Empirical Framework

Justin Skidmore

2022-11-29

df <- ceosal1  
head(ceosal1)

salary pcsalary sales roe pcroe ros indus finance consprod utility  
1 1095 20 27595.0 14.1 106.4 191 1 0 0 0  
2 1001 32 9958.0 10.9 -30.6 13 1 0 0 0  
3 1122 9 6125.9 23.5 -16.3 14 1 0 0 0  
4 578 -9 16246.0 5.9 -25.7 -21 1 0 0 0  
5 1368 7 21783.2 13.8 -3.0 56 1 0 0 0  
6 1145 5 6021.4 20.0 1.0 55 1 0 0 0  
 lsalary lsales  
1 6.998509 10.225389  
2 6.908755 9.206132  
3 7.022868 8.720281  
4 6.359574 9.695602  
5 7.221105 9.988894  
6 7.043160 8.703075

reg <- lm(salary~ roe + ros + sales,df)  
stargazer(reg, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 salary   
-----------------------------------------------  
roe 22.003\*   
 (11.517)   
   
ros -1.105   
 (1.450)   
   
sales 0.015\*   
 (0.009)   
   
Constant 864.118\*\*\*   
 (228.400)   
   
-----------------------------------------------  
Observations 209   
R2 0.032   
Adjusted R2 0.018   
Residual Std. Error 1,360.113 (df = 205)   
F Statistic 2.253\* (df = 3; 205)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

reg2 <- lm(log(salary)~ roe + ros + sales, df)  
stargazer(reg,reg2,type = "text")

===========================================================  
 Dependent variable:   
 ----------------------------  
 salary log(salary)   
 (1) (2)   
-----------------------------------------------------------  
roe 22.003\* 0.017\*\*\*   
 (11.517) (0.004)   
   
ros -1.105 -0.001   
 (1.450) (0.001)   
   
sales 0.015\* 0.00001\*\*\*   
 (0.009) (0.00000)   
   
Constant 864.118\*\*\* 6.612\*\*\*   
 (228.400) (0.089)   
   
-----------------------------------------------------------  
Observations 209 209   
R2 0.032 0.140   
Adjusted R2 0.018 0.127   
Residual Std. Error (df = 205) 1,360.113 0.529   
F Statistic (df = 3; 205) 2.253\* 11.096\*\*\*   
===========================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

We would use log-level based on our inputes since is .032 compared to the .14 for level-level.

bptest(reg)

studentized Breusch-Pagan test  
  
data: reg  
BP = 0.82922, df = 3, p-value = 0.8425

bptest(reg2)

studentized Breusch-Pagan test  
  
data: reg2  
BP = 0.2907, df = 3, p-value = 0.9618

reg3 <- lm(lsalary~ roe + ros + sales + pcroe + finance + indus + utility, df)  
stargazer(reg3, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 lsalary   
-----------------------------------------------  
roe 0.010\*\*   
 (0.005)   
   
ros -0.001\*\*   
 (0.001)   
   
sales 0.00001\*\*\*   
 (0.00000)   
   
pcroe 0.001\*   
 (0.0004)   
   
finance -0.055   
 (0.109)   
   
indus -0.234\*\*   
 (0.097)   
   
utility -0.518\*\*\*   
 (0.121)   
   
Constant 6.940\*\*\*   
 (0.137)   
   
-----------------------------------------------  
Observations 209   
R2 0.243   
Adjusted R2 0.217   
Residual Std. Error 0.501 (df = 201)   
F Statistic 9.223\*\*\* (df = 7; 201)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

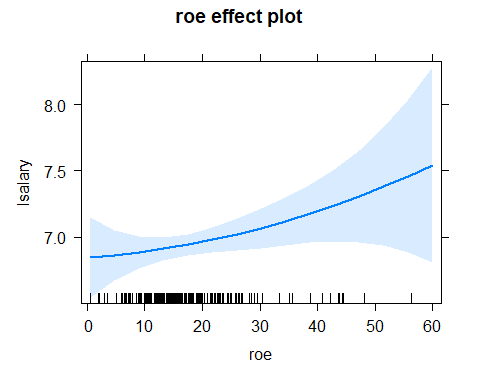
bptest(reg3)

studentized Breusch-Pagan test  
  
data: reg3  
BP = 8.906, df = 7, p-value = 0.2595

reg4 <- lm(lsalary~ roe + I(roe^2) + ros + sales + pcroe + finance + indus + utility, df)  
stargazer(reg4, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 lsalary   
-----------------------------------------------  
roe 0.003   
 (0.015)   
   
