Home assignment # 2
(Suggested solutions)

1. Your are given the following model

$$y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$$

Assume that you want to test the following set of restrictions:

(a) $\beta_2 - \beta_3 = 1$

(b) $\beta_4 = \beta_6$ and $\beta_5 = 0$

Construct models that incorporate restrictions (a) and (b), separately and together. Describe what test you will use to test the restrictions, including its distribution and parameters (i.e., describe how would you test: the restriction (a), the restriction (b), and both of them together).

Solution:

- First, let us test the restrictions separately.

(a) We can express

$$\beta_2 = 1 + \beta_3$$

and plug into the unrestricted model:

$$\begin{aligned} y &= \beta_1 + (1+\beta_3)X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \\ y - X_2 &= \beta_1 + \beta_3(X_3 + X_2) + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \quad . \end{aligned}$$

We have here $J = 1$ restriction, n observations and $k = 6$ parameters. Hence, to test the restrictions, we should run the unrestricted model

$$y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$$

and the restricted model

$$y - X_2 = \beta_1 + \beta_3(X_3 + X_2) + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \quad ,$$

save SSE in both cases and test

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(n-k)} = \frac{(SSE_R - SSE_U)/1}{SSE_U/(n-6)} \sim F_{1,n-6} \quad .$$

Note that since we have only one restriction, we could also use the fact that

$$\sqrt{F} \sim t_{n-6}$$

and if n is large, also

$$\sqrt{F} \sim N(0,1) \quad .$$

(b) When we plug these restrictions into the original model, we obtain

$$y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_6 (X_4 + X_6) + \varepsilon \quad .$$

We would apply the same method as in (a) with the only difference that now we have $J = 2$ restrictions and so

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(n-k)} = \frac{(SSE_R - SSE_U)/2}{SSE_U/(n-6)} \sim F_{2,n-6} \quad .$$

- Second, let us test the restrictions together. The restricted model now becomes

$$y - X_2 = \beta_1 + \beta_3 (X_3 + X_2) + \beta_6 (X_4 + X_6) + \varepsilon$$

and $J = 3$, so we have

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(n-k)} = \frac{(SSE_R - SSE_U)/3}{SSE_U/(n-6)} \sim F_{3,n-6} \quad .$$

2. Imagine you are interested in the determinants of the revenues in shoe stores in Prague. Suppose you have specified the following model:

$$Rev_t = \alpha + \beta Inc_t + \delta Price_t + \theta Popul_t + \eta Weekend_t + \varepsilon_t \quad ,$$

where Rev_t denotes the amount of revenues in the Prague shoe stores on a particular day t, Inc_t is per capita income in Prague, $Price_t$ is a price index for shoes relative to other goods in Prague, $Popul_t$ is number of people living in Prague, and $Weekend_t$ is a dummy variable for weekend days.

(a) This specification recognizes that people might go shopping for shoes more often on weekends than on working days. Explain how would you test for such a hypothesis.

(b) What are the predicted revenues (in terms of the coefficients of the model) for weekends and for working days?

(c) Explain how would you alter the specification to account for the fact that people buy more shoes during the sales period, which is in January and July.

(d) If people have higher income, they buy more shoes on weekends (i.e., the effect of per capita income on revenues is larger on weekends compared to working days). Is this incorporated in your specification? If not, how would you do it? How would you test for the hypothesis that if people have higher income, they buy more shoes on weekends?

Solution:

(a) This specification allows for different intercept for weekends and for working days (ceteris paribus). This could reflect the fact most people go shopping on weekends, possibly making the revenues larger (so $\eta > 0$), which could be a hypothesis to be tested using a one-sided t-test:

$$H_0 : \eta \leq 0, H_A : \eta > 0$$

$$t = \frac{\widehat{\eta}}{s.d.(\widehat{\eta})} \sim t_{n-5,0.95}$$

(b) For weekend days the model looks like:

$$\hat{Rev_t} = (\hat{\alpha} + \hat{\eta}) + \hat{\beta} Inc_t + \hat{\delta} Price_t + \hat{\theta} Popul_t \quad ,$$

whereas during working days it is:

$$\hat{Rev_t} = \hat{\alpha} + \hat{\beta} Inc_t + \hat{\delta} Price_t + \hat{\theta} Popul_t \quad ,$$

(c) To recognize the effect of a sales period (January and July), we have to create a dummy variable indicating if the day in question is in Junuary or July or not:

$$Sales_t = \begin{cases} 1 & \text{if day t is in January or July} \\ 0 & \text{otherwise} \end{cases}$$

