Ad 495.01: Sp. Tp. In Finance: Quantitative Research Methods In Corporate Finance

Problem set 1. Survey Data Analysis and Classical Linear Regression Model

Taylan Mavruk

2025

Instructions

1) The assignment should be solved in groups of **three** students, you choose your own colleague.
2) It is highly recommended that the exercises are solved using STATA.
3) Your reports should include how you achieved the results as well as **interpretation** of the results.
4) Use the format Times New Roman 12pt. and 1.5 spacing and write maximum of 15 pages including the tables and the figures.
5) **The deadline for submission is 7/7/2025 at 13:00 hours.**
6) Submit your lab work to the course webpage under Assignments. If you have problems with submission contact me at: taylan.mavruk@pt.bogazici.edu.tr **NO LATER than the deadline**. If you miss the deadline you need to complete the assignment with additional work.

For each question useful Stata codes are given. If you want to know more about how these code work, write "findit your_code" in Stata. You will then receive information on the codes and some examples of the use of codes. Copy and paste one of the examples in your do file and change the variable names to your own variable names.

Questions

Survey Data Analysis

1) Briefly explain the following concepts:
   a) What is factor analysis?
   b) What is principal component analysis?
   c) What are the differences/similarities between factor analysis and principal component analysis?

d) Eigenvalue
e) Eigenvector
f) Factor loadings
g) Uniqueness
h) Communalities
i) Rotated factor loadings
j) Varimax rotation
k) Promax rotation

Use data **Chapter_4_OC_Y2009.dta** (available at www.cambridge.org/Mavruk, Student Resources) on ownership concentration (OC) measures for firms in Sweden in year 2009. The OC measures of interest are herfindal_index largest_owner first_second one_sumtwo_four bzd5l ssd5_l ssd5_o bzd5_l sum_five_cf sum25_cf herfindahl_index_cf largest_owner_cf first_second_cf first_sumtwofour_cf. Copy and paste these measures into your do file (note that when you copy and paste you may have unnecessary gaps between the words) .

You may find the definitions of these measures in the article:
Mavruk, Taylan, Conny Overland and Stefan Sjögren, 2019, Keeping it real or keeping it simple? Ownership concentration measures compared, *European Financial Management*, https://onlinelibrary.wiley.com/doi/epdf/10.1111/eufm.12249

2) Analyze principal components, eigenvalues, and proportion of variance. Interpret your results. Run a scree plot and interpret the figure.

   Useful stata commands:
   describe, summarize, corr, pca, screeplot.

3) Compute the component loadings and focus on loadings above **+0.2** or below **-0.2**. Interpret the results. Run component loadings with i) no rotation, ii) varimax, and iii) promax. Explain and compare the results.

   Useful stata commands:
   pca, rotate, varimax, promax.
   For example: rotate, varimax comp(2) blanks(.2)

4) Predict the components and show their descriptive statistics. What do you observe? Explain briefly. What is a KMO measure of sampling adequacy? Explain. What is the KMO in your PCA analysis? Run a Pearson correlation test on the predicted components and interpret the significance, the sign, and the magnitude of the correlation coefficients. Explain how these predicted components could be used in further analysis (see Mavruk, Overland and Sjögren, 2019).

   Useful stata commands:
   predict, sum, estat kmo, pwcorr.

Classical Linear Regression Model

The CAPM explains the equilibrium state of an efficient market in which investors are acting under the assumption of homogenous expectations in accordance with Markowitz's mean-variance analysis framework. Since its introduction, the CAPM has been the dominant model to determine the market prices of equities and other securities, albeit under severe scrutiny (see the course book, Lindblom, Mavruk, and Sjögren, 2017, for a detailed review of the model and the literature).

Using regression to determine the CAPM coefficients

To estimate the co-variance and correlation between an equity and the market portfolio, the historical observed returns over a specific period are often used. The single equity's observed returns are regressed against the observed returns for the market portfolio in accordance with the formula shown in Equation 1, where the intercept alpha (α_j) indirectly reveals if the equity is over- or underpriced (abnormal return) and ($\varepsilon_{j,t}$) is the error term:

$$r_{j,t} - r_{f,t} = \alpha_j + \beta_j(r_{M,t} - r_{f,t}) + \varepsilon_{j,t}, \quad (1)$$

where $r_{j,t}$ is the return on individual stock, $r_{f,t}$ is risk-free rate of return, and $r_{M,t}$ is the return on market, OMX. The intercept alpha (α_j) and beta coefficient (β_j) are estimated using the ordinary least squares (OLS) estimation technique. The β_j is the slope of the best fitted line and indicates the extent to which movements in the single equity j are associated with movements in the overall market.

