Assignment 2: Empirical research in accounting
Deadline: Sunday May 25, 2025, 23:59

Below are the questions for Assignment 2 of the course *Research laboratory: Business Economics*. Please submit your group solutions in a PDF file. For questions 2 and 3, which require analyses in Stata, also make sure to write down all of your Stata commands in a "do-file" and separately upload this do-file with the solutions document. Also make sure to include comments in the do-file in order for the reader to understand the steps taken (but do <u>not</u> include a log file of all Stata output). Note: all uploaded documents are automatically checked for plagiarism. Duplications of other groups' code and solutions will be identified (even when the wording has been changed) and all these cases will be forwarded to the examination board.

You should be able to answer these questions based on the material presented to you in the lectures and tutorials of block 5, as well as the Stata commands explained in Chapters 2–3 of the "*Introductory Guide to Using Stata in Empirical Financial Accounting Research*" by David Veenman that is posted on Canvas.

QUESTION 1: READING AND UNDERSTANDING AN EMPIRICAL PAPER

Carefully read the paper titled "Hubris and Unethical Decision Making: The Tragedy of the Uncommon." by McManus (2018), which was published in the *Journal of Business Ethics*.[1] Answer the following questions:

  a) What is the main research question?
  b) Why is it important to have an answer to this question? (do not use more than 100 words)
  c) Provide "Libby boxes" for the paper. Note: you can ignore the role of control variables in these Libby boxes.

QUESTION 2: PANEL DATA – EFFECTS OF CHANGES IN REGULATION

The Stata file "**shockdata2025.dta**", which is posted on Canvas, is a dataset that contains panel data for a large set of stock-listed companies from several European countries over the years 2001–2022. The dataset was constructed from the Compustat Global Fundamentals Annual database, which were obtained from WRDS. Next, the dataset was modified to reflect the effects of a *hypothetical* regulation change called "Regulation Random". This hypothetical regulation change is designed to have taken place in some, but not all, of the countries at the start of the year 2012. Assume that the primary reason for this regulation change was to make companies operate more efficiently and become more profitable.[2] We can therefore test the following hypothesis:

  HYPOTHESIS 1 (null): Regulation Random did not change the profitability of companies.

The dataset consists of the following variables:
➢ *companyid*: unique identifier for each company;
➢ *countryid*: unique id for each country;
➢ *country*: name of the country;
➢ *year*: reporting year for which company variables are measured;
➢ *profit*: measure of the company's profitability for the year (randomly created);

---

[1] Link: https://doi.org/10.1007/s10551-016-3087-9
[2] The regulation and its effects are *hypothetical* in the sense that we created the effects in the data using a simulation analysis. These data can therefore not be used to make actual inferences about the performance of companies in these countries.

➢ *companysizegroup*: a categorical variable capturing the relative size of the company (group 1 reflects the smallest companies, while group 5 reflects the largest companies);
➢ *regulation*: dummy variable set equal to 1 for companies in countries that were subject to Regulation Random, and 0 for all other companies.

Required:
a) Inspect the data to determine which countries were affected by Regulation Random.
b) Estimate a linear regression with OLS using *profit* as the dependent variable and *regulation* as the explanatory variable. Make sure to let Stata compute regression standard errors that are robust to heteroscedasticity and clustering at the company-level. Present the results from this estimation and explain what we can conclude from the results. Does this analysis help you to test HYPOTHESIS 1? Why, or why not?
c) Create a graph with two lines that display the time patterns of the average (i.e., mean) profitability of companies for (a) the group of countries subject to the regulation change, and (b) the group of countries not subject to the regulation change.[3] Explain what we can infer from this graph with respect to HYPOTHESIS 1.
d) Based on the dataset, complete the following table by filling in the average of the profitability variable for each group, and by computing the difference-in-difference in average profitability. What does the result tell you about the effect of the regulation?