I(roe2) 0.0001   
 (0.0003)   
   
ros -0.001\*\*   
 (0.001)   
   
sales 0.00001\*\*\*   
 (0.00000)   
   
pcroe 0.001   
 (0.0004)   
   
finance -0.058   
 (0.110)   
   
indus -0.232\*\*   
 (0.097)   
   
utility -0.529\*\*\*   
 (0.123)   
   
Constant 7.009\*\*\*   
 (0.191)   
   
-----------------------------------------------  
Observations 209   
R2 0.244   
Adjusted R2 0.214   
Residual Std. Error 0.502 (df = 200)   
F Statistic 8.074\*\*\* (df = 8; 200)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

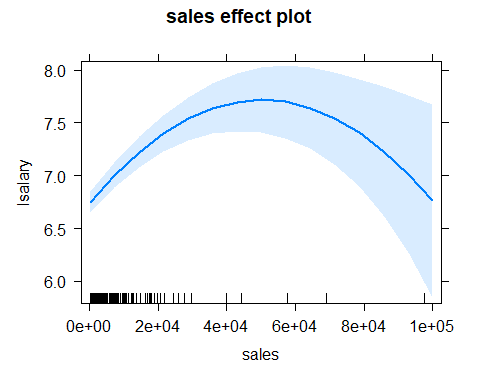
plot(effect("roe", reg4))

 We wanted to test if

reg5 <- lm(lsalary~ roe + ros + sales+ I(sales^2) + pcroe + finance + indus + utility, df)  
stargazer(reg5, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 lsalary   
-----------------------------------------------  
roe 0.011\*\*   
 (0.005)   
   
ros -0.001\*\*   
 (0.001)   
   
sales 0.00004\*\*\*   
 (0.00001)   
   
I(sales2) -0.000\*\*\*   
 (0.000)   
   
pcroe 0.0005   
 (0.0004)   
   
finance -0.054   
 (0.106)   
   
indus -0.244\*\*\*   
 (0.094)   
   
utility -0.479\*\*\*   
 (0.118)   
   
Constant 6.790\*\*\*   
 (0.138)   
   
-----------------------------------------------  
Observations 209   
R2 0.296   
Adjusted R2 0.268   
Residual Std. Error 0.485 (df = 200)   
F Statistic 10.515\*\*\* (df = 8; 200)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

plot(effect("sales", reg5))



linearHypothesis(reg5, c("roe","pcroe"))

Linear hypothesis test  
  
Hypothesis:  
roe = 0  
pcroe = 0  
  
Model 1: restricted model  
Model 2: lsalary ~ roe + ros + sales + I(sales^2) + pcroe + finance +   
 indus + utility  
  
 Res.Df RSS Df Sum of Sq F Pr(>F)   
1 202 48.772   
2 200 46.967 2 1.8046 3.8422 0.02305 \*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

We decided to keep the pcroe. It might not be independently significant but it is jointly significant with pcroe.

vif(reg5)

roe ros sales I(sales^2) pcroe finance indus   
 1.354280 1.241860 5.651509 5.469470 1.046013 1.707912 1.697045   
 utility   
 1.752789

bptest(reg5)

studentized Breusch-Pagan test  
  
data: reg5  
BP = 8.9752, df = 8, p-value = 0.3444

#THE PAPER The first assumption (MLR.1) is that our model is linear in parameters. All of our regressions are linear in parameters. The second assumption (MLR.2) is that of random sampling. We can reasonably assume that our study was conducted with random sampling. The third assumption (MLR. 3) is no perfect collinearity and although we have a mild multiplicity between a few of our variables there is no perfect collinearity. The fourth assumption (MLR. 4) is zero conditional mean. This will be true for our model. The fifth assumption of homoscedasticity will hold, we tested by the BP test for reg5 (see results). The sample size should be large enough that we don’t need assumption # 6. Because all of the assumptions hold, OLS is the Best Linear Unbiased Estimator. Our regression will use OLS. ## Models 1 & 2 We know that there are many factors that effect the CEO salary, so these first two models were to be a baseline to compare against future models, we can also determine whether to use level-level or the log-level form of the models. Model 1: Model 2: When we compare the two models, for model 1 was .032, while for model 2 it was .14. This proves that the log-level model is preferred because it can model the diminished returns. For future models we now know to us the log-level model. ##Model 3 & 4 In our Model 3 we added all the variables from ceosal1 to see what was significant and what wasn’t. To our surprise only one was not significant. We decided to take out pcroe, to our surprise again decreased. We did the LinearHypothesis test with “roe” and “pcroe,” and it turned out they were jointly significant. We decided to keep pcroe for future models. Model 3: Model 4: ## Model 5 & 6 In our Model 5 we tried to include a quadratic variable to see the effect on salary. We squared ROE but seemed to have no effect. In our Model 6 we squared sales which raised our to .296. Model 5: Model 6: