Základy ekonometrie

Ordinary least-squares method (OLS) – metoda nejmenších čtverců

Lineární regresní model:

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

kde Y_t je závislá proměnná (vysvětlovaná), X_t a Z_t jsou nezávislé proměnné (vystvětlující).

Předpoklady:

- Vztah mezi Y_t a X_t, Z_t je lineární, daný výše uvedenou rovnicí
- X_t a Z_t jsou nestochastické proměnné. Rezidua ϵ_t jsou nezávislá na X_t, tedy nekorelovaná $E(X_t\,\epsilon_t) = 0$. To stejné platí pro Z_t.
- Neexistuje lineární vztah mezi (dvěma či více) nezávislými proměnnými; neexistuje *multikolinearita* $E(X_t\,Z_t) = 0$.
- Rezidua mají nulovou střední hodnotu (pro všechna pozorování) $E(\epsilon_t) = 0$
- Rezidua mají konstatní rozptyl (pro všechna pozorování) $E(\epsilon_t^2) = \sigma^2$, jsou *homoskedastická.*
- Rezidua odpovídajcí různým pozorováním jsou nezávislá, tedy nekorelovaná $E(\epsilon_t\,\epsilon_s) = 0$ pro všechna $t \neq s$
- Rezidua mají normální rozložení.

Maticový zápis:

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

kde vektory mají následující rozměry: $\mathbf{Y}$ je $N \times 1$, $\mathbf{X}$ je $N \times k$, $\boldsymbol{\beta}$ a $k \times 1$ je $\boldsymbol{\epsilon}$ je $N \times 1$, kde k je počet vysvětlujících proměnných plus 1, N je počet pozorování. Matice $\mathbf{X}$ se nazývá *matice plánu.*

Odhad

Chceme najít vektor parametrů $\hat{\boldsymbol{\beta}}$ který minimalizuje

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

kde

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

a

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

$\hat{\boldsymbol{\beta}}$ je vektor odhadnutých parametrů, $\hat{\boldsymbol{\epsilon}}$ představuje vektor reziduí, $\hat{\mathbf{Y}}$ je vektor vyrovaných hodnot a $\mathbf{Y}$ jsou skutečné hodnoty.

BLUE (Best Linear Unbiased Estimator) je pak dán:

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

Verifikace

Měli bychom testovat všechny předpoklady metody – homoskedasticitu a nezávislost reziduí, jejich rozložení, (multi)kolinearitu vysvětuljících proměnných atd. To je ale obsahem jiných kurzů a zabralo by to příliš mnoho času. Pro naše účely bude postačující testovat **statistickou významnost** parametrů.

Chceme testovat nulovou hypotézu $\beta = 0$, jinými slovy, že neexistuje vztah mezi proměnnými X_t and Y_t. (Test této hypotézy je založen na Studentově rozložení.) Testovací statistika (t-statistika) je dána vztahem

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

Směrodatná odchylka odhadnutých parametrů $s_{\hat{\beta}_i}$ je

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

kde $(\mathbf{X}'\mathbf{X})^{-1}_{ii}$ jsou diagonální prvky matice $(\mathbf{X}'\mathbf{X})^{-1}$ a s^2 je *rozptyl reziduí* [1]

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

[1] *Směrodatná odchylka reziduí* je $s = \sqrt{s^2}$.

Porovnáváme t-statistku s *kritickou hodnotou* t-rozdělení (t_{crit}). Ta je definována pro hladinu významnosti α, s $N - k$ stupni volnosti: $t_{\alpha/2}(N - k)$.[2] Pro velké vzorky dat a 5% hladinu význmanosti je: $t_{crit} = 1.96$.

Pokud je t-statistka v absolutní hodnotě větší než kritická hodnota t_{crit}

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

zamítneme nulovou hypotézu ($\hat{\beta}_i = 0$) ve prospěch alternativní hypotézy, že parametr je statisticky různý od nuly.

Zjednodušeně řečeno, pokud je tato podmínka splněna, parametr $\hat{\beta}_i$ je statisticky významný a nezávislá proměnná (X_t) má v regresní rovnici nějakou vysvětlovací schopnost; existuje nějaký lineární vztah mezi X_t a Y_t.

Reference

PINDYCK, R. S., RUBINFELD, D. L. *Econometric Models and Economic Forecasts,* Irwin/McGraw-Hill, 4ed., 1998, Kapitola 3 a 4, STM-126.

[2] α je dělena dvěmi, protože test je oboustranný.