Statistics for MFEs – Problem Set 7

Professor Martin Lettau

Due January 31, 2024, 2:00pm PST, to be submitted via bCourses

Note: Use basic Python commands (e.g. matrix multiplication) for all questions in this problem set. Do NOT use the built-in package (e.g. statsmodels or pandas regression commands)!

Use the same data set you used in Problem Set 3 (`https://raw.githubusercontent.com/mlettau/Data/master/Statistics/Data_PG_UN.csv`). The spreadsheet includes monthly returns for the CRSP-VW index (R_m), Proctor & Gamble (R_{PG}), Unilever (R_{UL}) and a Consumer Goods index (R_{HH}). In the lecture, we discussed regressions for Proctor & Gamble using the finite sample results under normality. The problem set asks you to do the same analysis for Unilever returns.

1. Run the regression
$$R_{ULt} - R_{ft} = \alpha + \beta(R_{mt} - R_{ft}) + e_t$$
   (a) Report the coefficient estimates, the R^2 and the adjusted $\bar{R}^2$.
   (b) Construct a scatterplot of UL returns (on the y-axis) and CRSP-VW returns (on the x-axis) as well as the regression line.
   (c) Compute the variance-covariance matrix of the OLS coefficients under the assumption of homoskedasticity.
   (d) Use t-tests to test the null hypothesis that each regression coefficient is individually equal to 0.
   (e) Assess whether there is significant evidence for heteroskedasticity.
   (f) Compute standard errors and the 90%, 95% and 99% confidence intervals under the assumption of homoskedasticity.
   (g) Compute the variance-covariance matrix of the OLS coefficients under the assumption of heteroskedasticity.
   (h) Compute standard errors and the 90%, 95% and 99% confidence intervals using the White variance-covariance matrix.
   (i) Use t-tests to test the null hypothesis that each regression coefficient is individually equal to 0 under the assumption of heteroskedasticity.
   (j) Compute the AIC, BIC and Hannah-Quinn ICs.
   (k) Compute the Durbin-Watson and Breusch-Godfrey test statistics. What do these tests tell you?

(l) Use QQ plots and formal tests to check whether the errors normally distributed.

(m) Check whether the regressions exhibit multicollinearity.

(n) Check whether there is evidence against the linear model specification.

(o) Run rolling regressions with 60-month windows and plot the β coefficients along with their 95% confidence intervals. What do you learn from these regressions?

2. Run the regression

$$R_{ULt} - R_{ft} = \alpha + \beta(R_{mt} - R_{ft}) + \gamma(R_{HHt} - R_{ft}) + e_t$$

and repeat (a)-(o).

3. Based on the results in 1. and 2., evaluate and compare the two regressions above. What is you preferred model? Why? Think about alternative specification(s) aside from 1. and 2.

4. You work for a hedge fund and your boss asks you what the current market βs are for PG and UL. Using the evidence in the lecture notes and this problem set to give him/her a comprehensive and well-reasoned answer keeping in mind the your boss is extremely nitpicky and does not accept "opinions" without facts.