4th set assignments Introductory Econometrics

Task 1

Confidence Intervals:

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

\[
\begin{bmatrix}
0.01 & 0.023 & 0.0017 & 0.0005 \\
0.023 & 0.0025 & 0.015 & 0.0097 \\
0.0017 & 0.015 & 0.64 & 0.0006 \\
0.0005 & 0.0097 & 0.0006 & 0.001
\end{bmatrix}
\]

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 \( \beta_k \) 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 \( \alpha \)% 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 \( \beta_k \) 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 \( \beta_k \) 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:

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

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 \( R^2_{uc} \):

\[
R^2_{uc} = \frac{\hat{Y}' \hat{Y}}{Y'Y} = 0.84
\]

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 \( \text{sqrft} \) (measured as square feet), the number of bedrooms \( \text{bdrms} \) and the size of the lot \( \text{lotsize} \) (also measured as square feet). Therefore, we estimate the following equation with OLS:

\[
\log(\text{price}) = \beta_0 + \beta_1 \log(\text{sqrft}) + \beta_2 \text{bdrms} + \beta_3 \log(\text{lotsize})
\]

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-1.29704</td>
<td>0.65128</td>
<td>-1.99152</td>
<td>0.0497</td>
</tr>
<tr>
<td>LSQRFT</td>
<td>0.70023</td>
<td>—</td>
<td>7.54031</td>
<td>0.0000</td>
</tr>
<tr>
<td>BDRMS</td>
<td>0.03696</td>
<td>0.02753</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LLOTSIZE</td>
<td>0.16797</td>
<td>0.03828</td>
<td>4.38771</td>
<td>0.0000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.64297</td>
<td>Mean dependent var</td>
<td>5.6332</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.63021</td>
<td>S.D. dependent var</td>
<td>0.3036</td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>0.18460</td>
<td>Akaike info criterion</td>
<td>-0.4968</td>
<td></td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>2.86256</td>
<td>Schwarz criterion</td>
<td>-0.3842</td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>25.860666</td>
<td>F-statistic</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson stat</td>
<td>2.08900</td>
<td>Prob(F-statistic)</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

(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 \( F - \text{Statistic} \)? Compute the missing value and interpret the result.

(f) Interpret the value of \( R\text{-squared} \).

(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?