ECON3150/4150: Introductory Econometrics – Re-take Exam Spring 2024

Be brief and to the point and always motivate your answers. All sub-questions have equal weight.

```
# Results to Question 1.
# salary                    1990 salary, thousands $
# pcsalary                  % change salary, 89-90
# sales                     1990 firm sales, millions $
# roe                       return on equity (roe), 88-90 avg
# pcroe                     % change roe, 88-90
# ros                       return on firm's stock (ros), 88-90
# manuf                     =1 if firm sector = manufacturing
# finance                   =1 if firm sector = finance
# consprod                  =1 if firm sector = consumer product
# utility                   =1 if firm sector = transport. or utilties
# lsalary                   natural log of salary
# lsales                    natural log of sales
```

```
##                  mean            SD        min          max   N
## salary   1281.1196172  1372.3453080 223.000000 14822.000000 209
## pcsalary   13.2822967    32.6339212 -61.000000   212.000000 209
## sales    6923.7932823 10633.2710884 175.199997 97649.898438 209
## roe        17.1842105     8.5185087   0.500000    56.299999 209
## pcroe      10.8004784    97.2193994 -98.900002   977.000000 209
## ros        61.8038278    68.1770516 -58.000000   418.000000 209
## manuf       0.3205742     0.4678178   0.000000     1.000000 209
## finance     0.2200957     0.4153057   0.000000     1.000000 209
## consprod    0.2870813     0.4534861   0.000000     1.000000 209
## utility     0.1722488     0.3785031   0.000000     1.000000 209
## lsalary     6.9503861     0.5663741   5.407172     9.603868 209
## lsales      8.2922648     1.0131606   5.165928    11.489144 209
```

```
##                          reg1           reg2           reg3           reg4
## Dependent Var.:       lsalary        lsalary        lsalary        lsalary
##
## lsales         0.257 (0.032)  0.244 (0.034) 0.275 (0.030)  0.257 (0.032)
## finance                       0.124 (0.081)                0.158 (0.083)
## consprod                      0.239 (0.091)                0.181 (0.096)
## utility                      -0.353 (0.071)               -0.283 (0.073)
## roe                                         0.018 (0.003)  0.011 (0.004)
## Constant        4.82 (0.271)   4.89 (0.290)  4.36 (0.260)   4.59 (0.287)
## _______________ ______________ ______________ _____________ ______________
## S.E. type       Heteros.-rob. Heterosk.-rob. Heteros.-rob. Heterosk.-rob.
## R2                    0.21082        0.33556       0.28199        0.35687
## Observations              209            209           209            209
```

```
# Results to Question 1i.
wald(reg2, keep="finance|consprod|utility", vcov="iid")
var(predict(reg2))/var(ceosal1$lsalary)
```

1. We are interested in understanding the determinants of CEO salaries using the dataset and regression models in the R-output above:

   a. Interpret the size of the coefficient on `lsales` in the 1st results column (`reg1`).

   ANSWER HINT: It's an elasticity, i.e. a 1% increase in sales is associated with a 0.257% increase in salary.

   b. Construct and interpret the 99% confidence interval for the coefficient on `lsales` in `reg1`.

   ANSWER HINT: Find z such that we put 0.5 in upper tail, ie $P(Z \leq z) = 0.995$. From the included table we see that 2.57 or 2.58 are closest. Using 2.58 we get a 99% of 0.257 $\pm$ 2.58 * 0.032= (0.17444,0.33956) (or with 2.57: 0.17476, 0.33924)

   c. Interpret the coefficient on `finance` in the 2nd column (`reg2`). Is this estimate significant at the 1% level?

   ANSWER HINT: Salaries in finance are about i) 12.4% higher (or more precisely exp(0.124) - 1=0.132, that is 13.2%), ii) than in manufacturing, iii) keeping sales constant.

   d. Sales in manufacturing are 1,7 billion USD. Use `reg2` to compute the average CEO salary in manufacturing in USD.

