4th set assignments Introductory Econometrics

Task 1

Confidence Intervals:

Suppose you have estimated a parameter vector $\mathbf{b} = (\ 0.55 \quad 0.37 \quad 1.46 \quad 0.01\ )'$ with an estimated variance-covariance matrix

$$\widehat{Var(\mathbf{b}|\mathbf{X})} = \begin{bmatrix} 0.01 & 0.023 & 0.0017 & 0.0005 \\ 0.023 & 0.0025 & 0.015 & 0.0097 \\ 0.0017 & 0.015 & 0.64 & 0.0006 \\ 0.0005 & 0.0097 & 0.0006 & 0.001 \end{bmatrix}$$

a) Compute the 95% confidence interval each parameter b_k.

b) What does the specific confidence interval computed in a) tell you?

c) Why are the bounds of a confidence interval for β_k random variables?

d) Another estimation yields an estimated b_k with the corresponding standard error $se(b_k)$. You conclude from computing the t-statistic $t_k = \frac{b_k - \bar{\beta}_k}{se(b_k)}$ that you can reject the null hypothesis $H_0 : b_k = \bar{\beta}_k$ on the α% significance level. Now, you compute the $(1-\alpha)$% confidence interval. Will $\bar{\beta}_k$ lie inside or outside the confidence interval?

Task 2

More about confidence intervals:

Suppose, computing the lower bound of the 95% confidence interval yields $b_k - t_{\alpha/2}(n-K)se(b_k) = -0.01$. The upper bound is $b_k + t_{\alpha/2}(n-K)se(b_k) = 0.01$ Which of the following statements are correct?

1. With probability of 5% the true parameter β_k lies in the interval -0.01 and 0.01.
2. The null hypothesis $H_0 : \beta_k = \bar{\beta}_k$ cannot be rejected for values $(-0.01 \leq \bar{\beta}_k \leq 0.01)$ on the 5% significance level.
3. The null hypothesis $H_0 : \beta_k = 1$ can be rejected on the 5% significance level.
4. The true parameter β_k is with probability $1 - \alpha = 0.95$ greater than -0.01 and smaller than 0.01.
5. The stochastic bounds of the $1 - \alpha$ confidence interval overlap the true parameter with probability $1 - \alpha$.
6. If the hypothesized parameter value $\bar{\beta}_k$ falls within the range of the $1 - \alpha$ confidence interval computed from the estimates b_k and $se(b_k)$ then we do not reject $H_0 : \beta_k = \bar{\beta}_k$ at the significance level of 5%.

Task 3

Goodness of fit:

a) Show that if the regression includes a constant:

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i$$

then the variance of the dependent variable can be written as:

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2 + \frac{1}{N}\sum_{i=1}^{N} e_i^2$$

Hint: $\bar{y} = \bar{\hat{y}}$

b) Take your result from a) and formulate an expression for the coefficient of determination R^2.

c) Suppose, you estimated a regression with an $R^2 = 0.63$. Interpret this value.

d) Suppose, you estimate the same model as in c) without a constant. You know that you cannot compute a meaningful centered R^2. Therefore, you compute the uncentered R^2_{uc}:

$$R^2_{uc} = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}}}{\mathbf{y}'\mathbf{y}} = 0.84$$

Compare the two goodness of fit measures in c) and d). Would you conclude that the constant can be excluded because $R^2_{uc} > R^2$?

Task 4

Regression with EViews:

In a hedonic price model the price of an asset is explained with its characteristics. In the following we assume that housing pricing can be explained by its size *sqrft* (measured as *square feet*), the number of bedrooms *bdrms* and the size of the lot *lotsize* (also measured as *square feet*. Therefore, we estimate the following equation with OLS:

$$log(price) = \beta_0 + \beta_1 log(sqrft) + \beta_2 bdrms + \beta_3 \log(lotsize)$$

Results of the estimation can be found in the following table:

Dep. Variable: LPRICE
Incl. observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.29704	0.65128	-1.99152	0.0497
LSQRFT	0.70023	—	7.54031	0.0000
BDRMS	0.03696	0.02753	—	—
LLOTSIZE	0.16797	0.03828	4.38771	0.0000
R-squared	0.64297	Mean dependent var		5.6332
Adjusted R-squared	0.63021	S.D. dependent var		0.3036
S.E. of regression	0.18460	Akaike info criterion		-0.4968
Sum squared resid	2.86256	Schwarz criterion		-0.3842
Log likelihood	25.86066	F-statistic		—
Durbin-Watson stat	2.08900	Prob(F-statistic)		0.0000

(a) Interpret the estimated coefficients $\hat{\beta}_1$ und $\hat{\beta}_2$.

(b) Compute the missing values for *Std. Error* and *t-Statistic* in the table and comment on the statistical significance of the estimated coefficients (H_0: $\beta_j = 0$ vs. H_1: $\beta_j \neq 0$, $j = 0, 1, 2, 3$).

(c) Test the null hypothesis H_0: $\beta_1 = 1$ vs. H_1: $\beta_1 < 0$.

(d) Estimate the p-value for $\hat{\beta}_2$ as close as possible and interpret.

(e) What is the null hypothesis of this specific $F - Statistic$? Compute the missing value and interpret the result.

(f) Interpret the value of *R-squared*.

(g) An alternative specification of the model that excludes the lot size as an explanatory variable provides you with values for the Akaike information criterion of -0.313 and a Schwartz criterion of -0.229. Which specification would you prefer?