BOX-COX transformace

- pokud chyby nemají normální rozdělení, hledáme transformaci T, která by nejenom linearizovala model, ale také transformovala chyby, aby byly přibl. normální
- jako užitečná se ukazuje násl. třída transf. (power family)

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \log y & \lambda = 0 \end{cases}$$

která předpokládá, že data y jsou pouze kladná
(pokud ne, musíme přičíst konstantu ke všem poz. a analyzovat tato posunutá data)

POZN: $\lim\limits_{\lambda \to 0} \dfrac{y^{\lambda}-1}{\lambda} = \log y$

- pro nalezení vhodného λ budeme předpokládat, že transformované veličiny $Y_i^{(\lambda)}$, $i = 1, \dots, n$, splňují podmínky LM, tj.

$$Y_i^{(\lambda)} = x_i^T \beta + e_i, \qquad e \sim N_n(0, \sigma^2 I_n)$$

$\left(Y_i^{(\lambda)} \sim N(x_i^T\beta, \sigma^2)\right.$ a $\left.Y_1^{(\lambda)}, \dots, Y_n^{(\lambda)} \text{ nez.}\right)$

úkol je odhadnout zároveň λ, β, σ^2, použijeme MLE

pomocí transformace získáme hustotu

$$f_{Y_i}(y_i) = f_{Y_i^{(\lambda)}}(y_i^{(\lambda)}) \frac{d y_i^{(\lambda)}}{d y_i} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i^{(\lambda)} - \mu_i)^2} \cdot y_i^{\lambda - 1},$$

kde $\mu_i = E Y_i^{(\lambda)} = x_i^T \beta$

věrohodnostní funkce pro poz. $y_1, \dots, y_n$ bude

$$L = \prod_{i=1}^{n} f_{Y_i}(y_i) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i^{(\lambda)} - \mu_i)^2} \cdot J(\lambda), \qquad J(\lambda) = \prod_{i=1}^{n} y_i^{\lambda-1} = \left(\prod_{i=1}^{n} y_i\right)^{\lambda - 1}$$

log-likelihood:

$$l = \ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i^{(\lambda)} - \mu_i\right)^2 + \ln J(\lambda)$$

věrohodnostní rovnice nemají explicitní analytické řešení, pro nalezení MLE si všimneme, že pro pevné λ je l proporcionální log. věrohodnosti pro odhad (β, σ^2) na základě $y^{(\lambda)} = \left(y_1^{(\lambda)}, \dots, y_n^{(\lambda)}\right)^T$, tedy

$$\hat{\beta}(\lambda) = (X^T X)^{-1} X^T y^{(\lambda)}$$

$$\hat{\sigma}^2(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i^{(\lambda)} - \hat{y}_i^{(\lambda)}\right)^2, \qquad \hat{y}_i^{(\lambda)} = x_i^T \hat{\beta}^{(\lambda)}$$

$$= \frac{1}{n}\, y^{(\lambda)T} (I_n - H)\, y^{(\lambda)}$$

dosazením do l dostaneme její hodnotu maximalizovanou vzhledem k (β, σ^2), tzv. profile log-likelihood

$$l_p(\lambda) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \hat{\sigma}^2(\lambda) - \frac{n}{2} + \ln J(\lambda) =$$

$$= C - \frac{n}{2}\ln \hat{\sigma}^2(\lambda) + (\lambda - 1)\sum_{i=1}^{n}\ln y_i$$

