Autokorelace náhodných složek

Druhou nesnází, která provází odhad zobecněného lineárního regresního modelu, je případná ***autokorelace náhodných složek*** regresní rovnice. Tento dost častý úkaz se vyskytuje daleko častěji u jednorovnicového modelu, jehož pozorování tvoří časové řady (u průřezových údajů je pozorován vzácně)[1]. Jeho důsledkem je, že **odhady parametrů modelu** pořízené obyčejnou metodou nejmenších čtverců OLS sice **zůstávají nestranné, ale ztrácejí vydatnost**. Při znalosti kovarianční matice náhodných složek[2] je pro získání vydatného odhadu nutno uplatnit zobecněnou metodu nejmenších čtverců GLS. Odhad s^2 rozptylu náhodných složek σ^2 metodou OLS je však – při vzájemně autokorelovaných náhodných složkách -vychýlený.

Indikace přítomnosti autokorelovanosti náhodných složek

O přítomnosti autokorelace náhodných složek (jejichž teoretické hodnoty neznáme) se lze přesvědčit jen nepřímo, vyšetřením reziduálních hodnot.

Velmi názorný obrázek o míře autokorelovanosti náhodných složek podává

A) Durbin-Watsonův koeficient autokorelace reziduí určený výrazem

(1)
$$DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^{\,2}}$$

který je **definován jako podíl součtu čtverců diferencí dvou po sobě jdoucích reziduálních hodnot a součtu čtverců všech reziduí.** Rozsah přípustných hodnot DW- koeficientu se pohybuje v rozmezí **< 0 , 4 >**, přičemž obě krajní hodnoty signalizují maximální možnou korelovanost dvou následujících reziduálních hodnot. **Pro případ DW ≅ 0 jde o kladnou autokorelaci 1.řádu, v případě DW ≅ 4 o zápornou autokorelaci,** zatímco prostřední **hodnota DW ≅ 2 signalizuje nepřítomnost autokorelace 1.řádu.**

Poznámka 1 Po umocnění výrazu v čitateli vzorce pro DW vidíme, že

a) při nepřítomnosti autokorelace reziduí bude skalární součin vektorů e_t a e_{t-1} blízký nule, takže zbytek čitatele bude přibližně rovný dvojnásobku jmenovatele

b) při silné kladné autokorelaci bude tento skalární součin blízký $\sum_{t=2}^{T} e_t^{\,2}$ a výraz $-2.\sum_{t=2}^{T} e_t.e_{t-1}$ bude přibližně roven součtu $-\sum_{t=2}^{T} (e_t^{\,2} + e_{t-1}^{\,2})$

c) konečně při silné záporné autokorelaci bude zmíněný skalární součin blízký $\sum_{t=2}^{T} e_t^{\,2}$ a výrazy $-2.\sum_{t=2}^{T} e_t.e_{t-1}$ a $\sum_{t=2}^{T} (e_t^{\,2} + e_{t-1}^{\,2})$ budou přibližně stejné co do absolutní velikosti i co do znamének

[1] Příčinou je zejména.to, že v průřezových vzorcích jsou hodnoty jednotlivých případů/pozorování řazeny zpravidla nahodile, takže není sebemenší důvod usuzovat, že může existovat souvislost (která závisí na pořadí) mezi nimi (s jakoukoliv náhodnou záměnou pořadí pozorování by se tato souvislost musela nutně změnit). Hodnoty pozorování v časových řadách jsou naproti tomu – až na naprosté výjimky - řazeny chronologicky.

[2] To je ovšem dost výjimečná situace.

Vše platí za předpokladu, že rozdíl v počtu členů sumací (ve jmenovateli je o 1 člen více) nebude při dostatečně početném datovém vzorku podstatný.

Poznámka 2 Zřetelná kladná autokorelace (1.řádu) je charakteristická delšími řetězci shodných znamének reziduálních hodnot ležícími střídavě nad a pod vyrovnávající regresní přímkou/nadrovinou.

Silná záporná autokorelace (1.řádu) je naopak typická téměř pravidelným střídáním znamének reziduálních hodnot (tj. zřetelnou oscilací dvou po sobě jdoucích reziduálních hodnot kolem vyrovnávající regresní přímky/nadroviny).