	Pre	Post	Difference
Regulation=1	…	…	…
Regulation=0	…	…	…
Difference	…	…	…

e) Estimate a linear regression with OLS using *profit* as the dependent variable and an interaction between the variables *regulation* and *post*, where the *post* variable is a dummy variable that is set equal to one for the years 2012–2022, and 0 for the years 2001–2011. Again let Stata calculate standard errors that are robust to heteroscedasticity and clustering at the company-level. Present the results from this regression and explain what we can conclude from the results. Does this analysis help you to test HYPOTHESIS 1? Why, or why not?
f) What is the name of the type of test you executed in questions d and e)?
g) Explain what the benefit is of using such a test, instead of simply looking at the changes in profitability for the companies that are subject to the regulation change.

QUESTION 3: ACCRUALS, CASH FLOWS, AND STOCK RETURNS

An important part of accounting research focuses on how to best measure and evaluate company performance. Performance over a period can be quantified in many different ways. For example, we can look at a company's net profit, which is the bottom-line number produced by an income statement. Alternatively, we can look at a company's cash flows from a cash flow statement. The difference between the net profit and the cash flow is what accountants call "accruals". Accruals arise because when recognizing the benefits and costs of doing business, companies can recognize these benefits before cash has been received, or recognize expenses before cash has been paid. For example, if a company sells goods or services on credit, it does not yet receive cash from its customer, but it does recognize revenues and profits by booking a receivable. This difference between the recognized profit and the company's cash flow is, thus, what we call the accruals.

---

[3] You do not have to use Stata to create this graph. You may use Excel, or any other program you prefer, and paste the graph into your solutions document.

Accounting theories predict that the company's cash flows and accruals should have a negative correlation with each other. In practice, however, there are many reasons for why this negative relation might not exist. This means that, empirically, it is useful to test the following hypothesis:

HYPOTHESIS 2 (null): There is no relation between accruals and cash flows.

We will test this relation using the dataset "**accruals.dta**", posted on Canvas. This dataset consists of the following variables:

- *gvkey*: unique company identifier variable;
- *permno*: unique company/stock identifier variable;
- *datadate*: date of the fiscal year-end;
- *conm*: company name;
- *at*: total assets of the company at year-end;
- *cfo*: cash flows as a fraction of previous-year total assets;
- *tacc*: accruals as a fraction of previous-year total assets;
- *year*: the year of the fiscal-year end date;
- *month*: the month of the fiscal-year end date;

Required:
a) Use the combined information in this dataset to estimate the following linear regression with OLS to test Hypothesis 2:

$$TACC_{it} = \alpha + \beta CFO_{it} + \varepsilon_{it}$$

Make sure to let Stata calculate regression standard errors that are robust to heteroscedasticity and clustering at the company-level. Report the regression results. What do the results tell you with respect to HYPOTHESIS 2? *Note*: Before estimating the regression, make sure to correct for outliers by "winsorizing" the variables.

b) Now estimate the same regression on two subsamples of companies with positive and negative cash flows, separately. What do you observe? What do these results tell you about HYPOTHESIS 2?

c) Now estimate an interaction regression that gives you the same insights as the separate regressions performed in the previous question. Carefully explain how to interpret the interaction regression results.

d) Finally, we examine the relation between companies' cash flows and their stock returns measured over the same year. Stock returns are defined as the relative change in the company's stock price over the year. Required:
   - Merge the "**accruals.dta**" dataset with dataset "**returns.dta**", which is also available on Canvas. The "**returns.dta**" dataset contains the following variables:
     i. *permno*: unique company/stock identifier variable;
     i. *year*: the calendar year;
     ii. *month*: the calendar month;
     iii. *ret*: the stock return measured over the yearly period that ends in the current month.
   - Create 10 portfolios of observations based on the values of companies' cash flows (portfolio 1 has the companies with the lowest cash flows, portfolio 10 has the companies with the highest cash flows). Now create a graph (e.g., in Excel or in Stata, depending on your preference) that shows the average (mean) stock returns of companies' in each of the 10 portfolios.