**Financial Econometrics**

**Project #2**

Name: \_Neil Yejjey\_

TU ID: \_TUK12930\_\_\_\_\_\_\_\_\_

Notice: please submit the completed Word document though the link on Canvas before the deadline. Please name your document in the format of “Project 2 – Your Name”.

**Question 1**

Question 1 examines how firms’ stock returns (y) are affected by different independent variables using multiple linear regression models. There are 4 independent variables, including firms’ PE ratio (denoted as pe), firms’ dividend payment (denoted as dividend), firms’ revenue amount (denoted as revenue), and firms’ industry (denoted as industry). The dataset “project2\_data\_1” is randomly created for you to practice and to understand how to do econometric analysis by using R. Each row of data represents the observation of a stock in the sample.

1. Create two dummy variables. The first dummy variable denotes the high revenue firm. If firm’s revenue is greater than or equal to 2 million, then the firm has high revenue, and its first dummy variable is equal to one. If firm’s revenue is less than 2 million, the firm has low revenue, and its first dummy variable is equal to zero. The second dummy variable denotes the materials firm. For the firm whose industry is materials, the second dummy variable is equal to one. For the firm whose industry is non-materials, the second dummy variable is equal to zero.
2. Run the multiple linear regression without dummy variables interaction (i.e. the four independent variables are x1, x2, and two newly created dummy variables). Please write the regression equation below.

Y = 0.690 + 0.03X1 – 0.181X2 + 0.016L + 0.026M+e

Where X1 denotates PE, X2 denotates dividend, L denotates the dummy variable that equals to 1 for high revenue firm, and M denotates the dummy variable that equals to 1 for material firms

1. Analyze the summary of regression results (i.e. state the estimated values of regression coefficients, interpret the use of R-squared, determine if each population regression coefficient is different from zero by using t-value or P-value. Interpret if the high revenue firms’ stock returns are different from low revenue firms, and if the materials firms’ stock returns are different from non-materials firms). Use 10% as the significance level.

The P value for X1 is 0.0158 is less than the 10% , indicating that X1 significantly affects Y for the population

The P value for X2 is almost 0, which is less than 10% indicating that X2 significantly effects Y for the population

The P value for dummy variable large is 0.0626 which is less than 10% and indicates that dummy variable large significantly effects Y for the population. High revenue firms’ stock returns are different from low revenue firms.

The P value for dummy variable materials is 0.0472 which is less than 10% and indicates that dummy variable materials significantly affect Y for the population. The stock price of firms in the material industry is different from non-materials firms

The adjusted R-squared is 0.9912 or 99.12% , indicating that the included independent variable explains 99.12% of the change in stock price.

1. Run the multiple linear regression with dummy variables interaction (i.e. the five independent variables are x1, x2, two newly created dummy variables and their interaction term). Please write the regression equation below.

Yi = 0.694 + 0.003X1 – 0.181X2 + 0.013L + 0.024M+0.004L\* M + e

1. Interpret if the dummy variables interaction exists. Use 10% as the significance level.

The P value for the interaction term of two dummy variables is 0.8445 what is less than 10% indicating that the dummy variables interaction has an effect on stock return. The interaction exists.

1. Screenshot the code and results

Text

Description automatically generated

A close-up of a document

Description automatically generated with medium confidence

**Question 2**

Question 2 examines if attending an after-school program has a causal effect on the improvement of students’ exam scores (y) using a difference-in-difference model. Students in class A, B, and C attend the after-school program, while students in class D do not. The after-school program runs in year 2018. The dataset “project2\_data\_2” is randomly created for you to practice and to understand how to do econometric analysis by using R. Each row of data represents the observation of a student in the sample.

1. Create two dummy variables. The first dummy variable is the treatment dummy variable, which indicates whether the student receives the treatment (i.e. whether the student attends the after-school program). If the student attends the program (i.e. is in class A, B, or C), then the first dummy variable is equal to one, otherwise, it is equal to zero. The second dummy variable is the post-treatment period dummy variable, which indicates whether the exam is taken in the post-program period. If the exam is taken in the post-program period (i.e., in 2019), then the second dummy variable is equal to one, otherwise, it is equal to zero.
2. Run the difference-in-difference model with control variables. Please write the regression equation below.

Y = 52.67 – 6.83d x p + 8d + 20.89p + u

1. Analyze the summary of regression results (i.e. state the estimated values of regression coefficients, interpret the use of R-squared, determine if each population regression coefficient is different from zero by using t-value or P-value) and determine if attending the after-school program has a causal effect on improving students’ exam scores. Use 10% as the significance level.

The P value for the interaction term of the two dummy variables is 0.616 which is greater than 10% indicating that there does not exist a causal relationship between attending the program and the exam score. The causality does not exist.

1. Screenshot the code and results

Text

Description automatically generated