Matematické Metody v Ekonometrii 4.

Multiple Regression Analysis - statistic inference

Blanka Šedivá

KMA

zimní semestr 2016/2017

Matrix notation of linear regression models

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \\ \vdots \\ \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \\ \vdots & \vdots & \ddots & \vdots \\ \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \\ \vdots \\ \\ \varepsilon_n \end{bmatrix},$$

where

- $\boldsymbol{Y} = [Y_1, Y_2, \dots, Y_n]^T$ is column vector of dependent variable;
- $\boldsymbol{X}$ matrix $n \times p$ of explanatory variables (DATA MATRIX)
- $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^T$ is column vector p of unknown parameters (coefficients);
- $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]^T$ column vector of stochastic disturbances.

Gauss Markov theorem

Suppose we have linear regression model in matrix notation

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with assumptions (P1)-(P4). Then

(i) $\boldsymbol{b} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ is the ordinary least squares estimator (OLS);

(ii) $\mathrm{E}\left(\boldsymbol{b}|\boldsymbol{X}\right) = \boldsymbol{\beta}$

(iii) $\mathbf{Var}\left(\boldsymbol{b}|\boldsymbol{X}\right) = \sigma^2\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$;

(iv) $s^2 = \dfrac{RSS}{n-p}$ is unbias estimator of σ^2, $\left(\mathrm{E}\left(s^2|\boldsymbol{X}\right) = \sigma^2\right)$
where $RSS = \left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}\right)^T\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}\right)$ is residual sum of square;

(v) the estimator $\boldsymbol{X}\boldsymbol{b}$ is BLUES of $\boldsymbol{Y}$.

https://en.wikipedia.org/wiki/Gauss-Markov_theorem

Gauss Markov theorem for normally distributed disturbances

Model $\boldsymbol{Y} \sim N\left(\boldsymbol{X}\beta, s^2\boldsymbol{I}\right)$,

(i) $\boldsymbol{b} \sim N\left(\beta; \sigma^2\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right)$;

(ii) $\dfrac{RSS}{\sigma^2} = \dfrac{s^2(n-p)}{\sigma^2} \sim \chi^2$ with $\nu = n - p$ degree of free (df)

(iii) $\boldsymbol{b}$ a RSS are independent

(iv) $\mathrm{E}\left(\boldsymbol{a}^T\boldsymbol{b}\right) = \boldsymbol{a}^T\beta$; where $\boldsymbol{a} = (a_0, a_1, \ldots, a_k)^T \neq \boldsymbol{0}$

(v) $\mathbf{Var}\left(\boldsymbol{a}^T\boldsymbol{b}\right) = s^2\boldsymbol{a}^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{a}$;

(vi) $T = \dfrac{\boldsymbol{a}^T\boldsymbol{b} - \boldsymbol{a}^T\beta}{\sqrt{s^2\boldsymbol{a}^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{a}}} \sim$ t-distribution with $\nu = n - p$ df.

Confidence intervals for model $\boldsymbol{Y} \sim N\left(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right)$

- $\boldsymbol{a}^T \boldsymbol{b} \pm t_{1-\alpha/2,(\nu=n-p)} s \cdot \sqrt{\boldsymbol{a}^T \boldsymbol{V} \boldsymbol{a}}$, where $\boldsymbol{V} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1}$
- for $\boldsymbol{a} = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$

$$b_j \pm s \cdot \sqrt{v_{jj}} \cdot t_{1-\alpha/2,(\nu=n-p)},$$

- confidence set for $\boldsymbol{\beta}$ is

$$\mathcal{K} = \left\{\boldsymbol{\beta} \in \mathbf{R}^p; (\boldsymbol{\beta} - \boldsymbol{b})^T \boldsymbol{X}^T \boldsymbol{X} (\boldsymbol{\beta} - \boldsymbol{b}) < p \cdot s^2 \cdot F_{1-\alpha/2,(\nu_1=p,\nu_2=n-p)}\right\}.$$

