

Impacts of Violation of Statistical Assumptions

Slide #1 of Mod07B

What if $E(\varepsilon) \neq 0$? This might suggest the model needs to include another regressor(s), the form of regression function is nonlinear in the regressor(s), or other reasons. Ignoring $E(\varepsilon) \neq 0$, the OLS estimator is most likely biased.

What if $E(\varepsilon) = 0$ but $Var(\varepsilon)$ is not constant? For instance, if y is a count of a certain event within a short time, the variable y can be a Poisson variable and hence its variance is equal to its mean. If $E(y | x) = \beta_0 + \beta_1 x$ is a regression model for the relationship between y and the regressor x , then $Var(y)$ and $Var(\varepsilon)$ can be proportional to $E(y | x)$. In this instance, the OLS estimator treating $Var(\varepsilon)$ as if it were constant is still unbiased for the respective regression coefficient. However, the variance of the OLS estimator treating $Var(\varepsilon)$ as if it were constant is erroneous. Consequently, T ratio statistic and F ratio statistic used in Modules 1 – 6 are erroneous. So are the thresholds for these tests.

Exercise

Construct a hypothetical numerical example to illustrate the points in last page.