   ANSWER HINT: 1,7 billion is 1700 million. Predicted log(salary) in manufacturing is 0.244*log(1700)+4.89 = 6.7. A naive conversion to levels is exp(6.7) = 816.4 thousand USD (note that this is strictly speaking missing a multiplicative correction $E[exp(residual) \mid X]$ which in practice means that we are about 10% off).

   e. Suppose sales in manufacturing increase by 17 million. Use `reg2` to compute the new average CEO salary in USD. What do you need to assume for this to be a good prediction?

   ANSWER HINT: This is a 1% increase in sales, and should therefore lead to a 0.257% (ie 816.4*0.257/100 = 2.1 thousand) increase in salary. This assumes that the estimated effect is causal otherwise changes in sales do not need to change salaries, or could do so by different amounts because of omitted-variable-bias (OVB).

   f. Perform a joint two-sided F-test (assuming homoskedasticity) of the null-hypothesis that CEO salaries do not differ across sectors in the 4th results column (`reg4`). Do you reject the null hypothesis at the 5% level?

   ANSWER HINT: Testing that the sector coeffcients in `reg4` are 0 means that we need to compare it to `reg3`. use the R2 version of the F: ((0.35687 - 0.28199)/3) / ((1 - 0.35687)/(209-6)) = 7.88 which is larger than the F(3,infty) = 2.6049 critical value and we reject the null-hypothesis.

   g. Explain why the coefficient on `finance` in larger in the `reg4` compared to `reg2`.

ANSWER HINT: Because finance has lower `roe` than manufacturing (and `roe` is positively correlated with salary). To see this remember that coef `reg2` = coef `reg4` + OVB, since `reg4` adds the "omitted variable" return to equity(`roe`). The sign of OVB depends on the sign of the correlation of the omitted variable which the outcome, which is positive, and the sign of the correlation of the finance dummy (which is relative to manufacturing) with must be negative.

h. How would the coefficients in `reg4` change if both salary **and** sales were measured in single USD (instead of thousands and millions USD respectively)?

ANSWER HINT: Let the original regression be: log(salary) = a + b * log(sales) + x*c log(salary* 1000) = log(salary) + log(1000) = log(1000) + a + b * log(sales) + x*c = log(1000) - b*log(1000000) + a + b * log(sales*1000000) + x*c. And we see that the only coefficient that is affected is the intercept, which changes by 3*(1-2*b)*log(10).

i. What two numbers (rounded to 3 decimals) does the R-code above in the output (marked Question 1i.) produce?

ANSWER HINT: Fstat= ((0.33556 - 0.21082)/3) / ((1 - 0.33556)/(209-5)) = 12.76612 and r2 of reg2: 0.33556

j. You are interested in testing whether the effect of `roe` on log(salary) is decreasing as `roe` increases. What is the regression specification that you would estimate?

ANSWER HINT: The easiest way to test for a decreasing slope in `roe` is to add a quadratic term to the regression which then should be negative.

k. Write the R-code that implements the estimation proposed in 1j.

ANSWER HINT: `feols(lsalary ~ roe + I(roe^2), ceosal1)`

```
# Dataset for exercise 2:
# firm time  productivity
# A      0     105
# A      1     125
# A      2     185
# B      0      80
# B      1     105
# B      2     112.5
# C      0      97.5
# C      1     107.5
# C      2     165
```

2. Suppose we are interested in evaluating the impact of a new training program on workers' productivity. Consider the following simplified scenario where we have 3 years of data for three different firms: Firm A, which implemented a training program after period t=1, and two other firms B and C which did not implement any training program. The dataset above reports average productivity for all workers in each firm:

   a. Based on the table above, do you think we can compare the outcome for firm A in period t=2 to that of the two other firms to estimate the causal effect of the training on productivity?

      ANSWER HINT: Comparing t=2 outcomes only gives us an unbiased estimate if the other firms are a good control group. This only happens if the training program would have been randomly assigned (or implemented in firm A for reasons unrelated to the firms productivity). Period 1 productivity, for example, looks however very different in firm A compared to the other firms, so that one must conclude that this comparison is unlikely to give a causal estimate.

   b. An alternative approach is difference-in-differences (DID). Discuss the key assumption of this approach in the context of this application, and evaluate it based on the averages provided for Firms A, B, and C.