- confidence interval for $y = \beta_1 + \beta_1 \cdot x_1 + \cdots + \beta_k \cdot x_k$ is when $\boldsymbol{a} = (1, x_1, \ldots, x_k)$

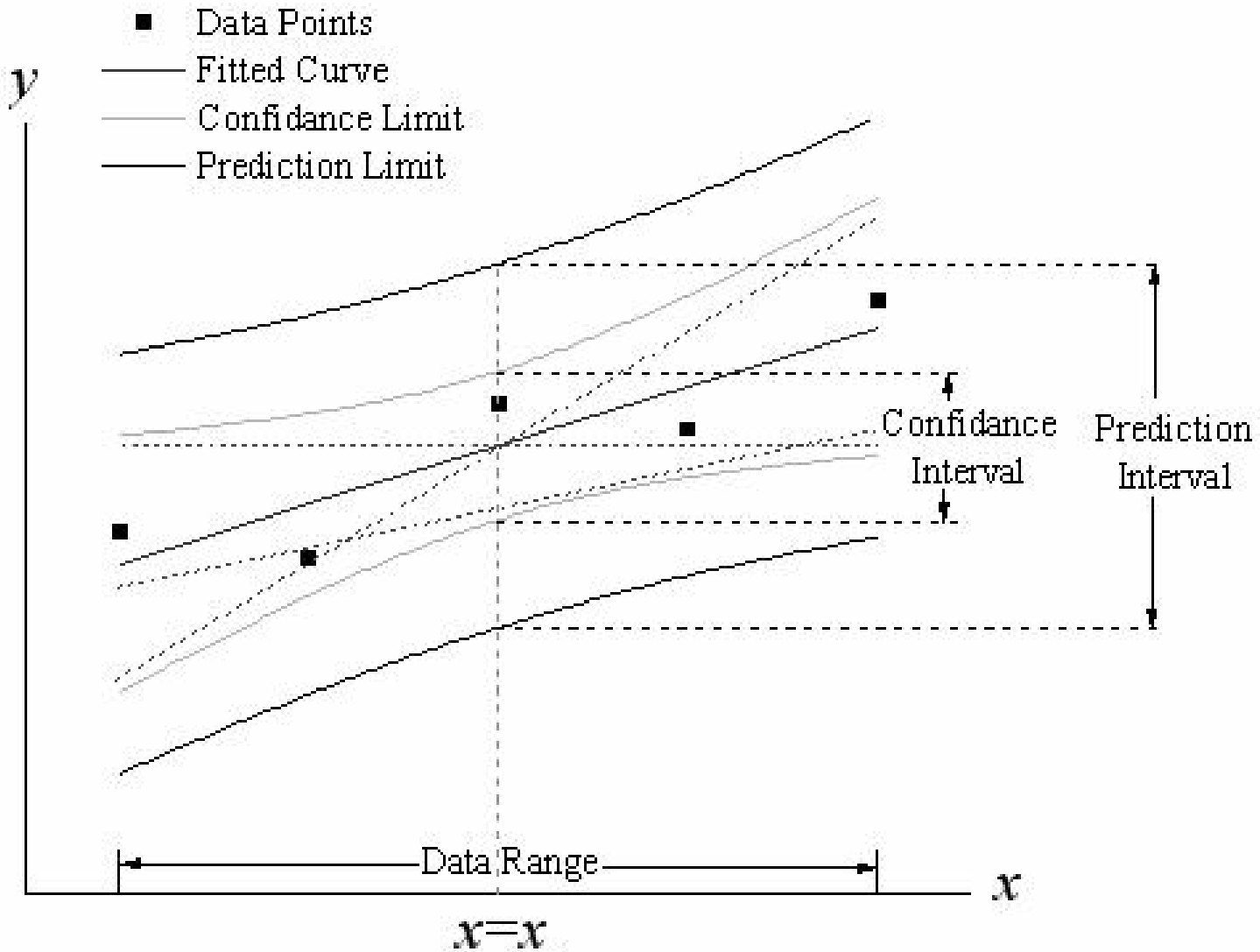
$$\boldsymbol{a}^T \boldsymbol{b} \pm t_{1-\alpha/2,(\nu=n-p)} \cdot s \cdot \sqrt{\boldsymbol{a}^T \boldsymbol{V} \boldsymbol{a}}$$