POZN: kvůli komplikované závislosti l_p na λ bude třeba numerické metody pro maximalizaci

lze přepsat do tvaru, kdy bude možné využít metody LR

$$l_p(\lambda) = C - \frac{n}{2}\ln \hat{\sigma}^2(\lambda) - \frac{n}{2}\ln J(\lambda) = C - \frac{n}{2}\ln \frac{\hat{\sigma}^2(\lambda)}{\left(J(\lambda)^{\frac{1}{n}}\right)^2}$$

$$\left(J(\lambda)\right)^{\frac{1}{n}} = \left[\left(\prod_{i=1}^{n} y_i\right)^{\frac{1}{n}}\right]^{\lambda-1} = (\dot{y})^{\lambda-1}$$ kde $\dot{y}$ je geometrický průměr

tz. $$l_p(\lambda) = C - \frac{n}{2}\ln \frac{\hat{\sigma}^2(\lambda)}{\left[(\dot{y})^{\lambda-1}\right]^2} = C - \frac{n}{2}\ln s_\lambda^2$$

kde $$s_\lambda^2 = \frac{\hat{\sigma}^2(\lambda)}{\left[(\dot{y})^{\lambda-1}\right]^2} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i^{(\lambda)}}{(\dot{y})^{\lambda-1}} - \frac{\hat{y}_i^{(\lambda)}}{(\dot{y})^{\lambda-1}}\right)^2$$

Pozn. s_λ^2 je reziduální součet čtverců (SSE) v modelu

$$\frac{y_i^{(\lambda)}}{(\dot{y})^{\lambda-1}} \text{ v závislosti na } \mathbb{X}_i^T$$

(Pozn. s_λ^2 lze snadno získat pomocí funkce lm( ))

celkem: max. $\ell_p(\lambda)$ $\Leftrightarrow$ min s_λ^2

algoritmus:

1) zvolit oblast hodnot λ, $I = \langle \lambda_{min}, \lambda_{max} \rangle$, a body $\lambda \in I$
   (typicky $I = \langle -2, 2 \rangle$ a 10–20 rovnoměrně rozdělených bodů)

2) nafitovat model $\frac{y^{(\lambda)}}{(\dot{y})^{\lambda-1}} \sim x$ a spočítat $\frac{1}{n} SSE = s_\lambda^2$

3) z grafu (λ, s_λ^2) vybrat $\hat{\lambda}$, které minimalizuje s_λ^2

4) pro zvolené $\hat{\lambda}$ nafitovat model $y^{(\hat{\lambda})} \sim x$ a pokračovat standardní analýzou

IS pro λ:

snadno lze odvodit LRT test pro test $H_0: \lambda = \lambda_0$

($H_0: \lambda = 1$ tedy je třeba transformace, pokud zamítneme H_0 $\Rightarrow$ transformace pomocí $\hat{\lambda}$)

LRT statistika $\Lambda = -2 \ln \frac{L(\lambda_0)}{L(\hat{\lambda})} = 2(\ell_p(\hat{\lambda}) - \ell_p(\lambda_0))$

víme $\Lambda \xrightarrow{\mathcal{L}} \chi^2(1)$

invertováním přípustné oblasti LRT testu, dostaneme as. $100(1-\alpha)\%$ IS pro λ

$\Lambda \le \chi^2_{1-\alpha}$

pokud $\hat{\lambda}$ je MLE λ, $\overset{as.}{\Rightarrow}$ $100(1-\alpha)\%$ IS pro λ je

$$2\left(\frac{n}{2} \ln s_{\lambda_0}^2 - \frac{n}{2} \ln s_{\hat{\lambda}}^2\right) \le \chi^2_{1-\alpha}$$

$$n \ln \frac{s_{\lambda_0}^2}{s_{\hat{\lambda}}^2} \le \chi^2_{1-\alpha}$$

$$\left\{ \lambda \in \mathbb{R} \;\middle|\; n \cdot \ln \frac{s_\lambda^2}{s_{\hat{\lambda}}^2} \le \chi^2_{1-\alpha}(1) \right\}$$

POZN: Kvůli jednoduchosti interpretace se často doporučuje zaokrouhlit $\hat{\lambda}$ na nejbližší $\frac{1}{4}$ nebo $\frac{1}{3}$.

