Empirical Asset Pricing, Assignment 2

Due: 10:00 AM on February 14, 2024 (Wednesday)

The assessment on this course will be based on three graded assignments. The assignments are due on:

- January 29, 10:00.
- February 14, 10:00.
- February 23, 10:00.

Computer language(s)

You should be prepared to do applied modeling on a computer, using a computer language that can perform matrix operations efficiently. I use Matlab, and therefore this is the language where I can provide the most help if needed, but any language that can perform matrix operations efficiently is acceptable. It is your responsibility to make sure you have access to adequate software.

Types of problems

You fill find three types of problems in the assignments:

a. **Compulsory; group**: These are compulsory problems that can be solved in groups of *two* students (you can also choose to do them individually, but no extra credit is given for working alone). You are encouraged to work in groups of two on these problems and write them up as a group.

b. **Compulsory; individual**: These are compulsory problems which should be addressed on an individual basis.

c. **Optional; individual**: These are optional problems which should be addressed on an individual basis. A genuine attempt on the optional problems helps in getting a pass with distinction grade.

Every exercise is labeled as one of the above types. **The requirement for a pass grade is that all compulsory problems in the assignments have been done satisfactory**.

Datasets

The assignments involve various datasets. Some of the data used throughout the assignments come from commercial data vendors and should therefore only be used for the assignments. Do not use the data outside this course without my prior consent.

Report and submission

You have to prepare a written report (typed, not hand-written) that contains the answers to the problems. Send the report (as a PDF file) to `adam.farago@cff.gu.se.` before the due date. Late assignments will not be graded. Results should be organized into tables and graphs for readability. You do not need to make fancy graphs and tables, but they should be clear and readable. The computer code for each assignment should also be attached to the report. Please provide ample documentation in the code describing what you are trying to achieve.

You should write up and submit the group problems as a group, while the solutions to the individual problems should be written up and submitted individually. The cover page should state: Course name, date, name(s), and which problem's you submit the solution to.

Contact me if you have questions (if you are not sure what exactly is required at a certain problem, it is better to clarify than to guess and be wrong).

Consumption-Based Asset Pricing (Lecture 4)

1. (**Compulsory; individual**) Consider the canonical asset pricing equation

$$\mathrm{E}_{t-1}\left[M_t R_{i,t}\right] = 1,$$

where M_t is a stochastic discount factor, $R_{i,t}$ is the gross return on asset i, and $\mathrm{E}_{t-1}[.]$ denotes the conditional expectations given information at date $t-1$. In the case of an representative investor with time discount factor δ who maximizes a time-separable utility function defined over aggregate consumption $U(C_t)$, the stochastic discount factor equals the investor's intertemporal marginal rate of substitution (i.e., $M_t = \delta \frac{U'(C_t)}{U'(C_{t-1})}$). The asset pricing equation is then also the investor's first-order condition. With power utility, this first-order condition is:

$$\mathrm{E}_t\left[\delta\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_{i,t}\right] = 1\ . \tag{1}$$

Note that by the law of iterated expectations the first-order condition also holds unconditionally. Under the additional assumption that the consumption growth and returns are jointly log-normally distributed, expected excess returns can be written

$$\mathrm{E}\left[r_{i,t} - r_{f,t}\right] + \frac{\sigma_i^2}{2} = \gamma\sigma_{ic}\ , \tag{2}$$

and the risk-free rate can be written

$$r_{f,t} = -\ln\delta + \gamma\mathrm{E}\left[\Delta c_t\right] - \frac{\gamma^2\sigma_c^2}{2}\ , \tag{3}$$

where $r_{i,t}$ is the log return on asset i between date $t-1$ (e.g., end of year $t-1$) and date t (end of year t), $r_{f,t}$ is the log risk-free rate for the same period, σ_i is the standard deviation of the log return on asset i, σ_{ic} is the covariance between the log return on asset i and log consumption growth, and σ_c is the standard deviation of log consumption growth. Note that $r_{i,t} = \ln(R_{i,t})$, $r_{f,t} = \ln(R_{f,t})$, $c_t = \ln(C_t)$, and $\Delta c_t = c_t - c_{t-1}$.