Use **Chapter_5_CLRM.dta** (available at www.cambridge.org/Mavruk, Student Resources) data to answer the following questions. Data include weekly prices on OMX 30 index, Ericson, H&M (hm), SEB, and SKF stocks and the risk-free rate; 3 months T-bill rate obtained from Riksbanken presented as a simple annual rate.

Begin with generating a weekly date variable. Stata commands (copy and paste these comments in your do file):

```
gen date=w(2004w35)+_n-1
format date %tw
tsset date
```

5) Calculate the return on the OMX index and the (arithmetic) return, $\frac{(P_t - P_{t-1})}{P_{t-1}}$, on **each** stock. Plot the returns against the time variable, date. Interpret the graphs. Construct a risk premium measure for each company and the overall market risk premium. Note that the risk-free rate is presented **annually** in the data while your analysis window is **weekly**. You need to divide the risk free rate by 52 (simple rate). If you use the return on the stocks in decimal form, then you also need to use the risk-free rate in decimal form. Remember that the units of the return variables need to match.
6) Plot these risk premiums over time and describe what you see.

   Useful stata commands:
   gen, twoway.

7) Choose only one stock (or the index) and re-calculate the return using logarithmic returns. To compare your results with the arithmetic returns run descriptive statistics on both variables and test the mean differences. What do you observe? Which one would you suggest using in regressions? Why?

   Useful stata commands:
   sum, mvtest means.

   For the rest of the questions, use only arithmetic returns.

8) Run a descriptive statistics and Pearson correlation analysis on all the return variables (not the risk premium) in the data. Explain your results.

   Useful stata commands:
   sum, pwcorr.

9) Estimate the CAPM model in equation (1) for each company. Interpret α_j , β_j, R2 and F-value.
   a) Is α_j, significantly different from zero?
   b) Is β_j significantly different from one? Why are we testing β_j against one?

   When answering these questions formulate the null and alternative hypothesis. State the test statistics, find the rejection region, show the observed value of the test statistic from Stata and state your conclusion.

   Useful stata commands:
   reg, test, outreg or outreg2.

   Note that the last two commands (outreg) will be very helpful to present your regression results in the assignment or in your thesis later on. Use **findit outreg2** to find out more about the details and presentations.

10) Choose one regression (one stock) and find the residuals from the model and plot them in a histogram. Do we have normally distributed error terms in the model? Test for normality by using a Shapiro-Wilk W test. Explain what you observe.

   Useful stata commands:
   reg, predict, resid, histogram, swilk.

11) Plot the residuals from the regression over time. Do you see any pattern in the residual plot or do you regard them as random? Calculate the correlation coefficient of the residual and lagged residual. Is the correlation coefficient significantly different from

zero? What does that imply?

Useful stata commands:
twoway, gen, pwcorr.
Example for lag residual (lag_ehat):
gen lag_ehat=l.ehat
pwcorr ehat lag_ehat, sig /*autocorrelation at lag 1*/

12) Test for Homoscedasticity of Residuals by using the same regression on the chosen stock above. Explain how the test works and interpret your results.

Useful stata commands:
rvfplot, estat hettest.

13) Check for Multicollinearity by examining the variance inflation factor. For this exercise, include the risk premium of the other stocks as independent variables in the model and run the tests. Interpret your results.

Useful stata commands:
pwcorr, reg, vif.

Note if any of the codes do not exist in your version of STATA you need to install the code. Just write for example "findit vif" in the command box and click install, which should take just a few seconds.

14) Interaction effects: In your CAPM regression include a crises dummy, D_crises, variable by assuming that crises begin in 2008 and we treat the whole year as event time (exclude it from the analysis):

Use the following codes (copy and paste these codes):

```
gen D_crises=0
replace D_crises=. if date>= tw(2008w1) & date <=tw(2008w52)
replace D_crises =1 if date >= tw(2009w1)
tab D_crises
```

Next generate an interaction term between D_crises and the market risk premium (OMX return minus the risk free rate, let's call it Rm_rf.

gen D_crises_Rm_rf= D_crises*Rm_rf

Include D_crises and the interaction term in your CAPM regression for the chosen stock and run the regression again.
Interpret your results.

GOOD LUCK!