We introduce this dummy variable in our model as an additional independent variable:

$$Rev_t = \alpha + \beta Inc_t + \delta Price_t + \theta Popul_t + \eta Weekend_t + \gamma Sales_t + \varepsilon_t \quad ,$$

(d) If we want to incorporate in our specification that people buy more shoes on weekends if they have higher incomes, we need to allow for different coefficients of Inc for weekends and working days. In other words, we need to allow for both the different intercept and slope coefficient of Inc for weekends and working days. During weekends the model should look like

$$Rev_t = (\alpha + \eta) + (\beta + \omega) Inc_t + \delta Price_t + \theta Popul_t + \varepsilon_t$$

and during working days it should look like

$$Rev_t = \alpha + \beta Inc_t + \delta Price_t + \theta Popul_t + \varepsilon_t \quad .$$

We can achieve that by using the dummy variable $Weekend$ in the following way:

$$Rev_t = \alpha + (\beta + \omega Weekend_t) Inc_t + \delta Price_t + \theta Popul_t + \eta Weekend_t + \varepsilon_t \quad ,$$

so that finally we obtain

$$Rev_t = \alpha + \beta Inc_t + \omega Weekend_t Inc_t + \delta Price_t + \theta Popul_t + \eta Weekend_t + \varepsilon_t \quad ,$$

If we assume that people buy more shoes on weekends if they have higher incomes, the coefficient ω should be significant and positive, which is a hypothesis that could be tested using one-sided t-test ($H_0 : \omega \leq 0, H_A : \omega > 0$).

3. Use data *ceosal2.gdt* for this exercise. Consider an equation to explain salaries of CEOs in terms of annual firm sales:

$$\ln(salary) = \beta_0 + \beta_1 \ln(sales) + \beta_2 roe + \beta_3 neg\_ros + \varepsilon \quad ,$$

where

salary	...	CEO's salary in thousands USD
sales	...	firm's sales in millions USD
roe	...	firm's return on equity
neg_ros	...	dummy, equal to 1 if return on firm's stock is negative

(a) Define the variables you need and estimate the equation.

(b) What is the interpretation of the coefficients β_1, β_2, and β_3?

(c) Test for the presence of a significant impact of firm's sales on CEO's salary by hand (using only the estimated coefficient and the standard error from the Gretl output) and then compare your results to the results of this test in Gretl. Define the null and alternative hypothesis, the test statistic, its distribution, and interpret the results of the test.

(d) You wonder if the impact of firm's return on equity on the CEO's salary is indeed linear. You decide to test for the presence of a non-linear relationship, which you approximate by a third order polynomial of roe (i.e., $\alpha_1 roe + \alpha_2 roe^2 + \alpha_3 roe^3$).

i. Define the null and alternative hypothesis, the test statistics and its distribution. Describe all specifications you need to be able to conduct the test, construct the necessary variables, and estimate these specifications in Gretl.

ii. Calculate the test statistics by hand, compare to the critical value at 99% significance level, and interpret the results.

iii. Conduct the test in Gretl and compare the results.

Solution:

(a) We estimate the model in Gretl and obtain the following results:

```
Model 1: OLS, using observations 1-189
Dependent variable: l_salary

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const        4.49516     0.323538     13.89     1.60e-030 ***
  l_sales      0.263336    0.0371875     7.081    2.88e-011 ***
  roe          0.0168715   0.00406539    4.150    5.07e-05  ***
  neg_ros     -0.183226    0.118667     -1.544    0.1243

Mean dependent var   6.956783   S.D. dependent var   0.557778
Sum squared resid    43.63565   S.E. of regression   0.485663
R-squared            0.253962   Adjusted R-squared   0.241864
F(3, 185)            20.99220   P-value(F)           9.48e-12
Log-likelihood      -129.6544   Akaike criterion     267.3089
Schwarz criterion    280.2759   Hannan-Quinn         272.5621

Log-likelihood for salary = -1444.49

Excluding the constant, p-value was highest for variable 7 (neg_ros)
```

(b) We obtain the result $\beta_1 = 0.26$. This means that 1% increase in firm's sales increases the CEO's salary by 0.26%. $\beta_2 = 0.016$, if firm's return on equity increases by 1, the CEO's salary increases by 1.6%. $\beta_3 = -0.18$, this means that CEOs in firms with negative return on stock have by 18% lower salary than those in firms with positive return.