Př. data TREES

Transformace vysvětlujících proměnných x

- pokud diagnostiky modelu naznačují, že vztah mezi Y a X není lineární pro jeden nebo více regresorů, může být vhodné přeformulovat model pomocí transformací proměnných X
- předp., že v modelu $Y = \beta_0 + \sum_{j=1}^{m} \beta_j x_j + e$ máme podezření na nelinearitu v j-té proměnné x_j
- jednou z možností jak postupovat je nahrazení x_j proměnnou $z_j = f(x_j)$, model tedy bude $Y = \beta_0 + \beta_1 x_1 + \dots + \beta_{j-1} x_{j-1} + \beta_j z_j + \dots + \beta_m x_m + e$
- pokud je f známá, jedná se o model LR a lze ho analyzovat standardně
  pokud je tato transformace vhodná, mělo by se to projevit ve zlepšení statistik jako je R^2, s, F a zlepšením grafů reziduí pro z_j oproti těm pro x_j
- bohužel f většinou známá není, možný přístup je parametrizovat nějak tuto funkci a pak odhadovat tyto parametry společně s β

– typická parametrizace: $z_j = x_j^{\lambda}$, $\lambda \in \mathbb{R}$ vhodné
($x_j > 0$ pokud $\lambda \in \mathbb{R}$, pokud x_j může být i záporné, množina hodnot λ je omezená)

– aproximace f pomocí polynomu vhodného stupně, tzn.

$$z_j = \sum_{k=1}^{l} \gamma_k x_j^k ,$$ kde γ_k mají být odhadnuty

výsledný model ale v tomto případě nebude lineární v parametrech β_j, $j = 0, \dots, m$ a γ_k, $k = 1, \dots, l$

- další možnost je použít trigonometrických funkcí nebo splines (piecewise polynomials)

Zaměříme se na $z_j = x_j^{\lambda}$:

- možnost je opět zvolit jistou množinu hodnot λ, vyladit modely pro všechna λ a vybrat model s nejlepší shodou s daty, např. s nejmenší SSE nebo největší R^2 nebo F
- může být časově náročné, můžeme minout vhodnou hodnotu λ, pokud nebyla v původní množině (nevíme jak R^2, F, SSE závisí na λ)

BOX – TIDWELL metoda

předp., že λ se příliš neliší od $\lambda = 1$, Taylorův rozvoj 1. řádu kolem $\lambda = 1$ dává

$$x^{\lambda} \approx x^1 + (\lambda - 1) \left.\frac{dx^{\lambda}}{d\lambda}\right|_{\lambda=1} \qquad \frac{dx^{\lambda}}{d\lambda} = x^{\lambda} \ln x \quad a \quad \left.\frac{dx^{\lambda}}{d\lambda}\right|_{\lambda=1} = x \ln x$$

$$x^{\lambda} \approx x + (\lambda - 1)\, x \ln x$$

dosazením do modelu

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_{j-1} x_{j-1} + \beta_j (x_j + (\lambda - 1) x_j \ln x_j) + \dots + \beta_m x_m + e$$

$$= \beta_0 + \sum_{l=1}^{m} \beta_l x_l + \underbrace{\beta_j (\lambda - 1)}_{\beta_{m+1}(\lambda)} x_j \ln x_j + e$$

máme lineární model s ~~...~~ s parametry β_l, $0 \le l \le m+1$

protože $\beta_{m+1} = (\lambda - 1)\beta_j$ můžeme (λ, β_j) odhadnout následovně:

1) vyladíme původní model a spočteme LSE $\hat{\beta}_j$ parametru β_j
2) vyladíme rozšířený model s $x_{m+1} = x_j \ln x_j$ a spočtem $\hat{\beta}_{m+1}$
3) z rovnosti $\hat{\beta}_{m+1} = (\hat{\lambda} - 1)\hat{\beta}_j$ dostaneme $\hat{\lambda} = \frac{\hat{\beta}_{m+1}}{\hat{\beta}_j} + 1$

tento postup umožňuje testovat potřebu transformace

$$H_0: \lambda = 1 \quad \times \quad H_1: \lambda \neq 1$$

pomocí t-testu pro $H_0: \beta_{m+1} = 0$

~~navíc 100(1−α)% IS pro λ je (β̂_{m+1} + 1~~

POZN: pokud model s $\hat\lambda$ vyjde neodpovídající, lze postupovat iteračně a získat posl. $\hat\lambda(l)$, $l \geq 1$

