Statistics

2023 Lectures
Part 12 - Introduction to Econometrics I

Institute of Economic Studies
Faculty of Social Sciences
Charles University

Goal of regression

- New setting: observed values of Y (dependent variable) are influenced by some other variable(s) x (independent or explanatory variable(s))
- Easiest setting: Let $Y_1, \ldots, Y_n$ be independent random variables with $E(Y_i) = \alpha + \beta x_i, i = 1, \ldots, n$ and $Var(Y_i) = \sigma^2$, where $(\alpha, \beta) \in \mathbb{R}^2$, $\sigma^2 > 0$ are unknown parameters, $x_1, \ldots, x_n$ are known constants with $\sum(x_i - \bar{x}) > 0$.
- typically we have observed data in the form of pairs $(Y_i, x_i), i = 1, \ldots n$, where some values can repeat
- Goal: to find (interval) estimators of α, β, σ^2 and test hypotheses about the parameters.

Simple regression model

- Consider the following model: $Y_i = \alpha + \beta x_i + \varepsilon_i$, ε_i iid with $E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2, i = 1, \ldots, n$. Setting $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, the assumptions on the moments of the disturbances are condensed to $E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2 I$.
- Obvious generalizations:

$$
\begin{aligned}
Y_i =& \beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in} + \varepsilon_i, i = 1, \ldots, n; \\
Y_i =& \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, i = 1, \ldots, n; \\
Y_i =& \frac{1}{1 + \alpha x^\beta} + \varepsilon_i, i = 1, \ldots, n; \\
&\vdots \\
Y =& X\beta + \varepsilon
\end{aligned}
$$

- The last model is a linear regression model in a matrix form (linear in parameters β_i).

Least Squares Estimators

In our model of simple linear regression the (ordinary) least squares estimators of α and β, based on minimization of

$$S(\alpha, \beta) = \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i)^2,$$

are

$$\begin{aligned} \hat{\alpha}_{OLS} &= \bar{Y} - \hat{\beta}\bar{x} \\ \hat{\beta}_{OLS} &= \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \end{aligned}$$

These estimators are unbiased and if $\frac{1}{n} \sum x_i^j \to c_j, j = 1, 2$, such that $c_2 - c_1^2 > 0$, then they are also consistent. Under normality, these estimators coincide with MLEs.
Moreover,

$$\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2.$$

The very last example

Example 104: Model of trend in GDP (in billions of CZK)

year	1966	67	68	69	70	71	72	73
t	1	2	3	4	5	6	7	8
r_t	140	146	162	171	176	187	197	209

Theoretical model: $r_t = \alpha + \beta t + \varepsilon_t$

Estimated model: $\hat{r}_t = 129.68 + 9.74t.$

MORE IN ECONOMETRICS I