The variance and standard deviation of the OLS Estimators

- The technique of Ordinary Least Squares guarantees that the estimated regression line is the best-fitting line that can be drawn through the data. In the sense that is has the smallest possible sum of squared residuals.
- By minimizing the sum of squared residuals we have been able to estimate the unknown parameters of the simple regression model, $\hat{\beta}_0$ and $\hat{\beta}_1$.
- However, these are estimates, since the true values are unknown. As estimates, they are random variables with statistical properties of their own. We concentrate here in finding out whether they are good estimates, meaning how precisely estimated are these parameters.

- The latter means we need a measure of the variance (or the standard deviation of the estimates): $Var(\hat{\beta}_1)$, and $Var(\hat{\beta}_0)$.

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{1}$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2 n^{-1} \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{2}$$

- We will concentrate here on the first one, the variance of the slope parameter.

- Notice that the numerator is the error variance. It makes sense that the large the error variance the larger the variance of the parameter. This means that the larger the variation in the unobservables affecting the dependent variable the the more difficult is to estimate β_1.

- The denominator shows that more variability in the independent variable is preferred. The more spread out is the sample of the x the easier is to trace out the relationship between y and the exogenous variable.

- For the purposes of constructing confidence intervals around our parameter estimate we will compute the standard deviation of the OLS estimators,

$$sd(\hat{\beta}_1) = \frac{\sigma}{\sqrt{SST_x}} \tag{3}$$

- Notice, however, that we do not know σ, so we will have to use an estimate, which is familiar to us from the measures of fit. The Standard Error of the Regression: SER.

- Remember that we start from a set of data on Y and X, the estimated regression line yields a set of fitted values, or predictions, for the actual Y_i values. These are given by
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i. \tag{4}$$
The associated errors of fit are given by the residuals
$$e_i = Y_i - \hat{Y}_i, \tag{5}$$
these residuals will be used to compute the SER.

- The n residuals we just computed constitute a data variable in itself, e. And it can be described with the tools we have learned regarding any other variable. We are going to use a statistic that uses e to compute the typical error of fit of the estimated regression.

- The **standard error of the regression** (SER) is defined by
$$SER = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} \tag{6}$$
and its value is interpreted as a measure of the typical error of fit.

- Notice that to compute the SER we first have to find the estimates of the coefficients β_0 and β_1, and then compute the fitted or predicted values of our dependent variable and then compute the n residuals.

- The units of measurement of the SER are always the same as those of the dependent variable Y, because each residual is equal to the actual value Y_i minus the fitted value, $\hat{Y}_i$.

- Now we are ready to compute an estimate of the standard deviation of the slope parameter, which we denote as $se(\hat{\beta}_1)$:

$$se(\hat{\beta}_1) = \frac{SER}{\sqrt{SST_x}} \tag{7}$$

- And this is called the standard error of $\hat{\beta}_1$. This is a measure provided by statistical packages when we use regression analysis to compute the OLS estimators, and will be the basis to be able to test different hypothesis about our economic model, and provide confidence intervals of our estimates.