Consider the annual data for the US in sheet "Consumption" of the Excel file "Assignment2Data_Gx.xlsx". The dataset begins in 1930 and contains the log consumption growth of non-durable goods and services (i.e., Δc_t, in the column "ConsGrowth"), log returns on an aggregated stock market portfolio ($r_{m,t}$, in the column "Market") and a T-bill portfolio ($r_{f,t}$, in the column "Rfree"). The final six columns ("S1" to "B3") contain log returns on six size/book-to-market sorted portfolios (returns on these six portfolios are only needed for the optional problem 2). Note that all variables are in logs and real terms.

(a) Provide means, standard deviations, and first-order autocorrelation coefficients for Δc_t, $r_{m,t}$, $r_{f,t}$, and $r_{m,t} - r_{f,t}$. Also, report the sample correlation between $r_{m,t}$ and Δc_t.

(b) Consider equation (2) for the aggregate stock market return (i.e., $i = m$). From the estimates in problem 1a, back out two measures of the risk aversion coefficient γ. The first measure uses the sample covariance between the stock return and consumption growth as an estimate of σ_{mc}. The second measure assumes that the correlation between excess returns on stocks and consumption growth equals one, which means that an estimate of their covariance is $\hat{\sigma}_{mc} = \hat{\sigma}_m \hat{\sigma}_c$. Interpret your results.

(c) Consider now equation (3) for the (average) risk-free rate. Use the two measure of relative risk aversion from 1b together with the estimates from 1a to back out measures of the (constant) time discount factor parameter δ and the time preference rate $-\ln\delta$. Comment on the results.

(d) The problem with the calculations in 1b and 1c is that the γ and δ estimates do not come with standard errors. Using the moment conditions

$$\mathrm{E}\begin{bmatrix} \Delta c_t - \mu_c \\ r_{m,t} - \mu_m \\ r_{m,t} - r_{f,t} + \frac{1}{2}(r_{m,t} - \mu_m)^2 - \gamma(r_{m,t} - \mu_m)(\Delta c_t - \mu_c) \\ r_{f,t} + \ln\delta - \gamma\Delta c_t + \frac{1}{2}\gamma^2(\Delta c_t - \mu_c)^2 \end{bmatrix} = 0 ,$$

Estimate the parameters δ, γ, μ_c, and μ_m. Also report the Newey and West (1987) standard errors of the parameter estimates using one lag. Comment on the results, more specifically on the γ and δ estimates and their standard errors.

The above moment conditions correspond to the case from problems 1b and 1c where you used the sample covariance between the stock return and consumption growth as an estimate of σ_{mc}. What moment conditions would you use to estimate the parameters corresponding to the second approach (assuming that the correlation between excess returns on stocks and consumption growth equals one)?

(e) Estimate γ and δ using a different set of moment conditions instead, namely:

$$\mathrm{E}\begin{bmatrix} \exp(\ln\delta - \gamma\Delta c_t + r_{m,t}) - 1 \\ \exp(\ln\delta - \gamma\Delta c_t + r_{f,t}) - 1 \end{bmatrix} = 0 ,$$

which correspond to the asset pricing equation (1) applied for the aggregate market and the risk-free asset, respectively. Also report the Newey and West (1987) standard errors of the parameter estimates using one lag. Comment on the results, and relate them to the results in 1d. What implicit assumption do you make in 1d, that you are not making here in 1e?

2. (**Optional; individual**) This problem is a continuation of problem 1, so the setup is the same. Estimate γ and δ using the following set of moment conditions:

$$\mathrm{E}\left[\ \exp(\ln\delta - \gamma\Delta c_t + \mathbf{r}_t) - \mathbf{1}\ \right] = \mathbf{0}\ , \tag{4}$$

with

$$\mathbf{r}_t = \left[r_{m,t}, r_{f,t}, r_{S1,t}, r_{S2,t}, r_{S3,t}, r_{B1,t}, r_{B2,t}, r_{B3,t}\right]'\ ,$$

where $\mathbf{0}$ is an appropriately sized column vector of zeros, $\mathbf{1}$ is an appropriately sized column vector of ones, and the function exp(.) operates element-wise. That is, you have now eight moment conditions (corresponding to the pricing of eight assets) and only two parameters, so the system is overidentified.

