Basic Econometrics

Damodar N. Gujurati
Dawn C. Porter

Lecture 4: THE SIGNIFICANCE OF THE STOCHASTIC DISTURBANCE TERM

By: Zahir Mohamed Omar

Facebook Page: Zahir Mohamed (@zahirpinhani)

YouTube Channel: Zahir Mohamed

THE SIGNIFICANCE OF THE STOCHASTIC DISTURBANCE TERM

- The disturbance term u_i is a *surrogate for all those variables that are omitted* from the model but that collectively affect *Y.* Why *don't we introduce them* into the model explicitly? The reasons are many:

- 1. *Vagueness of theory: The theory, if any, determining the behavior of Y* may be, and often is, incomplete. *We might be* ignorant or unsure about the other variables affecting *Y.*

- 2. *Unavailability of data:* Lack of quantitative information about these variables, e.g., information on family wealth generally is not available.

- 3. *Core variables versus peripheral variables: Assume* that besides income X_1, *the number of children per family* X_2, *sex* X_3, *religion* X_4, *education* X_5, *and geographical region* X_6 *also affect* consumption expenditure. But the joint influence of all or some of these variables may be so small and it does not pay to introduce them into the model explicitly. One hopes that their combined effect can be treated as a random variable u_i.

CONTINUE…

- 4. *Intrinsic randomness in human behavior:* Even if we succeed in introducing all the relevant variables into the model, there is bound to be some "intrinsic" randomness in individual Y's that cannot be explained no matter how hard we try. The disturbances, the *u's,* may very well reflect this intrinsic randomness.

- 5. *Poor proxy variables:* for example, Friedman regards *permanent consumption (Y_p) as a function* of *permanent income (X_p). But since data on these variables are not directly* observable, in practice we use proxy variables, such as current consumption (*Y) and current income (X), there is the problem of errors of measurement, u may in this case then also represent the errors* of measurement.

- 6. *Principle of parsimony: we would like to* keep our regression model as simple as possible. If we can explain the behavior of *Y "substantially" with two or three explanatory variables and if* our theory is not strong enough to suggest what other variables might be included, why introduce more variables? Let *u_i represent all other variables.*

CONTINUE…

- 7. *Wrong functional form:* Often we do not know the form of the functional relationship between the regressand (dependent) and the regressors. Is consumption expenditure a linear (in variable) function of income or a nonlinear (invariable) function? If it is the former,

- *$Y_i = β_1 + B_2X_i + u_i$ is the proper functional relationship* between *Y and X, but if it is the latter,*

- *$Y_i = β_1 + β_2X_i + β_3X^2_i + u_i$* may be the correct functional form.

- In two-variable models the functional form of the relationship can often be judged from the scattergram. But in a multiple regression model, it is not easy to determine the appropriate functional form, for graphically we cannot visualize scattergrams in multiple dimensions.

- For all these reasons, the stochastic disturbances *ui assume an extremely* critical role in regression analysis, which we will see as we progress.