(c) Testing statistical significance of variable *sales* using a two-sided t-test at 95% significance level:

$$H_0 : \beta_1 = 0, H_A : \beta_1 \neq 0$$

$$t = \frac{\widehat{\beta_1}}{s.d.(\widehat{\beta_1})} \sim t_{n-k(1-\frac{\alpha}{2})}$$

$$\text{t-statistics: } t = \frac{0.2633}{0.0372} = 7.08$$

$$\text{critical value: } t_{n-k(1-\frac{\alpha}{2})} = t_{185(0.975)} = 1.960$$

$$|t| > t_{n-k(1-\frac{\alpha}{2})}$$

Therefore, we reject H_0, and confirm that the effect of sales on CEO's salary is significant. This is also confirmed by the results in the Gretl output, which show both the calculated t-statistic and the corresponding p-value, which is equal to $2.88 * 10^{-11}$. The p-value is thus smaller than the chosen significance level (0.05), we thus reject the null hypothesis and conclude that the effect is significant.

(d) i. We can test for a non-linear relationship between *roe* and *salary* with an F-test. Let's first define the unrestricted model:

$$\ln(salary) = \alpha_0 + \alpha_1 \ln(sales) + \alpha_2 roe + \alpha_3 roe^2 + \alpha_4 roe^3 + \alpha_5 neg\_ros + \varepsilon \quad ,$$

The null and alternative hypothesis are:

$$H_0 : \alpha_3 = 0 \ \& \ \alpha_4 = 0, H_A : \alpha_3 \neq 0 \vee \alpha_4 \neq 0$$

And we get the restricted model by plugging in the restrictions (under the null hypothesis). The restricted model is thus the same as the original equation:

$$\ln(salary) = \alpha_0 + \alpha_1 \ln(sales) + \alpha_2 roe + \alpha_5 neg\_ros + \varepsilon \quad ,$$

The F-statistics is as follows:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(n-k)} \sim F_{J,n-k} \quad ,$$

To be able to conduct the test, we need to run the restricted and unrestricted model and save the sum of squared residuals (SSE) from both models. We already have the results for the restricted model from part (a). Therefore, we create the variables roe^2 and roe^3, and run the unrestricted model. We obtain the following results:

```
Model 2: OLS, using observations 1-189
Dependent variable: l_salary

              coefficient    std. error    t-ratio     p-value
  -----------------------------------------------------------
  const        4.48376       0.388577       11.54      1.69e-023 ***
  l_sales      0.263537      0.0374221       7.042     3.70e-011 ***
  roe          0.0172534     0.0313685       0.5500    0.5830
  roe_2        3.46366e-05   0.00138242      0.02505   0.9800
  roe_3       -9.80521e-07   1.76649e-05    -0.05551   0.9558
  neg_ros     -0.182175      0.119645       -1.523     0.1296

Mean dependent var   6.956783   S.D. dependent var   0.557778
Sum squared resid    43.62930   S.E. of regression   0.488274
R-squared            0.254071   Adjusted R-squared   0.233690
F(5, 183)            12.46630   P-value(F)           2.03e-10
Log-likelihood      -129.6407   Akaike criterion     271.2813
Schwarz criterion    290.7318   Hannan-Quinn         279.1612

Log-likelihood for salary = -1444.47

Excluding the constant, p-value was highest for variable 8 (roe_2)
```

ii. We can now plug the SSE_R and SSE_U to the test statistic (number of restrictions $J = 2$, number of observations $n = 189$, and number of coefficients in the unrestricted model $k = 6$):

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(n-k)} = \frac{(43.63565 - 43.62930)/2}{43.62930/(189-6)} = 0.0133 \quad ,$$

Next, we compare the calculated statistics to the critical value at 99% confidence level:

$$F_{J,n-k(1-\alpha)} = F_{2,183(0.99)} = 4.61$$

$$F < F_{J,n-k(1-\alpha)}$$

Therefore, we do not reject the null hypothesis, the second and third order polynomials in *roe* are not jointly significant, and we thus conclude that there is no evidence for the presence of a non-linear relationship in *roe*.

iii. We obtain the following results of the test of linear restrictions in Gretl:

```
Restriction set
 1: b[roe_2] = 0
 2: b[roe_3] = 0

Test statistic: F(2, 183) = 0.0133305, with p-value = 0.986759

Restricted estimates:

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const        4.49516      0.323538     13.89      1.60e-030 ***
  l_sales      0.263336     0.0371875     7.081     2.88e-011 ***
  roe          0.0168715    0.00406539    4.150     5.07e-05  ***
  roe_2        0.000000     0.000000     NA        NA
  roe_3        0.000000     0.000000     NA        NA
  neg_ros     -0.183226     0.118667     -1.544     0.1243

  Standard error of the regression = 0.485663
```

Results from the F-test in Gretl show both the calculated F-statistic (the same value as the one calculated by hand) and the corresponding p-value. The p-value is equal to 0.987, which bigger than the chosen significance level (0.01). Therefore, we do not reject the null hypothesis and conclude that the variables are not jointly significant.