BUFN 640 – Problem Set 2

Due on Sunday, October 9 at 11:59 pm

This assignment will show you how to estimate and test the Fama-French 3-factor model and solve the Markowitz portfolio choice problem in Python. Please use the template Python notebook (*Problem Set 2 – template.ipynb*) in the shared Google Drive folder or on Canvas to load the data. Implement the rest of the file yourself.

Important: Please submit your homework using Canvas. Your submission needs to include **a single** file – a copy of your Python notebook (*.ipynb* Jupyter notebook file). To produce the latter, please click *File* → *Download .ipynb* in Google Colab, then save and upload the file on Canvas. Multiple-choice questions are to be answered in the notebook file, too. Each student has to submit his/her individual assignment.

Part I: Multiple-choice questions (25 points)

1. (5 points) If our regression equation is $y = X\beta + u$, where we have T observations and K regressors, what will be the dimension of β using the standard matrix notation

   (a) $T \times K$

   (b) $T \times 1$

   (c) $K \times 1$

   (d) $K \times K$.

2. (5 points) Suppose that the value of R^2 for an estimated regression model is exactly one. Which of the following are true?

   i. All of the data points must lie exactly on the line

   ii. All of the residuals must be zero

   iii. All of the variability of y around its mean has been explained by the model

   iv. The fitted line will be horizontal with respect to all of the explanatory variables.

   (a) (ii) and (iv) only

   (b) (i) and (iii) only

(c) (i), (ii), and (iii) only

(d) (i), (ii), (iii), and (iv).

3. (5 points) Which of these is a mathematical expression of the residual sum of squares?

   i. $\hat{u}'\hat{u}$

   ii. $[\hat{u}_1, \hat{u}_2, ..., \hat{u}_T]$

   iii. $\hat{u}_1 + \hat{u}_2 + ... + \hat{u}_T$

   (a) (i) only

   (b) (i) and (ii) only

   (c) (i) and (iii) only

   (d) (i), (ii), and (iii).

4. (5 points) Why is R^2 a commonly used and perhaps better measure of how well a regression model fits the data than the residual sum of squares (RSS)?

   (a) The RSS is often too large

   (b) The RSS does not depend on the scale of the dependent variable whereas the R^2 does

   (c) The RSS depends on the scale of the dependent variable whereas the R^2 does not

   (d) The RSS depends on the scale of the independent variable whereas the R^2 does not.

5. (5 points) In the following regression estimated on 64 observations:

$$y_t = \beta_1 + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \beta_4 X_{4,t} + u_t,$$

   Which of the following null hypotheses could we test using an F-test?

   i. $\beta_2 = 0$

   ii. $\beta_2 = 1$ and $\beta_3 + \beta_4 = 1$

   iii. $\beta_3 \beta_4 = 1$

   iv. $\beta_2 - \beta_3 - \beta_4 = 1$.

(a) (i) and (ii) only

(b) (ii) and (iv) only

(c) (i) and (iii) only

(d) (i), (ii), (iii), and (iv)

(e) (i), (ii), and (iv) only.

Part II: The Fama and French Model (40 points)

CAPM In the previous assignment you used a univariate linear regression to test the Capital Asset Pricing Model (CAPM). CAPM relates the expected return on portfolio p – denoted by $E[R_p]$ – and the expected return on the market – denoted by $E[R_m]$ – as follows:

$$E\left[R_p\right] - R_f = \beta_p \left(E[R_m] - Rf\right),$$

where R_f is the risk-free rate and $\beta_p = \frac{Cov[R_p,R_m]}{Var[R_m]}$. An empirical test of the CAPM is based on the following regression:

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_p \left(R_{m,t} - R_{f,t}\right) + u_{p,t}, \quad t = 1, \ldots, T$$

implying that, if CAPM holds, the intercept α_p should be equal to 0.

Fama-French 3-factor (FF3) Another very popular asset pricing model in the empirical finance literature is the Fama-French 3-factor (FF3) that was published in 1993. Eugene Fama and Kenneth French found that value stocks tend to outperform growth stocks (i.e., value), and that small-cap stocks outperform large-cap stocks (i.e., size). Thus, the FF3 mode adds in size and value as risk factors to the model as shown below

$$E\left[R_p\right] - R_f = \alpha_p + \beta_{p,M} \left(R_{m,t} - R_{f,t}\right) \underbrace{+\beta_{p,S} R_{SMB,t} + \beta_{p,h} R_{HML,t}}_{\text{new factors}} + u_{p,t}.$$

Setup

Use 25 Fama and French portfolios from Ken French's *website* at the **monthly** frequency, which contain returns on portfolios sorted on size (5 portfolios) and book-to-market (5 portfolios). The total is, therefore, 25 portfolios. Start your sample in 1926 (to present). The portfolios are named according to their first and second sorting variable. In particular:

- *SMALL LoBM, ME1 BM1, ..., ME1 HiBM* are Small portfolios (split into 5 book-to-market portfolios)
- *...*
- *Big LoBM, ME5 BM2, ..., ME5 HiBM* are Large portfolios (split into 5 book-to-market portfolios)

You may use the example from class to load data into Python using the DataReader interface (don't forget to change to monthly frequency). Specifically, the following piece of code downloads the data from the website, unpacks it, and load into Python as a Pandas table:

import pandas_datareader.data as web
r = web.DataReader('25_Portfolios_5x5', 'famafrench', start=1900)[0]/100

I have also prepared a Colab template notebook which guides you through these steps.