- prediction interval for specific predicted values

$$\boldsymbol{a}^T \boldsymbol{b} \pm t_{1-\alpha/2,(\nu=n-p)} \cdot s \cdot \sqrt{\boldsymbol{I} + \boldsymbol{a}^T \boldsymbol{V} \boldsymbol{a}}$$

Confidence and prediction band

$$\hat{y} \pm t_{\alpha/2}\left(\textit{Estimated standard error of } \hat{y}\right)$$



MATLAB: Confidence and prediction band

- [Y,DELTA] = polyconf(p,X,S,param1,val1,param2,val2,...)
- Y = polyconf(p,X) evaluates the polynomial p at the values in X. p is a vector of coefficients in descending powers.
- [Y,DELTA] = polyconf(p,X,S) takes outputs p and S from polyfit and generates 95% prediction intervals Y ± DELTA for new observations at the values in X.

simopt	predopt	Bounded Quantity
off	observation	$y_{n+1}(x_{n+1})$
off	curve	$p_{n+1}(x_{n+1})$
on	observation	$y(x)$ for all x
off	curve	$p(x)$ for all x

Example I

- $\boldsymbol{y} = (1, 2, 3, 4)^T$ and $\boldsymbol{x} = (-2, -1, 1, 2)^T$ matrix notation $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- $\boldsymbol{X} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ $\qquad \boldsymbol{Y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$
- $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_1 \end{bmatrix}$
- $\boldsymbol{b} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$
  - $\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 4 & 0 \\ 0 & 10 \end{bmatrix}$ $\qquad \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/10 \end{bmatrix}$
  - $\boldsymbol{X}^T\boldsymbol{y} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$

Example II

- $\boldsymbol{b} = \begin{bmatrix} 5/2 \\ 7/10 \end{bmatrix}$
- $\widehat{\boldsymbol{y}} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5/2 \\ 7/10 \end{bmatrix} = \begin{bmatrix} 11/10 \\ 18/10 \\ 32/10 \\ 39/10 \end{bmatrix}$
- residua $\boldsymbol{e} = \begin{bmatrix} 1/10 \\ -2/10 \\ 2/10 \\ -1/10 \end{bmatrix}$
- $RSS = \boldsymbol{e}^T \boldsymbol{e} = 10/100 = 1/10$
- $s_R^2 = RSS/2 = 1/20 = 0.05$
- $\boldsymbol{s}_b = \left(\sqrt{\frac{1}{20}} \sqrt{\frac{1}{4}}, \sqrt{\frac{1}{20}} \sqrt{\frac{1}{10}} \right) = (0.1118, 0.0707)$

Example III

- confidence interval $\boldsymbol{\beta}$
  - for $\alpha = 5\%$ and df $= 2$ is according tables $t_{0.975} = 4.3027$
  - $\mathcal{P}\left(2.0189 < \beta_1 < 2.9811\right) = 0.95$
  - $\mathcal{P}\left(0.3958 < \beta_2 < 1.0042\right) = 0.95$
- $H_0 : \beta_1 = 0$
  statistics $t = \frac{2.5}{\sqrt{1/80}} = 22.3607$
  the critical region for $\alpha = 5\%$ is $W = (-\infty, -4.3027) \cup (4.3027, \infty)$
  $t \in W \Rightarrow$ reject H_0
- $H_0 : \beta_2 = 0$
  $t = \frac{0.7}{\sqrt{1/200}} = 9.8995$
  the critical region for $\alpha = 5\%$ is
  $W = (-\infty, -4.3027) \cup (4.3027, \infty)$ $t \in W \Rightarrow$ reject H_0