(a) Using the notation from Lecture 1, you can estimate the parameter vector $\theta = [\gamma, \delta]'$ by finding the solution to $A_T g_T(\theta) = 0$. How would you define A_T to get the same parameter estimates as in problem 1e but using the moment conditions defined in equation (4)?

Carry out the estimation using this A_T and report the result of a test for overidentifying restrictions (test statistic and associated p-value). Interpret the results.

(b) Alternatively, the parameters of the overidentified system in (4) can be estimated by finding the θ that minimizes $Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$, where W_T is a weighting matrix. Estimate γ and δ using the moment conditions in (4) and the identity weighting matrix, $W = I$. Also report the Newey and West (1987) standard errors of the parameter estimates using one lag. Finally, report the result of a test for overidentifying restrictions (test statistic and associated p-value). Comment on your results and compare them to those in 2a.

(c) Provide results for the GMM two-step estimator. I.e., in the second step, you should use the weighting matrix $W_T = \hat{S}_T^{-1}$, where $\hat{S}_T$ denotes the (Newey-West) estimator of S_0 that you used in 2b. Also report the Newey and West (1987) standard errors of the new parameter estimates using one lag. Finally, report the result of a test for overidentifying restrictions (test statistic and associated p-value). Comment on your results and compare them to those in 2a and 2b.

Cross-Section of Expected Returns (Lecture 5)

3. (**Compulsory; individual**) - For this exercise, use the data in the Matlab file "Assignment2Ex3Data.mat" (if needed, I can provide the data in text format as well). The dataset contains monthly data on 12,141 individual US stocks from 1973/01 to 2021/12 (588 months):

`months`	588×1 vector indicating the months in the sample.
`permno`	$1 \times 12,141$ vector of stock identifiers.
`ret`	$588 \times 12,141$ matrix of monthly stock returns during month t (measured from the beginning of month t to the end of month t).
`me`	$588 \times 12,141$ matrix of market equity values of the stocks measured at the beginning of month t.
`prc`	$588 \times 12,141$ matrix of share prices of the stocks measured at the beginning of month t.
`exchcd`	$588 \times 12,141$ matrix indicating which index the stock is listed on at the beginning of month t (1=NYSE, 2=AMEX, 3=NASDAQ).

You find further matrices in the data file that contain characteristic values, that I will refer to as X throughout the problem. **Each student should work with one characteristic throughout the problem (and each student should pick a different one)**. Indicate your choice in the first column of the table in the Google sheet under the following link (you pick a characteristic by putting your name in the first column if the given characteristic is not yet picked by an other student):

`http://tinyurl.com/mf73pcd9`

I will refer to table in the above Google sheet as "*Table G*" throughout the problem.

Some notes on the data:

- Missing values are indicated by `NaN`. Some characteristics might have missing values for a large number of stocks even when the stocks are traded.
- The order of the stocks in the columns are the same across all matrices (i.e., column i refers to the same stock i in all matrices).
- All data are properly "lined up" row-wise. That is, when constructing the portfolio returns for month t, you should use the data in row t from all the matrices.
- I use actual data from US stocks, but made some selection choices. Therefore, do not use the this data for projects outside of this course.

