Janusz Wywiał
Katedra Statystyki
Akademia Ekonomiczna w Katowicach

Regression estimators

Difference estimator:

$$t_{dS} = \bar{y}_S + b\left(\bar{x}_S - \bar{x}\right)$$

$$D^2(t_{dS}) = \frac{N-n}{Nn}\left(v_*(\mathrm{y}) + 2b\sqrt{v_*(\mathrm{x})v_*(\mathrm{y})}\,r(\mathrm{x},\mathrm{y}) - b^2 v_*(\mathrm{x})\right)$$

Regression estimator:

$$\mathbf{t_{BS}} = \bar{\mathrm{y}}_S - (\bar{\mathrm{x}}_S - \bar{\mathrm{x}})\mathrm{B}$$

$$\mathrm{B} = \frac{\mathrm{v_{xy}}}{\mathrm{v_x}} \qquad \mathbf{v_x > 0}$$

$$E\left(t_{BS}\right) = \bar{\mathrm{y}},$$

$$\mathbf{D^2(t_{BS})} = \frac{\mathrm{N-n}}{\mathrm{Nn}}\mathrm{v}_{*\mathrm{y}}\left(1-\rho^2_{\mathrm{xy}}\right)$$

$$e_w\left(\mathrm{t}_{BS}/\bar{\mathrm{y}}_S\right) = \mathrm{D}^2(\mathrm{t}_{BS})/\mathrm{D}^2(\bar{\mathrm{y}}_S) = 1 - \rho^2_{yx}$$

The regression coefficients estimated:

$$\tilde{\mathrm{B}} = \frac{\mathrm{v_{xyS}},}{\mathrm{v^2_{xS}}}, \ \mathbf{v_{xS} > 0}$$

$$\tilde{\mathrm{t}}_{\mathrm{BS}} = \bar{\mathrm{y}}_{\mathrm{S}} - (\bar{\mathrm{x}}_{\mathrm{S}} - \bar{\mathrm{x}})\tilde{\mathrm{B}}$$

$\tilde{t}_{BS}$ **is asymptotically unbiased.**

$$\tilde{D}^2(t_{BS}) = \frac{\mathrm{N - n}}{\mathrm{Nn}} \mathrm{v}_{*\mathrm{yS}}\left(1 - r^2_{\mathrm{xyS}}\right) + O(n^c), \qquad c < -1$$

$$\mathrm{r}_{\mathrm{xyS}} = \frac{\mathrm{v}_{*\mathrm{xyS}}}{\sqrt{\mathrm{v}_{*\mathrm{xS}}\mathrm{v}_{*\mathrm{yS}}}}$$

Podejście modelowe

Model m_1:

$Y_i = x_i\beta + \xi_i$, i=1,...,N,

$E(\xi_i)=0$, $D^2(\xi_i)=x_i\sigma^2$, $Cov(\xi_i, \xi_k)=0$.

Cel: predykcja: $G = \sum_{i \in U} Y_i$

Predyktor ilorazowy: $G_{IS} = g_x \frac{\overline{Y}_S}{\overline{x}_S}$

gdzie: $g_x = \sum_{i \in U} x_i$

G_{IS} jest m-nieobciążony, bo $E_m(G_{IS}-G)=0$.

$$\mathrm{M}_{\mathrm{m}_1} = \mathrm{E}_{\mathrm{m}_1}(G_{IS} - \mathrm{G})^2 = \frac{N(N-n)}{n} \frac{\overline{x}_U \overline{x}_{U-S}}{\overline{x}_S} \sigma^2$$

Niech:

$$\overline{x}_{S1} = \max_{s \in \mathbf{S}} \{\overline{x}_S\}$$

Celowy wybór próby s_1, czyli $P_1(s_1)=1$. Wtedy:

$$M_{P_1,m_1} = E_{P_1} E_{m_1} (G_{IS} - G)^2 \le E_P E_{m_1} (G_{IS} - G)^2$$

<u>Model m_2:</u>

$Y_i=x_i\beta+\alpha+\xi_i$, i=1,...,N,

$E(\xi_i)=0$, $D^2(\xi_i)=x_i\sigma^2$, $Cov(\xi_i,\xi_k)=0$.

Wtedy:

$$E_{m_2}(G_{IS}) = N\alpha \frac{g_x}{\bar{x}_s} + \beta g_x$$

$$\Delta_{m_2}(G_{IS}) = E_{m_2}(G_{IS} - G) = N\alpha\left(\frac{g_x}{\bar{x}_s} - 1\right)$$

Niech:

$$\bar{x}_{S2} = g_x$$

Celowy wybór próby (zrównoważonej) s_2, czyli $P_2(s_2)=1$. Wtedy:

$$E_{P_2} E_{m_2} (G_{IS} - G)^2 = 0$$

Czyli dla próby celowej s_1 predyktor G_{IS} jest optymalny w przypadku modelu m_1 (m-nieobciążony i ma minimalny błąd śr. kwadratowy).

Dla próby celowej, zrównoważonej s_2 predyktor G_{IS} jest m-nieobciążony dla modelu m_2 lecz w przypadku modelu m_1 nie ma już minimalnego błędu śr. kwadratowego.

Dla próby celowej, zrównoważonej s_2 predyktor G_{IS} jest odporny na występowanie modelu m_2 zamiast modelu m_1 w tym sensie, że zachowuje m-nieobciążoność.