15.1: OLS Error Assumptions Revisited

As described in earlier chapters, there is a set of key assumptions that must be met to justify the use of the tt and FF distributions in the interpretation of OLS model results. In particular, these assumptions are necessary for hypotheses tests and the generation of confidence intervals. When met, the assumptions make OLS more efficient than any other unbiased estimator.

> ***OLS Assumptions***
>
> *Systematic Component*
>
> - *Linearity*
> - *Fixed XX*
>
> *Stochastic Component*
>
> - *Errors have constant variance across the range of XX*
>
> $E(\epsilon 2i)=\sigma 2\epsilon E(\epsilon i2)=\sigma\epsilon 2$
>
> - *Errors are independent of XX and other* $\epsilon i\epsilon i$
>
> $E(\epsilon i)\equiv E(\epsilon|xi)=0E(\epsilon i)\equiv E(\epsilon|xi)=0$
>
> *and*
>
> $E(\epsilon i)\neq E(\epsilon j)E(\epsilon i)\neq E(\epsilon j)$ *for* $i\neq ji\neq j$
>
> - *Errors are normally distributed*
>
> $\epsilon i\sim N(0,\sigma 2\epsilon)\epsilon i\sim N(0,\sigma\epsilon 2)$

There is an additional set of assumptions needed for "correct" model specification. An ideal model OLS would have the following characteristics: - YY is a linear function of modeled XX variables - No XX's are omitted that affect E(Y)E(Y) and that are correlated with included XX's. Note that exclusion of other XXs that are related to YY, but are not related to the XXs in the model, does not critically undermine the model estimates. However, it does reduce the overall ability to explain YY. All XX's in the model affect E(Y)E(Y).

Note that if we omit an XX that is related to YY and other XXs in the model, we will bias the estimate of the included XXs. Also consider the problem of including XXs that are related to other XXs in the model, but not related to YY. This scenario would reduce the independent variance in XX used to predict YY.

Table 15.1 summarizes the various classes of assumption failures and their implications.

Problem	Biased B	Biased SE	Invalid t/F	Hi Var
Non-linear	Yes	Yes	Yes	—
Heteroscedasticity	No	Yes	Yes	Yes
Autocorrealtion	No	Yes	Yes	Yes
Non-normal error	No	No	Yes	Yes
Multicollinearity	No	No	No	Yes
Omit relevant X	Yes	Yes	Yes	—
Irrelevant X	No	No	No	Yes
X measurement Error	Yes	Yes	Yes	—

Figure 15.1.1: Summary of OLS Assumption Failures and their Implications

When considering the assumptions, our data permit empirical tests for some assumptions, but not all. Specifically, we can check for linearity, normality of the residuals, homoscedasticity, data "outliers" and multicollinearity. However, we can't check for correlation between error and XX's, whether the mean error equals zero, and whether all the relevant XX's are included.