Revision 1

Problem 1. Consider the model:

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

The estimated model is:

$$\widehat{\log(price)} = 11.67 + 0.000379\ sqrft + 0.0289\ bdrms \quad (1)$$

$$(0.10) \quad (0.000043) \qquad (0.0296) \quad (2)$$

$$n = 88, \quad R^2 = 0.588$$

(a) What is the percentage change in price when a 150-foot bedroom is added to the house? Denote this percentage change by $\hat{\theta}100\%$.

$\hat{\theta} = 0.000379 \times 150 + 0.0289 = 0.0858$

(b) What regression do you have to run to obtain directly a standard error for $\hat{\theta}$?

We have $\theta = 150\beta_1 + beta_2$. Replace for $\beta_2 = \theta - 150\beta_1$ and we end up with the regression:

$$\log(price) = \beta_0 + \beta_1(sqrft - 150bdrms) + \theta\, bdrms + u$$

(c) What is the predicted average house price when $sqrft = 300$ and $bdrms = 5$?

$11.67 + 0.000379 \times 3 + 0.0289 \times 5$

(d) You are worried about the functional form of your model and would like to investigate the possibility of having quadratic terms in your model. Therefore you conduct a RESET test. The R^2 from the restricted regression is 0.60, and the R^2 from the unrestricted regression is 0.72.

i. Write the restricted model and the unrestricted model.

ii. Write the null and alternative hypotheses.

iii. Compute the relevant statistic for the RESET test and decide whether you reject or fail to reject the null hypothesis at 5% significance level.

See slide 26 from Lecture 5.

Problem 2. Consider the simple linear regression model:

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

where u_t is a white noise with mean zero and variance σ_u^2. The OLS estimator can be written as

$$\hat{\beta}_1 = \beta_1 + SST_x^{-1} \sum_{t=1}^{T} x_t u_t,$$

where $SST_x = \sum_{t=1}^{T} x_t^2$.

(a) Compute $Var(\hat{\beta}_1)$ conditional on the regressors $x_1, \ldots, x_T$.

Because $Cov(u_t, u_s) = 0$ for $s \neq t$ (u_t is a white noise which means it is uncorrelated with u_s), we have conditional on $x_1, \ldots, x_T$:

$$Var(\hat{\beta}_1) = Var\left(\frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2}\right) = \frac{Var(\sum_{t=1}^{T} x_t u_t)}{(\sum_{t=1}^{T} x_t^2)^2} \tag{3}$$

$$= \frac{\sum_{t=1}^{T} x_t^2 Var(u_t)}{(\sum_{t=1}^{T} x_t^2)^2} + \frac{2\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T} x_t x_s Cov(u_t, u_s)}{(\sum_{t=1}^{T} x_t^2)^2} \tag{4}$$

$$= \frac{\sum_{t=1}^{T} x_t^2 Var(u_t)}{(\sum_{t=1}^{T} x_t^2)^2} = \frac{\sigma_u^2 \sum_{t=1}^{T} x_t^2}{(\sum_{t=1}^{T} x_t^2)^2} = \frac{\sigma_u^2}{SST_x} \tag{5}$$

(b) Assume that:

$$u_t = \rho u_{t-1} + e_t,\ t = 1, \ldots, T-1, \quad |\rho| < 1, e_t\ i.i.d.(0,1).$$

i. What is the name of this model?

ii. Derive $Cov(u_t, u_{t+h})$. [*Hint: Use the fact that* $Cov(u_t, u_{t+h}) = Cov(u_t, u_{t-h})$].

By recursive substitution we have (assuming u_0 non-random):

$$u_t = \rho^t u_0 + \sum_{i=0}^{t} \rho^i e_{t-i} \tag{6}$$

$$u_{t+h} = \rho^{t+h} u_0 + \sum_{i=0}^{t+h} \rho^i e_{t+h-i} \tag{7}$$

So, using the fact that $E(u_t) = \rho^t u_0$ and $E(u_{t+h}) = \rho^{t+h} u_0$, we have:

$$Cov(u_t, u_{t+h}) = E[(u_t - E(u_t))(u_{t+h} - E(u_{t+h}))] \tag{8}$$

$$= E((\sum_{i=0}^{t} \rho^i e_{t-i})(\sum_{i=0}^{t+h} \rho^i e_{t+h-i})) = \sum_{i=0}^{t} \rho^{2i}\rho^h \to \frac{\rho^h}{1-\rho^2} \text{ as } t \to \infty \text{ because } |\rho| < 1$$

Above the cross terms involving the error e with different subscripts are zero because e_t is i.i.d.

iii. Compute $Var(\hat{\beta}_1)$. Briefly compare it with the variance in (a).

Because of ii. there is an extra term in the numerator of the expression for $Var(\hat{\beta}_1)$. We need:

$$Cov(u_t, u_s) = \sum_{i=0}^{\min(t,s)} \rho^{2i}\rho^{|t-s|}. \tag{9}$$

Following the same reasoning as for Problem 2(a) we have:

$$Var(\hat{\beta}_1) = \frac{\sum_{t=1}^{T} x_t^2 Var(u_t)}{(\sum_{t=1}^{T} x_t^2)^2} + \frac{2\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T} x_t x_s Cov(u_t, u_s)}{(\sum_{t=1}^{T} x_t^2)^2} \tag{10}$$

$$= \frac{\sigma_u^2}{SST_x} + \frac{2\left(\rho^{|t-s|}\sum_{i=0}^{\min(t,s)}\rho^{2i}\right)\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T} x_t x_s}{(\sum_{t=1}^{T} x_t^2)^2} \tag{11}$$

iv. What regression do you need to run to obtain the GLS estimator?

See slides 23-26 from Lecture 11 with the difference that now you only have one regressor (assumed exogenous)