$\hat\lambda(0) = \hat\lambda$ a rozvineme x_j^λ kolem $\hat\lambda(0)$, tzn.

$$x_j^\lambda \approx x_j^{\hat\lambda(0)} + (\lambda - \hat\lambda(0))\, x_j^{\hat\lambda(0)} \ln x_j$$

dosazením do rovnice modelu

$$y = \beta_0 + \sum_{\substack{l=1 \\ l \neq j}}^{m} \beta_l x_l + \beta_j x_j^{\hat\lambda(0)} + \underbrace{\beta_j(\lambda - \hat\lambda(0))}_{\beta_{m+1}} x_j^{\hat\lambda(0)} \ln x_j + e$$

vyřešíme tento model s a bez přidané proměnné $x_{m+1} = x_j^{\hat\lambda(0)} \ln x_j$

ozn. $\hat\beta_j(1)$ a $\hat\beta_{m+1}(1)$ příslušné odhady

potom
$$\hat\lambda(1) = \hat\lambda(0) + \frac{\hat\beta_{m+1}(1)}{\hat\beta_j(1)}$$

můžeme dále iterovat do konvergence nebo skončit po pevném počtu iterací

POZN: další užívané transformace $\mathbb{X}$, y

a) centrované proměnné

$\mathbb{X}_c$ $\quad (\mathbb{X}_c)_{ij} = x_{ij} - \bar{x}_j \qquad i = 1, \dots, n \quad j = 1, \dots, m$

kde $\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$ průměr j-tého sloupce matice $\mathbb{X}$

y_c $\quad (y_c)_i = y_i - \bar{y}$

1) $\hat\beta_1, \dots, \hat\beta_m$ je řešením $(\mathbb{X}_c^T \mathbb{X}_c)\begin{pmatrix}\beta_1 \\ \vdots \\ \beta_m\end{pmatrix} = \mathbb{X}_c^T y_c$ $\qquad$ 2) $\hat\beta_0 = \bar{y} - \sum_{j=1}^{m} \hat\beta_j \bar{x}_j$

b) centrované a škálované proměnné

škálování sloupců tak, aby jejich norma byla 1

tzn. hodnoty prvků j-tého sloupce matice $\mathbb{X}_c$ podělíme $s_j = \left(\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \right)^{\frac{1}{2}}$

centrovaná a škálovaná matice $\mathbb{X}_{sc}$ pak bude

$$\mathbb{X}_{sc} = \mathbb{X}_c \cdot \mathbb{S}, \qquad \mathbb{S} = \operatorname{diag}\left(\frac{1}{s_1}, \dots, \frac{1}{s_m}\right)$$

model potom bude

(platí $\hat{\beta}_s = \mathbb{S}^{-1} \hat{\beta}_c$)

$$\mathbb{Y}_c = \mathbb{X}_{sc} \beta_s + e$$

lze použít i $\mathbb{Y}_{sc}$, tedy centrované a škálované $\mathbb{Y}$ (více str. 135)