Questions

Using Python:

1. (10 points) Estimate the intercept α_p and the slope β_p for all 25 portfolios using OLS for both the CAPM and the FF3. Are the annualized α_p's economically large? (You can annualize monthly α_p simply by multiplying it by 12). How different are the magnitudes of alpha in the CAPM vs FF3 model? How different are the R^2 for the two models? Hint: you can use my regression function from the *FamaFrench* notebook.

2. (10 points) For the Small HiBM portfolio test the null that $\beta_{SMB} = \beta_{HML}$ using the F-test. What's your conclusion, can you reject the hypothesis that the two coefficients are equal? Hint: we have tested a similar restriction in the *FamaFrench* notebook.

3. (5 points) Test the null hypothesis $H_0 : \alpha_p = 0$ for each of the 25 portfolios, using a two-sided alternative hypothesis and the FF3 model. Specify the test statistic for a level of significance equal to 5%. Compute standard errors and the p-values for all α_p. Hint: this should be easy and requires no manual calculations. You may rely on the output of a Python's regression function or my regression function from the *FamaFrench* notebook.

4. (5 points) Are the estimated betas in the FF3 model what you expected them to be for all the portfolios? That is, what are the average values of $\beta_{p,M}, \beta_{p,s}, \beta_{p,h}$? Are some betas higher or lower for small/big/value/growth stocks?

5. (5 points) Which portfolios seem to be the best performing in terms of expected returns? Which ones are the worst performing? Do you see any pattern between size and/or book-to-market and performance?

6. (5 points) If you were an asset manager, what kind of portfolios would you invest in? If you could short some of the portfolios, would you do that?

Part III: An Optimal Portfolio of Industries (35 points)

Modern Portfolio Theory Modern Portfolio Theory (MPT) is an investment theory developed by Harry Markowitz and published under the title ”Portfolio Selection” in the Journal of Finance in 1952.

Fundamentally, this is a theory of “risk-return trade-off.” Higher risk is associated with greater probability of higher return and lower risk with a greater probability of smaller return. MPT assumes that investors are risk-averse, meaning that given two portfolios that offer the same expected return, investors will prefer the less risky one. Thus, an investor will take on increased risk only if compensated by higher expected returns.

Another factor comes in to play in MPT is “diversification.” Modern portfolio theory says that it is not enough to look at the expected risk and return of one particular stock. By investing in more than one stock, an investor can reap the benefits of diversification – chief among them, a reduction in the riskiness of the portfolio.

What you need to understand is “risk of a portfolio is not equal to average/weighted-average of individual stocks in the portfolio.” In terms of return, yes, it is the average/weighted average of individual stock's returns, but that's not the case for risk. The risk is about how volatile the asset is, if you have more than one stock in your portfolio, then you have to take count of how these stocks movement correlates with each other. The beauty of diversification is that you can even get lower risk than a stock with the lowest risk in your portfolio, by optimising the allocation.

Setup

Start by downloading and importing **monthly** industry data from Ken French's website. This is the data you will use in your group project as well. To load the data, you can use the DataReader interface. Specifically, the following piece of code downloads the data from the website, unpacks it, and load into Python as a Pandas table:

```
import pandas_datareader.data as web
r = web.DataReader('12_Industry_Portfolios', 'famafrench', start=1900)[0]/100
```

Next load the risk-free rate, merge it with returns, and compute excess returns on every portfolio (in excess of the risk-free rate). The template Colab file shows how to perform these steps.

Questions

Using Python:

1. (5 points) Now that have obtain excess returns on every industry, compute and show their means. Which industries have performed well?

2. (15 points) Compute the minimum variance portfolio using these 12 industries. Recall, the minimum variance portfolio is the portfolio that delivers the lowest variance of any possible investment portfolio. Use the formula I showed in class to find this portfolio's weights. Which industries have the highest/lowest weights? Perhaps you found that some industries that had high mean returns in the previous question command a negative weight in the minimum-variance portfolio (that is, we would want to short sell these industries). Explain why this could be the case.

3. (15 points) Compute the weights of the portfolio with the highest possible Sharpe ratio using the formula I showed in class. Compare the weights to the weights of the minimum variance portfolio. Recall, the highest Sharpe ratio portfolio is the portfolio that delivers the highest possible expected excess return per unit of risk (as measured by the standard deviation of portfolio returns). Which industries have the highest/lowest weights? Perhaps you found that some industries that had high mean returns in the previous question command a negative weight in the maximum Sharpe Ratio portfolio (that is, we would want to short sell these industries). Explain why this could be the case. Hint: you may use the *Efficient Frontier and Modern Portfolio Theory* notebook I discussed in class. Specifically, the section "The Markowitz unconstrained solution (using matrix algebra)" performs this computation.