Econ 410 Supplementary Proofs
Lecture 6

Goal

We wish to prove that, in the context of a simple OLS regression model, the following equation:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}$$

is an unbiased estimator of σ^2.

Proof

Note that the residual can be written as:

$$\begin{aligned} \hat{u}_i &= y_i - \hat{y}_i \\ &= (\beta_0 + \beta_1 x_i + u_i) - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \\ &= (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)x_i + u_i \end{aligned} \tag{1}$$

Summing this equation over all observations, dividing by n, and noting that the average residual is equal to zero, we obtain:

$$0 = (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)\bar{x} + \bar{u} \tag{2}$$

Subtracting equation (2) from equation (1), we get:

$$\hat{u}_i = (\beta_1 - \hat{\beta}_1)(x_i - \bar{x}) + (u_i - \bar{u}) \tag{3}$$

As an aside, note that solving for $(u_i - \bar{u})$ gives us:

$$(u_i - \bar{u}) = \hat{u}_i - (\beta_1 - \hat{\beta}_1)(x_i - \bar{x}) \tag{4}$$

Squaring equation (3):

$$\hat{u}_i^2 = (\beta_1 - \hat{\beta}_1)^2(x_i - \bar{x})^2 + (u_i - \bar{u})^2 + 2(\beta_1 - \hat{\beta}_1)(x_i - \bar{x})(u_i - \bar{u})$$

Plugging in equation (4):

$$\begin{aligned} \hat{u}_i^2 &= (\beta_1 - \hat{\beta}_1)^2(x_i - \bar{x})^2 + (u_i - \bar{u})^2 + 2(\beta_1 - \hat{\beta}_1)(x_i - \bar{x})(\hat{u}_i - (\beta_1 - \hat{\beta}_1)(x_i - \bar{x})) \\ &= (\beta_1 - \hat{\beta}_1)^2(x_i - \bar{x})^2 + (u_i - \bar{u})^2 + 2(\beta_1 - \hat{\beta}_1)(x_i - \bar{x})\hat{u}_i - 2(\beta_1 - \hat{\beta}_1)^2(x_i - \bar{x})^2 \\ &= (u_i - \bar{u})^2 + 2(\beta_1 - \hat{\beta}_1)(x_i - \bar{x})\hat{u}_i - (\beta_1 - \hat{\beta}_1)^2(x_i - \bar{x})^2 \end{aligned}$$

Now let's plug this equation for the squared residual into our proposed estimator:

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{\sum_{i=1}^n \hat{u}_i^2}{n-2} \\
&= \frac{\sum_{i=1}^n \left((u_i - \bar{u})^2 + 2(\beta_1 - \hat{\beta}_1)(x_i - \bar{x})\hat{u}_i - (\beta_1 - \hat{\beta}_1)^2 (x_i - \bar{x})^2\right)}{n-2} \\
&= \frac{\sum_{i=1}^n (u_i - \bar{u})^2 + 2(\beta_1 - \hat{\beta}_1)\sum_{i=1}^n (x_i - \bar{x})\hat{u}_i - (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2}{n-2} \\
&= \frac{\sum_{i=1}^n (u_i - \bar{u})^2 - (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2}{n-2}
\end{aligned}
$$

Now we take the expected value to see if this is an unbiased estimator:

$$
\begin{aligned}
E(\hat{\sigma}^2|x) &= E\left(\frac{\sum_{i=1}^n (u_i - \bar{u})^2 - (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2}{n-2}\middle| x_1, ..., x_n\right) \\
&= \frac{E\left[\sum_{i=1}^n (u_i - \bar{u})^2 | x_1, ..., x_n\right] - E\left[(\beta_1 - \hat{\beta}_1)^2 | x_1, ..., x_n\right] \sum_{i=1}^n (x_i - \bar{x})^2}{n-2} \\
&= \frac{(n-1)E\left[\frac{\sum_{i=1}^n (u_i - \bar{u})^2}{n-1} | x_1, ..., x_n\right] - \frac{\sigma^2}{SST_x}\sum_{i=1}^n (x_i - \bar{x})^2}{n-2} \\
&= \frac{(n-1)\sigma^2 - \frac{\sigma^2}{SST_x}\sum_{i=1}^n (x_i - \bar{x})^2}{n-2} \\
&= \frac{(n-1)\sigma^2 - \sigma^2}{n-2} \\
&= \frac{(n-2)\sigma^2}{n-2} = \sigma^2
\end{aligned}
$$

So this is indeed an unbiased estimator of σ^2!