Vážené nejmenší čtverce (weighted least squares WLS)

budeme nyní předpokládat, že chyby e_i jsou normální, nezávislé, ale $\operatorname{var}(e_i) = \sigma_i^2$ závisí na i

konkrétně: $\sigma_i^2 = \dfrac{\sigma^2}{w_i}$, kde $w_i > 0$, $i = 1, \dots, n$ jsou známé váhy

uvažujeme tedy model

$$\mathbb{Y} = \mathbb{X}\beta + e, \quad \text{kde} \quad e \sim N_n(0, \sigma^2 \mathbb{W}) \qquad (1)$$

kde $\mathbb{W} = \operatorname{diag}\left(\frac{1}{w_1}, \dots, \frac{1}{w_n}\right)$

protože jsou váhy w_i známé, lze MLE odhady par. β a σ^2 nalézt následovně:

$$\mathbb{W} = \mathbb{K}\mathbb{K}^T, \quad \text{kde} \quad \mathbb{K} = \mathbb{W}^{\frac{1}{2}} = \operatorname{diag}\left(\frac{1}{\sqrt{w_1}}, \dots, \frac{1}{\sqrt{w_n}}\right)$$

def.

$$\mathbb{Z} = \mathbb{K}^{-1}\mathbb{Y} \qquad \mathbb{M} = \mathbb{K}^{-1}\mathbb{X} \quad \text{a} \quad \mathcal{E} = \mathbb{K}^{-1} e$$

dostaneme model $\underline{Z = M\beta + \varepsilon}$, kde $\underline{\varepsilon \sim N_n(0, \sigma^2 I_n)}$ (2)

transf. vektor $Z = (\sqrt{w_1}\, Y_1, \dots, \sqrt{w_n}\, Y_n)^T$

protože $Cov(\varepsilon) = K^{-1}\sigma^2 W (K^{-1})^T = \sigma^2 K^{-1} K K^T (K^T)^{-1}$
$= \sigma^2 I_n$

to už je standardní model LR, ve kterém platí

$$\hat\beta_w = (M^T M)^{-1} M^T Z = \Big(X^T \underbrace{(K^{-1})^T K^{-1}}_{= (K K^T)^{-1} = W^{-1}} X \Big)^{-1} X^T (K^{-1})^T K^{-1} Y = \underline{(X^T W^{-1} X)^{-1} X^T W^{-1} Y}$$

↖ WLS estimator

$$\hat\sigma^2_w = \frac{1}{n} \sum_{i=1}^{n} (\underset{\sqrt{w_i}\, Y_i}{Z_i} - \underset{\sqrt{w_i}\, x_i^T \hat\beta = \sqrt{w_i}\, \hat Y_i}{\hat Z_i})^2 = \underline{\frac{1}{n} \sum_{i=1}^{n} w_i (Y_i - \hat Y_i)^2} = \frac{1}{n} SSE_w$$

↑ vážený součet čtverců

platí: $E\hat\beta_w = (X^T W^{-1} X)^{-1} X^T W^{-1} \underbrace{EY}_{= X\beta} = \beta$ $\quad$ $\hat\beta_w$ je nestranný odhad β

$E\left(\frac{SSE_w}{n-m-1}\right) = \sigma^2$, tedy $\underline{s^2_w = \frac{SSE_w}{n-m-1}}$ je nestranný odhad σ^2

Věta: Nechť $\hat\beta_w$ je WLS odhad β, jestliže $Cov(e) = \sigma^2 W = \sigma^2 \operatorname{diag}\left(\frac{1}{w_1}, \dots, \frac{1}{w_n}\right)$.
Potom platí:

1) $\underline{Cov(\hat\beta_w) = \sigma^2 (X^T W^{-1} X)^{-1}}$

2) nechť v_i je i-tý diag. prvek $(X^T W^{-1} X)^{-1}$. Jestliže $e_i \sim N\left(0, \frac{\sigma^2}{w_i}\right)$, $i = 1, \dots, n$, potom
$$\underline{T_i = \frac{\hat\beta_{w,i} - \beta_i}{s_w \sqrt{v_i}} \sim t(n-m-1)}$$

3) pro $\hat Y_w = X\hat\beta_w$ platí $E\hat Y_w = X\beta$ a $Cov(\hat Y_w) = \sigma^2 X (X^T W^{-1} X)^{-1} X^T$

