MIT Sloan School of Management

Adv. Analytics of Finance — 15.457
Hui Chen — Fall 2025

Problem Set 3

Due: 2:30 PM, Tuesday, October 14

Instructions:

- Please submit your homework on Canvas. List the names of all of your team members and their IDs in your writeup. Each team only needs to submit once.
- For question(s) involving coding, you will receive full credit only when the code can be successfully executed.

1. **Interview questions.**

   (a) Describe how to estimate a linear regression model using (i) maximum likelihood, and (ii) maximum a posteriori (MAP) estimation. (Note: We did not cover how to apply the two methods for linear regression in class. You should try to follow their standard procedures and make the necessary assumptions.)

   (b) True or false: If the data are positively autocorrelated, the efficient GMM standard error that does not correct for the autocorrelation will be too small. With negative autocorrelation, the standard error will be too big.

2. **Comparing multiple strategies.** You work for the asset management department of a large bank. You have the historical returns of N portfolio managers over the same time period, x_t^n, with $n = 1, \cdots, N$ and $t = 1, \cdots, T$. You want to see whether all of the managers produce the same average returns.

   (a) State the null hypothesis of your test.

   (b) Assume that returns are independent and identically distributed over time but potentially correlated contemporaneously. Derive the estimator of the average returns for the N strategies, $\widehat{\mu}$ (this is an $N \times 1$ vector, with $\widehat{\mu}_i$ being the average return estimator for manager i). Derive also the asymptotic covariance matrix $\widehat{\Omega}$ for the estimated mean vector.

   (c) Argue that the estimator $\widehat{\mu}$ has multivariate normal distribution asymptotically.

(d) Define $\widehat{\delta}_k = \widehat{\mu}_k - \widehat{\mu}_1$, $k = 2, \cdots, N$. Argue that the asymptotic distribution of the vector $\widehat{\delta} = (\widehat{\delta}_2, ..., \widehat{\delta}_N)'$ is multivariate normal. What is the mean and var-cov matrix of this distribution under the null hypothesis? Derive the var-cov matrix of $\widehat{\delta}$ from $\widehat{\Omega}$, and explain how to estimate it directly from return data x_t^n.

(e) Denote the covariance matrix of the distribution of $\widehat{\delta}$ in (2d) by V. Argue that the test statistic

$$W = \widehat{\delta}' V^{-1} \widehat{\delta}$$

is distributed as $\chi^2(N-1)$.

(f) Using the χ^2 distribution, construct a test of the null hypothesis with 5% size. Your test should rely on the statistic W in (2e).

(g) How would your answer above change if you had ignored the fact that returns are correlated contemporaneously?

3. **Performance evaluation.** You run a large university endowment that invests with outside portfolio managers. Two final candidates have been presented to you. The PMs' track records from 2005 to 2024 are included in the spreadsheet "PM.xlsx." You might also need the data for monthly market excess returns, returns for the SMB and HML factors, and one-month Treasury bill rates from Ken French's website (https://goo.gl/xtDApU). Please answer the following questions.

(a) Summarize the performances of the two managers:

- Make scatter plots of the managers' monthly excess returns against the market excess returns.
- Make a boxplot of the two managers' monthly returns and comment on the differences. If you notice any outliers, discuss how you would treat them.
- Report the mean and standard deviation of monthly returns; compute the Sharpe ratio, information ratio, and maximum drawdown. For the information ratio, use market portfolio as the benchmark.

(b) For each manager, is there any evidence for the ability to outperform the market? Is there any evidence of ability to generate alpha (relative to the Fama-French 3-factor model)? What are their exposures to size and value factors?

$$R_{i,t}^e = \alpha_i + \beta_i R_{m,t}^e + \gamma_i R_{hml,t} + \delta_i R_{smb,t} + \epsilon_{i,t}$$

(c) Which PM is better at beating the market? Which can generate a higher Sharpe ratio? Which can generate a higher information ratio based on the 3-factor model (defined as $\alpha_i / \sigma(\epsilon_{i,t})$ from the model above)? Answer these questions through formal statistical tests.

4. **Event-driven strategy: PEAD** After learning about Bernard and Thomas (1985), you are wondering whether your own hedge fund can build a trading strategy based on the PEAD. You thought of a different way to identify stocks with positive and negative earnings surprises. Instead of using accounting data, you will sort stocks based on the **standardized excess return** on the day after the announcement (assuming that earning announcements are released after market close):

$$s_{i,t} = \frac{R_{i,t} - R_{m,t}}{\sigma_{i,t}},$$

where $R_{i,t}$ is stock i's return on day t, $R_{m,t}$ is the market return, and $\sigma_{i,t}$ is the volatility of stock i's excess return (which can be estimated using past returns, say the past 60 days).

Data: You can download the data PEAD.csv here. It contains daily return data for the 1,000 largest stocks (based on market cap at the end of 2010, the start of the sample).

The file has several columns. The column 'permno' is the firm identifier. The column 'date' is the calendar date. The field 'mktcap' is the market capitalization of the company. The column 'ret' includes equity returns of the firm and 'mkt' the corresponding market return for that date. Finally, the field 'ann' is a dummy variable that is equal to 1 on the day after the company's earnings announcement and zero otherwise.

(a) Design an event study to test the null hypothesis that "*stocks with the most positive (negative) standardized excess returns on the day after the announcement have the same average returns in the next 30 days.*" You can define "most positive (negative)" as those stocks with $s_{i,t} > k$ ($s_{i,t} < -k$) for some k (e.g., $k = 3$, but you can also experiment with the threshold).

(b) Is there any evidence that PEAD may be weakening in recent years? (For background readings, see here and here.)

(c) (**Optional**): Redo part (a) using the panel regression approach. Specifically, estimate the following model of returns:

$$\begin{aligned} R_{1,t} &= \alpha_1 + \beta_1 R_{m,t} + \delta^+ D^+_{1,t} + \delta^- D^-_{1,t} + u_{1,t} \\ R_{2,t} &= \alpha_2 + \beta_2 R_{m,t} + \delta^+ D^+_{2,t} + \delta^- D^-_{2,t} + u_{2,t} \\ &\vdots \\ R_{N,t} &= \alpha_N + \beta_N R_{m,t} + \delta^+ D^+_{N,t} + \delta^- D^-_{N,t} + u_{N,t} \end{aligned}$$

where $D^+_{i,t}$ is a dummy variable that equals 1 if stock i had its most recent earnings announcement on day u with $0 \leq t - u \leq 30$, and a standardized excess return $s_{i,u+1} > k$ on the day after the announcement, and zero otherwise; similarly, $D^-_{i,t}$ is a dummy variable that equals 1 if stock i had its most recent earnings announcement on day u with $0 \leq t - u \leq 30$, and a standardized excess return $s_{i,u+1} < -k$ on the day after the announcement, and zero otherwise. Test the null hypothesis that $\delta^+ = \delta^-$.