Basics of Econometrics

Ordinary least-squares method (OLS)

Linear regression model:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + \epsilon_t$$

where Y_t is dependent variable (explained varible), X_t and Z_t are independent variables (explanatory variables) and ϵ_t is error term. β_1 is constant term, β_2, β_3 are parameters.

Assumptions:

- Relationship between Y_t and X_t, Z_t is linear, given by equation above.
- The X_t and Z_t are nonstochasitc variables. Error term ϵ_t is independent of the X_t and therefore uncorrelated $E(X_t\,\epsilon_t) = 0$. The same for Z_t.
- No exact linear relationship exists between (two or more) independent variables; there is no *(multi)colinearity* $E(X_t\,Z_t) = 0$.
- The error has zero expected value for all variables $E(\epsilon_t) = 0$
- The error term has constant variance for all variables $E(\epsilon_t^2) = \sigma^2$, it is *homoscedastic.*
- Errors corresponding to different observations are independent and therefore uncorrelated $E(\epsilon_t\,\epsilon_s) = 0$ for all $t \neq s$
- The error term is normally distributed.

Matrix formulation:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where the vectors and matrixes has following dimension: $\mathbf{Y}$ is $N \times 1$, $\mathbf{X}$ is $N \times k$, $\boldsymbol{\beta}$ is $k \times 1$ and $\boldsymbol{\epsilon}$ is $N \times 1$, where k is number of independent variables plus 1, N is number of observations. The matrix $\mathbf{X}$ is called *design matrix.*

Estimation

Our objective is to find a vector of parameters $\hat{\boldsymbol{\beta}}$ which minimize

$$\sum_{t=1}^{N} \hat{\boldsymbol{\epsilon}}_t^2 = \hat{\boldsymbol{\epsilon}}' \hat{\boldsymbol{\epsilon}}$$

where

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \hat{\mathbf{Y}}$$

and

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

$\hat{\boldsymbol{\beta}}$ is estimated value of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\epsilon}}$ represents vector of *regressions residuals*, $\hat{\mathbf{Y}}$ represents vector of fitted values of $\mathbf{Y}$. The BLUE (Best Linear Unbiased Estimator) is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\mathbf{Y}).$$

Verification

We should test all assumptions of this method – homoscedasticity and independence of residuals, their distribution, (multi)colinearity of explanatory variables etc. but it takes lot of time and effort. For our purposes it will be sufficient to test only **statistical significance** of parameters.

Statistical test of a hypothesis associated with a regression coefficients (parameters) is based on the t-distribution. We want to test the null hypothesis $\beta = 0$ or equivalently that there is no relationship between variables X_t and Y_t. In this case the t-statistic is given by

$$t_i = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$$

Standard error of the estimated coefficients $s_{\hat{\beta}_i}$ is given by

$$s_{\hat{\beta}_i} = \sqrt{s^2 (\mathbf{X}'\mathbf{X})_{ii}^{-1}},$$

where $(\mathbf{X}'\mathbf{X})_{ii}^{-1}$ are diagonal elements of $(\mathbf{X}'\mathbf{X})^{-1}$ and s^2 is *residual variance*[1]

$$s^2 = \frac{\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}}{N-k}$$

[1] *Standard error of the regression* is $s = \sqrt{s^2}$.

We compare t-statistic with the *critical value* of t-distribution (t_{crit}). It is defined at significance level α, with $N - k$ degrees of freedom, also written as $t_{\alpha/2}(N - k)$.[2] For large samples and 5 percent significance level $t_{crit} = 1.96$.

If the t-statistic is greater, in absolute value, than the critical value t_{crit}

$$|t_i| > t_{crit}$$

we reject the null hypothesis ($\hat{\beta}_i = 0$) in favor of alternative hypothesis that the parameter is significantly different from zero.

Simply said, if this condition is fulfiled, the parameter $\hat{\beta}_i$ is statistically significant and the independent variable (X_t) has some explanatory ability in the regression equation; there exists some linear relationship between X_t and Y_t.

Futher readings

Pindyck, R. S., Rubinfeld, D. L. *Econometric Models and Economic Forecasts,* Irwin/McGraw-Hill, Fourth edition, 1998, Chapter 3 and 4, STM-126.

[2] α is devided by two, because it is two-tailed test.