4) nechť $\hat e_w$ ~~jsou rezidua~~ $= Y - \hat Y_w$ jsou rezidua v modelu (1)
a $\hat\varepsilon_w = Z - \hat Z = Z - M\hat\beta_w$ jsou rezidua v transformovaném modelu (2).
Potom $\hat\varepsilon_w = \sqrt{W^{-1}}\, \hat e_w = W^{-\frac{1}{2}} \hat e_w$ a $E(\hat e_w) = E(\hat\varepsilon_w) = 0$

5) Nechť $\mathbb{H}_w = X(X^TW^{-1}X)^{-1}X^TW^{-1}$ je vážená projekční matice.

Potom $\underline{\hat{e}_w = (I-\mathbb{H}_w)e}$ a $\underline{Cov(\hat{e}_w) = \sigma^2(I-\mathbb{H}_w)W}$.

To znamená $\underline{Cov(\hat{\varepsilon}_w) = \sigma^2 W^{-\frac{1}{2}}(I-\mathbb{H}_w)W^{\frac{1}{2}}}$.

Dk: 1) $Cov(\hat\beta_w) = (X^TW^{-1}X)^{-1}X^TW^{-1}\underbrace{Cov\, Y}_{=\sigma^2 W}W^{-1}X(X^TW^{-1}X)^{-1} = \sigma^2(X^TW^{-1}X)^{-1}$

2) $Var\, \hat\beta_{w,i} = \sigma^2\delta_i$, tzn. $\frac{\hat\beta_{w,i}-\beta_i}{\sigma\sqrt{\delta_i}} \sim N(0,1)$ a víme, že

$\hat\beta_{w,i}$ a s_w^2 jsou nezávislé, $\frac{s_w^2(n-m-1)}{\sigma^2} \sim \chi^2(n-m-1)$

$$\Rightarrow \frac{\hat\beta_{w,i}-\beta_i}{s_w\sqrt{\delta_i}} \sim t(n-m-1)$$

3) $E\hat{Y}_w = X E\hat\beta_w = X\beta$, $Cov(\hat{Y}_w) = X\, Cov\hat\beta_w X^T = \sigma^2 X(X^TW^{-1}X)^{-1}X^T$

4) protože $Z = W^{-\frac{1}{2}}Y$ a $M = W^{-\frac{1}{2}}X$ dostaneme

$$\hat\varepsilon_w = Z - \hat{Z} = W^{-\frac{1}{2}}Y - W^{-\frac{1}{2}}\underset{X}{\cancel{M}}\hat\beta_w = W^{-\frac{1}{2}}(Y - X\hat\beta_w) = W^{-\frac{1}{2}}\hat{e}_w$$

$$E\hat{e}_w = E(Y-\hat{Y}_w) = X\beta - X\beta = 0 \Rightarrow E\hat\varepsilon_w = 0$$

5) $\hat{e}_w = Y - \hat{Y}_w = Y - X\hat\beta_w = X\beta + e - X(X^TW^{-1}X)^{-1}X^TW^{-1}\overbrace{(X\beta+e)}^{=Y}$

$= X\beta - X\beta + e - \underbrace{X(X^TW^{-1}X)^{-1}X^TW^{-1}}_{\mathbb{H}_w}e = \underline{(I-\mathbb{H}_w)e}$

$Cov(\hat{e}_w) = (I-\mathbb{H}_w)\, Cov\, e\,(I-\mathbb{H}_w)^T = \sigma^2(I - X(X^TW^{-1}X)^{-1}X^TW^{-1})W(I - W^{-1}X(X^TW^{-1}X)^{-1}X^T)$

$= \sigma^2W - \sigma^2X(X^TW^{-1}X)^{-1}X^T - \sigma^2X(X^TW^{-1}X)^{-1}X^T + \sigma^2X(X^TW^{-1}X)^{-1}X^T$

$= \sigma^2(I-\mathbb{H}_w)W$

$Cov(\hat\varepsilon_w) = W^{-\frac{1}{2}}\, Cov(\hat{e}_w)\, W^{-\frac{1}{2}} = \sigma^2 W^{-\frac{1}{2}}(I-\mathbb{H}_w)W^{\frac{1}{2}}$ Q.E.D.