(a) Create the return series of 5 portfolios sorted on your chosen characteristic for the period 1973/01 to 2021/12 using *equal-weighted* returns in each portfolio. The steps are the following:

i. Start in month $t = 1$ (1973/01). Determine the set of eligible stocks to select into the portfolios: for a stock to be eligible, it has to have (i) a non-missing return observation for month t, (ii) a non-missing market equity (me) observation at the beginning of month t, and (iii) a non-missing characteristic (X) observation at the beginning of month t.

ii. Sort the eligible stocks based on their $X_{i,t}$ values into 5 equally sized portfolios (i.e., the breakpoints should be the 20th, 40th, 60th, and 80th percentiles).

iii. Calculate the return on each of the 5 portfolios for month t. The return for a given portfolio is the equal-weighted average return on the constituent stocks, i.e.

$$R_{Pj,t} = \frac{1}{n(D_{j,t})} \sum_{i \in D_{j,t}} R_{i,t} \ , \tag{5}$$

where $R_{i,t}$ denotes the return on stock i in month t, $R_{Pj,t}$ is the return on quintile j ($j = 1, ..., 5$), $D_{j,t}$ denotes the set of stocks that belong to quintile j, and $n(D_{j,t})$ is the number of stocks in quintile j.

iv. Repeat steps (i) to (iii) for each month $t = 2, 3, ...$ throughout the rest of the sample until month $t = 588$ (2021/12).

You should end up with the monthly return series of the 5 characteristic-sorted portfolios starting from 1973/01 to 2021/12.

Report the average returns of the 5 sorted portfolios and comment on the patterns you see.

(b) Create the return series of 5 portfolios sorted on your chosen characteristic for the period 1973/01 to 2021/12 using *value-weighted* returns in each portfolio. That is, you should follow the same steps as in 3a, but in step (iii) use market equity-weighted returns instead of equal-weighted returns. I.e., instead of equation (5), you should use

$$R_{Pj,t} = \sum_{i \in D_{j,t}} \frac{me_{i,t}}{\sum_{i \in D_{j,t}} me_{i,t}} R_{i,t} \ , \tag{6}$$

where $me_{i,t}$ is the market equity of stock i at the beginning of month t.

Report the average returns of the 5 sorted portfolios and comment on the patterns you see (compare it with the average returns from 3a).

(c) Create the monthly return series of the following long-short portfolio:

$$R_{LS,t} = Sign \times (R_{P5,t} - R_{P1,t}) \ ,$$

where $R_{P5,t}$ ($R_{P1,t}$) is the monthly return series of the portfolio containing stocks with the highest (lowest) $X_{i,t}$ values, and $Sign$ is either +1 or -1, as indicated in *Table G*. Create two versions of the long-short portfolio, the first one by using the equal-weighted sorted portfolios and the second by using the value-weighted sorted portfolios.

Perform the time-series test of the CAPM and at least one other popular linear factor model on the long-short portfolios (both on the equal- and value-weighted long-short portfolio, separately). That is, estimate

$$R_{LS,t} = \alpha_{LS} + \beta' F_t + \varepsilon_t \ ,$$

where F_t contains the market excess return in case of the CAPM, and the market excess return with some additional factors in case of the other models.[1] Test the null hypothesis

$$\mathrm{H}_0: \ \alpha_{LS} = 0 \ .$$

Use your preferred choice of standard error estimate for $\hat{\alpha}_{LS}$.

(d) Perform the same time-series tests as in problem 3c above, using two sub-sample periods:

- **Sample period of the original paper**: use the period starting in January of year *SampleStart* and ending in December of year *SampleEnd*, where *SampleStart* and *SampleEnd* for your chosen charateristic is indicated in *Table G*.
- **Post-publication period**: use the period starting in January of year *PubYear* and ending in 2021/12, where *PubYear* for your chosen charateristic is indicated in *Table G*.

Report the results and comment on them in light of the post-publication decline in long-short returns documented by Pontiff and McLean (2016).

[1] Data on factors for various models can be found in "Assignment2Data_Gx.xlsx". The market excess return can be found on the sheet "CAPM". You also find factor returns for the Fama-French 3-factor model (on sheet "FF3"), the Fama-French 5-factor model (on sheet "FF5"), the Carhart (1997) model (on sheet "Carhart"), and the Hou, Xue, and Zhang (2015) model (on sheet "HXZ").