      ANSWER HINT: The DID assumption does not rely on random assignment which makes firms comparable on average. Instead it assumes that changes over time in firm A would have been the same without the training program as the changes observed in the other firms. This is the common trend assumption. While this assumption cannot be tested, we can look at trends before t=1. Firm A: 125-105=20, Firm B=105-80=25, Firm C=107.5-97.5=10. The pre-trends do not exactly line up, but firm B seems to be the best comparison.

   c. Compute the DID estimate of the training program using the data above.

      ANSWER HINT: Using firm B as the comparison firm we get: (185-125) - (112.5 - 105) = 52.5

   d. Explain how you would set up the data (sample and variable definitions) and what specification you would estimate in a regression analysis that investigates the key assumption of the DID approach.

ANSWER HINT: test for whether firms have different trends in the pre-period (time<=1). `feols(productivity ~ i(firm) + i(firm)*I(time==1), df[time <= 1])` differential time trends shows up in the interactions `i(firm)*I(time==1)`.

e. Explain how you would set up the data (sample and variable definitions) and what specification you would estimate in a regression analysis that estimates the causal effect of the training program using the DID approach.

ANSWER HINT: `feols(productivity ~ i(firm) + i(firm)*I(time==2), df[(firm=="A" | firm=="B") & time >= 1])`

Critical Values for the $F_{m,\infty}$ Distribution

```
## Rows denote degrees of freedom (m), and columns significance level (%)

##          10%     5%     1%

##
##  1 :  2.7055 3.8415 6.6349
##  2 :  2.3026 2.9957 4.6052
##  3 :  2.0838 2.6049 3.7816
##  4 :  1.9449 2.3719 3.3192
##  5 :  1.8473 2.2141 3.0173
##  6 :  1.7741 2.0986 2.8020
##  7 :  1.7167 2.0096 2.6393
##  8 :  1.6702 1.9384 2.5113
##  9 :  1.6315 1.8799 2.4073
##  10 :  1.5987 1.8307 2.3209
##  11 :  1.5705 1.7886 2.2477
##  12 :  1.5458 1.7522 2.1847
##  13 :  1.5240 1.7202 2.1299
##  14 :  1.5046 1.6918 2.0815
##  15 :  1.4871 1.6664 2.0385
##  16 :  1.4714 1.6435 2.0000
##  17 :  1.4570 1.6228 1.9652
##  18 :  1.4439 1.6038 1.9336
##  19 :  1.4318 1.5865 1.9048
##  20 :  1.4206 1.5705 1.8783
##  21 :  1.4102 1.5557 1.8539
##  22 :  1.4006 1.5420 1.8313
##  23 :  1.3916 1.5292 1.8104
##  24 :  1.3832 1.5173 1.7908
##  25 :  1.3753 1.5061 1.7726
##  26 :  1.3678 1.4956 1.7554
##  27 :  1.3608 1.4857 1.7394
##  28 :  1.3541 1.4763 1.7242
##  29 :  1.3478 1.4675 1.7099
##  30 :  1.3419 1.4591 1.6964
```