- z odvození vyplývá, že odhady par. β a σ^2 lze získat použitím transformovaného modelu (2)

- protože ale transformovaný model neobsahuje intercept,
(první sloupec M je $(\sqrt{w_1}, \dots, \sqrt{w_n})^T$)
nefunguje klasický rozklad součtu čtverců a F statistiku nelze def. obvyklým způsobem stejně jako R^2 (viz regrese skrz počátek)

- nicméně princip „extra ~~sum~~ sum of squares" funguje ať má model intercept nebo ne

např. celkový F-test lze provést pomocí stat.

$$F_w = \frac{\frac{SSE_R - SSE_F}{m}}{s_w^2}$$

, kde SSE_F je reziduální součet čtverců s_w^2 plného modelu

a SSE_R je rez. s. č. redukovaného transformovaného modelu $Z = M_0\beta_0 + \varepsilon$,
$M_0 = (\sqrt{w_1}, \dots, \sqrt{w_n})^T$

pokud mají chyby normální rozdělení, platí za $H_0: \beta_1 = \dots = \beta_m = 0$,

že $F_w \sim F(m, n-m-1)$, H_0 zamítáme pokud $F_w > F_{1-\alpha}(m, n-m-1)$

- str. 110 (*)
- pro analýzu reziduí je třeba uvažovat vhodné grafy reziduí

– máme dva vektory reziduí $\hat{e}$ v původním modelu (1)
$\hat{\varepsilon}$ v transformovaném modelu (2)

a tedy dvě možnosti

– pro kontrolu konstantního rozptylu lze uvažovat i standardizovaná nebo studentizovaná rezidua (pomocí bodu 4 (vlastnosti) lze ukázat, že jsou v obou modelech stejná)

– je třeba být opatrný oproti jakým hodnotám budeme rezidua zobrazovat

- grafy $\hat{\varepsilon}_i$ proti slorzkám M a predikovaným hodnotám $\hat{Y}$ jsou OK, neboť např.

$$\sum_{i=1}^{n} \hat{Y}_i \hat{\varepsilon}_i = 0$$

(jsou OG, měl by být vidět rozptyl obláčku kolem osy x)

- dosazením $\hat{\varepsilon}_i = \sqrt{w_i}\, \hat{e}_i$ a $\hat{Y}_i = \sqrt{w_i}\, \hat{y}_i$ dostaneme

$$\sum_{i=1}^{n} w_i \hat{e}_i \hat{y}_i = 0$$

tzn. graf $\hat{e}_i$ proti $\hat{y}_i$ bude zavádějící

- graf $\sqrt{w_i}\, \hat{e}_i$ proti $\sqrt{w_i}\, \hat{y}_i$ je ale v pořádku
- podobné závěry platí i pro grafy $\hat{e}_i$ proti x_j^i, $i = 1, \dots, n$

⊛ původně je def. $R^2 = \rho^2(\hat{Y}, Y)$, kde $\rho(\hat{Y}, Y)$ je výběrový korelační koeficient, pro $W = I$ dostaneme standardní R^2

POZN: – pokud jsou váhy neznámé, bylo by třeba je odhadnout společně s β a σ^2 z dat

– to ale není obecně možné, protože máme více parametrů než dat

– někdy to možné je, pokud máme další informace o rozdělení chyb (tvar kovarianční matice, B_i, G_i)

POZN: celý postup WLS lze použít i na případ $\varepsilon \sim N_n(0, \sigma^2 W)$, kde W je známá, ale není diagonální

protože W je symetrická, ex. reg. K tak, že $W = KK^T$

stejná transformace jako u WLS opět vede na transf. model, kde $\varepsilon \sim N_n(0, \sigma^2 I_n)$.