Nevýhodou *Durbin-Watsonova koeficientu* je však skutečnost, že **empiricky získanou hodnotu DW nelze v úplnosti statisticky testovat** (rozdělení testové statistiky závisí na prvcích matice $\boldsymbol{X}$, které jsou rozdílné pro každý statistický výběr). V důsledku toho obsahuje interval přípustných hodnot pro DW-koeficient $<0,4>$ dvě „hluché oblasti“, v nichž nelze rozhodnout, zda hypotéza o nepřítomnosti autokorelace 1. řádu bude zamítnuta ve prospěch některé z alternativ (kladná či záporná autokorelovanost). Tak

v intervalu $<0,d_D>$	**se zamítá hypotéza neautokorelovanosti ve prospěch alternativy: existence kladné autokorelace 1. řádu**
v intervalu $<d_D,d_H>$	**nelze test rigorózně vyhodnotit**
v intervalu $<d_H,4-d_H>$	**se nezamítá (přijímá) hypotéza neautokorelovanosti**
v intervalu $<4-d_H,4-d_D>$	**nelze test rigorózně vyhodnotit**
v intervalu $<4-d_D,4>$	**se zamítá hypotéza neautokorelovanosti ve prospěch alternativy: existence záporné autokorelace 1.řádu**

„Mezní“ hranice d_D,d_H lze spočítat pro libovolný počet stupňů volnosti (T-k) a pro obvyklé hladiny významnosti (α=0,01 nebo 0,05). Příslušné hodnoty jsou tabelovány.

Poznámka 3. DW test není přímo použitelný v případě testování sériové korelace vyšších řádů nebo při nelineární formě autokorelace náhodných složek. Některé modifikace k zmírnění problému v těchto situacích navrhli **Nerlove, Wallis, Theil, Nagar** a **Geary.**

Postupy vedoucí k eliminaci autokorelovanosti náhodných složek

B) COCHRANE-ORCUTTova procedura[3] kterou lze popsat tímto iterativním opakováním následujících tří fází

B0) Předstupněm iteračního procesu je **výpočet parametrů $b^{(0)}$ modelu v původní specifikaci obyčejnou metodou nejmenších čtverců OLS a následné stanovení vyrovnaných hodnot závisle proměnné $\hat{y}_t$ a hodnot reziduí e_t**

B1) **Formulujeme autoregresní schéma 1. řádu pro rezidua e_t ve tvaru**

(2) $$e_t = \rho_1 e_{t-1} + v_t$$

kde v_t je příslušný *bílý šum* autoregresního procesu 1. řádu (jeho realizace jsou centrované, nekorelované, homoskedastické a stejně rozdělené náhodné veličiny). K zachování stacionarity procesu je, jak známo, nutné splnění podmínky $|\rho_1| < 1$.

B2) **Odhad $\hat{\rho}_1$ koeficientu autokorelace 1. řádu ρ_1 získáme pomocí výrazu**

(3) $$\hat{\rho}_1 = \frac{\sum_{t=1}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_{t-1}^{\,2}}$$

při dodefinování hodnotou $e_0 = 0$.

B3) **Takto získaný odhad $\hat{\rho}_1$ se použije v modifikovaném regresním modelu.** Modifikace je představována úpravou jednotlivých modelových proměnných pomocí ***metody zobecněných diferencí*.**

V případě např. 3 vysvětlujících proměnných, kde první „vysvětlující“ proměnnou představuje vektor jedniček, má tento vztah podobu (pro $t = 2,3,...,T$)

(4) $$y_t - \hat{\rho}_1 y_{t-1} = \beta_1(1-\hat{\rho}_1) + \beta_2(x_{t2} - \hat{\rho}_1 x_{t-1,2}) + \beta_3(x_{t3} - \hat{\rho}_1 x_{t-1,3}) + ... + \varepsilon_t - \hat{\rho}_1 \varepsilon_{t-1}$$

Nasazením metody OLS na takto modifikovaný model získáme upravený odhad $\hat{\beta}^{(1)}$ vektoru parametrů β. Tento odhad se dosadí do původního modelu a následně se spočtou (přes vyrovnané hodnoty) **upravená rezidua $e_t^{(1)}$**. S těmi se vstoupí do druhého kroku iterační procedury představované opakováním fáze ***B1)***. Následuje opět sekvence operací ***B2) , B3)*** atd.

