average of } \left[\left(\widehat{\tilde{r}}^i_t - \tilde{r}^i_t\right)^2\right] \\ \text{MSE}_{\text{baseline}} \equiv &\quad \text{sample average of } \left[\left(\overline{\tilde{r}}^i_t - \tilde{r}^i_t\right)^2\right] \end{aligned}$$

---

[4] Recall that we found in Homework 4 that the market premium went from being strongly positive to strongly negative when estimated in the cross-section with an intercept.

[5] You may find the following pandas command helpful: `.expanding().mean()`

[6] You may wish to use `from statsmodels.regression.rolling import RollingOLS`

[7] This will take longer to compute: we are estimating a multifactor regression at every month in time and for every security. So we are running roughly $T \times N$ regressions.

[8] See `.shift()` in pandas.

Warning! This calculation will be wrong if your forecasts have NaN values where the benchmark does not. For this reason, it is important to eliminate any date where either series has an NaN value.[9] If you are careful about this issue, then you can write the OOS r-squared as a ratio of SSE:

$$\begin{aligned}\text{OOS r-squared} =& 1 - \frac{\text{SSE}_{\text{forecast}}}{\text{SSE}_{\text{baseline}}}\\ =& 1 - \frac{\sum_t \left(\widehat{\tilde{r}}^i_t - \tilde{r}^i_t\right)^2}{\sum_t \left(\overline{\tilde{r}}^i_t - \tilde{r}^i_t\right)^2}\end{aligned}$$

(a) Report the OOS r-squared for each of the n security forecasts.

(b) Does the LPM do a good job of forecasting monthly returns? For which asset does it perform best? And worst?

(c) Re-do the exercise using a window of 36 months. And 96 months. Do either of these windows work better?

(d) Re-do the exercise using the FF 5-Factor Model instead of the AQR model. Re-do it with the CAPM. Do either of these models improve on forecasting?

[9] For instance, if you use the rolling regressions, your initial forecast values will be NaN. But your expanded mean calculation for the baseline will not have any NaN. Thus, it is important to require a minimum number of observations in the expanded mean. Or you can more explicitly enforce that both dataframes have NaNs in the same time periods.