The Cumulative Standard Normal Distribution Function, $\Pr(Z \leq z)$

```
## Rows denote 1st decimal value of z, and columns 2nd decimal value of z

## So for example, P(Z ⩽ 0.22) = 0.5871

##            0      1      2      3      4      5      6      7      8      9
## 0.0 : 0.5000 0.5040 0.5080 0.5120 0.5160 0.5199 0.5239 0.5279 0.5319 0.5359
## 0.1 : 0.5398 0.5438 0.5478 0.5517 0.5557 0.5596 0.5636 0.5675 0.5714 0.5753
## 0.2 : 0.5793 0.5832 0.5871 0.5910 0.5948 0.5987 0.6026 0.6064 0.6103 0.6141
## 0.3 : 0.6179 0.6217 0.6255 0.6293 0.6331 0.6368 0.6406 0.6443 0.6480 0.6517
## 0.4 : 0.6554 0.6591 0.6628 0.6664 0.6700 0.6736 0.6772 0.6808 0.6844 0.6879
## 0.5 : 0.6915 0.6950 0.6985 0.7019 0.7054 0.7088 0.7123 0.7157 0.7190 0.7224
## 0.6 : 0.7257 0.7291 0.7324 0.7357 0.7389 0.7422 0.7454 0.7486 0.7517 0.7549
## 0.7 : 0.7580 0.7611 0.7642 0.7673 0.7704 0.7734 0.7764 0.7794 0.7823 0.7852
## 0.8 : 0.7881 0.7910 0.7939 0.7967 0.7995 0.8023 0.8051 0.8078 0.8106 0.8133
## 0.9 : 0.8159 0.8186 0.8212 0.8238 0.8264 0.8289 0.8315 0.8340 0.8365 0.8389
## 1.0 : 0.8413 0.8438 0.8461 0.8485 0.8508 0.8531 0.8554 0.8577 0.8599 0.8621
## 1.1 : 0.8643 0.8665 0.8686 0.8708 0.8729 0.8749 0.8770 0.8790 0.8810 0.8830
## 1.2 : 0.8849 0.8869 0.8888 0.8907 0.8925 0.8944 0.8962 0.8980 0.8997 0.9015
## 1.3 : 0.9032 0.9049 0.9066 0.9082 0.9099 0.9115 0.9131 0.9147 0.9162 0.9177
## 1.4 : 0.9192 0.9207 0.9222 0.9236 0.9251 0.9265 0.9279 0.9292 0.9306 0.9319
## 1.5 : 0.9332 0.9345 0.9357 0.9370 0.9382 0.9394 0.9406 0.9418 0.9429 0.9441
## 1.6 : 0.9452 0.9463 0.9474 0.9484 0.9495 0.9505 0.9515 0.9525 0.9535 0.9545
## 1.7 : 0.9554 0.9564 0.9573 0.9582 0.9591 0.9599 0.9608 0.9616 0.9625 0.9633
## 1.8 : 0.9641 0.9649 0.9656 0.9664 0.9671 0.9678 0.9686 0.9693 0.9699 0.9706
## 1.9 : 0.9713 0.9719 0.9726 0.9732 0.9738 0.9744 0.9750 0.9756 0.9761 0.9767
## 2.0 : 0.9772 0.9778 0.9783 0.9788 0.9793 0.9798 0.9803 0.9808 0.9812 0.9817
## 2.1 : 0.9821 0.9826 0.9830 0.9834 0.9838 0.9842 0.9846 0.9850 0.9854 0.9857
## 2.2 : 0.9861 0.9864 0.9868 0.9871 0.9875 0.9878 0.9881 0.9884 0.9887 0.9890
## 2.3 : 0.9893 0.9896 0.9898 0.9901 0.9904 0.9906 0.9909 0.9911 0.9913 0.9916
## 2.4 : 0.9918 0.9920 0.9922 0.9925 0.9927 0.9929 0.9931 0.9932 0.9934 0.9936
## 2.5 : 0.9938 0.9940 0.9941 0.9943 0.9945 0.9946 0.9948 0.9949 0.9951 0.9952
## 2.6 : 0.9953 0.9955 0.9956 0.9957 0.9959 0.9960 0.9961 0.9962 0.9963 0.9964
## 2.7 : 0.9965 0.9966 0.9967 0.9968 0.9969 0.9970 0.9971 0.9972 0.9973 0.9974
## 2.8 : 0.9974 0.9975 0.9976 0.9977 0.9977 0.9978 0.9979 0.9979 0.9980 0.9981
## 2.9 : 0.9981 0.9982 0.9982 0.9983 0.9984 0.9984 0.9985 0.9985 0.9986 0.9986
```