Korelované chyby

- zejména v časových nebo ekonomických datech se často objevuje korelace jedn. hodnot
- potom není splněn předpoklad nezávislosti chyb,
- tento stav je třeba detekovat (někdy pomohou grafy reziduí)
- modely pro korelovaná data: Analýza časových řad

Pokud je přítomna autokorelace a chyby mají konstantní rozptyl platí:

1) OLS odhad $\hat{\beta}$ je nestranný, ale neplatí Gauss-Markovova věta, tzn. $\hat{\beta}$ nemá nejmenší rozptyl
2) $MSE = \frac{1}{n-m-1} SSE$ (odhad σ^2) může být podstatně menší než skutečná hodnota σ^2, což může dávat falešný pocit přesnosti
3) v důsledku 2) mohou být zvětšeny hodnoty T statistik, takže testy o parametrech a IS nefungují
4) protože jsou chyby závislé, F-test a t-testy nejsou přesně platné ani když jsou chyby normální

DURBIN - WATSON statistika

omezíme se na pozorování získaná v čase $t = 1, 2, \dots, n$ a případ, že chyby e_t splňují podmínky autoregresního procesu 1. řádu (AR1),

tj.
$$e_t = \rho e_{t-1} + u_t, \qquad |\rho| < 1$$

ρ je autokorelační koeficient, $u_t \sim N(0, \sigma_u^2)$ jsou nezávislé $t = 1, \dots, n$ a u_t je nezávislé na e_t, $t \geq 1$

častěji pro data časových řad platí $\rho > 0$ (pozitivní autokorelace) pro test $H_0 : \rho = 0 \quad \times \quad H_1 : \rho > 0$ se užívá

Durbin-Watsonova statistika

$$d = \frac{\sum_{t=2}^{n} (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{n} \hat{e}_t^2} \quad , \text{ kde } \hat{e}_t \text{ jsou rezidua modelu LR}$$

pokud zamítneme H_0, odhadneme se ρ pomocí $\hat{\rho} = \dfrac{\sum_{t=2}^{n} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=2}^{n} \hat{e}_t^2}$

POZN: dá se ukázat, že $d \approx 2(1-\hat{\rho})$

$$\sum_{t=2}^{n} (\hat{e}_t - \hat{e}_{t-1})^2 = \sum_{t=2}^{n} \hat{e}_t^2 + \sum_{t=2}^{n} \hat{e}_{t-1}^2 - 2 \sum_{t=2}^{n} \hat{e}_t \hat{e}_{t-1} \approx$$

$$\approx 2 \left(\sum_{t=2}^{n} \hat{e}_t^2 - \sum_{t=2}^{n} \hat{e}_t \hat{e}_{t-1} \right)$$

Cauchy-Schwarzova nerovnost $\Rightarrow$ $\dfrac{\sum_{t=2}^{n} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{n} \hat{e}_t^2}$ leží přibližně v $(-1,1)$

tzn. d leží přibližně v $(0,4)$

$\hat{\rho} \approx 1 \Rightarrow d \approx 0 \qquad \hat{\rho} \approx 0 \Rightarrow d \approx 2$ tzn. pro malé hodnoty d budeme zamítat H_0, pro velké nebudeme zamítat

– kritické hodnoty určené Durbinem a Watsonem jsou tabelované

TEST:
1) spočítat hodnotu d
2) nalézt kritické hodnoty (d_L, d_U) pro dané n a $m+1$
3) a) zamítnout H_0, pokud $d < d_L$
   b) nezamítnout H_0, pokud $d > d_U$
   c) pro $d_L < d < d_U$ test nerozhodne

POZN: pro test $H_0 : \rho = 0 \times H_1 : \rho < 0$ lze použít podobný test pro $d' = 4 - d$

metody pro korekci autokorelace: Cochrane-Orcutt