4. (**Optional; individual**) Walter, Weber, and Weiss (2023) study the effects of methodological choices on the results from portfolio sorting. Consider the same portfolio sorting exercise as in problems 3a and 3b with varying methodological choices inspired by Walter, Weber, and Weiss (2023). Possible variations you can consider with the data provided:

   - *Breakpoints*: creating 10 portfolios (deciles) as opposed to 5 portfolios (quintiles).
   - *NYSE breakpoints*: Using only NYSE listed stocks ($exchcd = 1$) when determining the breakpoints (e.g., 20th, 40th, ... percentiles) for portfolio sorting.
   - *Rebalancing*: in problem 3, you re-sort the stocks into portfolios based on X every month. You can consider less-frequent portfolio formation, e.g., quarterly, or annual.
   - *Price exclusion*: Exclude stocks from the portfolio sorting (i.e., from entering the portfolios) if the share price is below 1$ or 5$ at the beginning of the month when the sorting is done.
   - *Size exclusion*: Exclude stocks from the portfolio sorting if they are below the 20th percentile of the cross-sectional size (me) distribution in your sample at the beginning of the month when the sorting is done.

   You do not have to consider all the variations, and you do not have to consider all combinations of the variations.

   You should create long-short portfolio returns, similar as in problem 3c using variations in methodological choices, and do the same tests as in problems 3c and 3d for these additional long-short portfolios.

5. (**Optional; individual**) Consider the monthly data in sheet "CrossSection" in the Excel file "Assignment2Data_Gx.xlsx". It contains monthly excess returns (in excess of the risk-free rate, i.e., $R^e_{i,t} = R_{i,t} - R_{f,t}$) on $N = 10$ different US stock portfolios.

   (a) Run regressions of the excess returns on the 10 portfolios on the market excess return; that is, run the following regressions for each of the 10 portfolios:

   $$R^e_{i,t} = \alpha_i + \beta_{i,M} R^e_{M,t} + \varepsilon_{i,t} \ .$$

   Note that you can either run 10 separate OLS regressions, or jointly estimate all parameters using GMM via the moment conditions on slide 11 of Lecture 5; the two approaches should lead to the same parameter estimates (up to a certain numerical precision).

      i. Report the $\hat{\alpha}_i$ estimates and the t-statistics of the alphas using two different standard error estimates: the usual OLS standard errors (assuming homoskedasticity and no serial correlation) and Newey-West standard errors with three lags. Comment on the results.

      ii. Test the null hypothesis that the alphas are jointly zero using two different test: the Gibbons, Ross, and Shanken (1989) finite sample test and the Wald test (for the Wald test, use the Newey-West covariance matrix of the parameter estimates with three lags). Comment on the results.

      iii. Given the proxy for the market portfolio and the risk-free asset, the CAPM suggests that the alpha is zero and the risk premium of asset i is given by the product of its beta and the market risk premium. Plot actual mean excess returns ($\hat{\mathrm{E}}\left[R^e_{i,t}\right]$) against predicted mean excess returns ($\hat{\beta}_{i,M} \times \hat{\mathrm{E}}\left[R^e_{M,t}\right]$). How can you illustrate the pricing errors ($\hat{\alpha}_i$-a) on this plot? Comment on the results.

   (b) Run regressions for the 10 portfolios with several factors:

   $$R^e_{i,t} = \alpha_i + \beta_i F_t + \varepsilon_{i,t} \ ,$$

   where $\beta_i = [\beta_{i,1}, ..., \beta_{i,K}]$ is a $1 \times K$ vector of betas, $F_t = [f_{1,t}, ..., f_{K,t}]$ is a $K \times 1$ vector of factor returns (i.e., returns on zero-cost factor portfolios), and K is the number of factors. You can choose to use your preferred factor model (see footnote 1 regarding the data for the factors).

   Do the same as in parts (i) to (iii) of problem 5a. Note that predicted mean excess return for asset i from a multifactor model is $\hat{\beta}_i \hat{\mathrm{E}}\left[F_t\right]$.