Introductory Econometrics (MA) - Problem Set 1

1. Conditional expectation function (CEF) & best linear prediction (BLP)

a) Show that the CEF is the best linear predition in terms of the mean squared error (MSE).
Hint: Assume an arbitrary function $g(X)$ and show that $\mathbb{E}\left([y - g(X)]^2\right)$ is minimized if $g(X) = \mathbb{E}(y|X)$.

b) Show that the OLS estimator for β_0 and β_1 in the simple linear regression model minimizes the MSE if the true model is linear.

2. CEF & Linear Regression

Load the dataset `HousePrices` from package `AER` and use it to solve the following exercises.

a) Estimate the effect of the number of bedrooms on house prices using OLS (level-level specification). Do the estimation both manually and using the function `lm()` implemented in `R`!

b) Construct a plot of the OLS regression line and the corresponding residuals.

c) Verify the numerical properties of the residuals in an OLS regression. What happens if you remove the intercept?

d) Add *lotsize* as an additional regressor. Use the regression anatomy formula to compute the OLS estimate of the effect of *lotsize* on house prices and compare it to the results obtained when using `lm()` to estimate the multivariate linear regression model.

3. Mean & Variance of OLS

a) Show that the OLS estimator is unbiased. Which assumptions are required for the unbiasedness of OLS to hold?

b) Derive the variance-covariance matrix of OLS under the assumption of homoskedasticity.

c) Derive the variance-covariance matrix of OLS under the assumption of heteroskedasticity.