Poté, co **v průběžném r-tém kroku získáme odhady $\hat{\beta}_1^{(r)}, \hat{\beta}_2^{(r)}, \hat{\beta}_3^{(r)}, \hat{\rho}_1^{(r)}$** pro $\beta_1, \beta_2, \beta_3, \rho_1$, **porovnáme je s hodnotami veličin $\hat{\beta}_1^{(r-1)}, \hat{\beta}_2^{(r-1)}, \hat{\beta}_3^{(r-1)}, \hat{\rho}_1^{(r-1)}$ získanými v předchozím $r-1-$*tém* kroku**. Jestliže rozdíly ve dvou po sobě jdoucích krocích nepřekročí předepsanou odchylku (stanovenou např. ve formě maxima z odchylek u jednotlivých parametrů anebo jako v absolutní hodnotě vzatý rozdíl odhadů autoregresního koeficientu 1.řádu ρ_1 tj. $\left|\hat{\rho}_1^{(r)} - \hat{\rho}_1^{(r-1)}\right|$), můžeme výsledky dosažené v daném iteračním kroku považovat za uspokojivé a příslušné odhady převzít jako konečné.

---

[3] **Cochrane, D., Orcutt, G.H.: Application of Least-Squares Regressions to Relationships Containing Autorrelated Error Terms. JASA 44/1949 str.32-61**.

Určitou podobnost s předchozím postupem vykazuje

C) DURBINova dvoustupňová metoda[4] použitelná i v případě přítomnosti autokorelace vyšších řádů u náhodných složek

C1) **Model se nejprve** – obdobně jako v kroku ***B3)*** předchozí metody – **převede na tvar zobecněných diferencí**

(5) $$y_t - \rho_1 y_{t-1} = \beta_1(1-\rho_1) + \beta_2(x_{t2} - \rho_1 x_{t-1,2}) + \beta_3(x_{t3} - \rho_1 x_{t-1,3}) + \ldots + \varepsilon_t - \rho_1 \varepsilon_{t-1}$$

resp. **po substitucích** $\beta_1(1-\rho_1) = \gamma_1$, $-\beta_2.\rho_1 = \gamma_2$, $-\beta_3.\rho_1 = \gamma_3$ atd. a $w_t = \varepsilon_t - \rho_1.\varepsilon_{t-1}$ **získá tvar**

(6) $$y_t = \rho_1 y_{t-1} + \gamma_1 + \beta_2 x_{t2} + \gamma_2.x_{t-1,2} + \ldots + \beta_k x_{tk} + \gamma_k.x_{t-1,k} + w_t$$

Odtud se **pomocí metody OLS získá** konzistentní **odhad autoregresního koeficientu 1.řádu** ρ_1 **příslušejícího zpožděné hodnotě proměnné** y_{t-1} .

C2) Tento **odhad** $\hat{\rho}_1$ **se dosadí do výchozího tvaru modelu (5)**[5] **a opětovným použitím OLS se získají zpřesněné odhady** $\hat{\beta}_1$ pro β_1, $\hat{\beta}_2$ pro β_2, $\hat{\beta}_3$ pro β_3 atd. s uspokojivými asymptotickými (tj. pro velký rozsah výběru T) vlastnostmi.

Poznámka 3

Odhad parametru ρ_1 získávaný v prvním kroku procedury *Durbinovy dvoustupňové metody* se zde tedy neprování podle **(3)** jako v případě *Cochran-Orcuttovy metody*, ale regresí **(6)** s maticí vysvětlujících proměnných Z tvaru:

$$y = \begin{pmatrix} y_2 \\ y_3 \\ \ldots \\ y_T \end{pmatrix} = Z = \begin{pmatrix} y_1 & 1 & x_{22} & x_{12} & \ldots & x_{1k} \\ y_2 & 1 & x_{32} & x_{22} & \ldots & x_{3k} \\ \ldots & .. & \ldots & .. & \ldots & \\ y_{T-1} & 1 & x_{T2} & x_{T-1,2} & \ldots & x_{Tk} \end{pmatrix} . \begin{pmatrix} \rho_1 \\ \gamma_1 \\ \beta_1 \\ \gamma_2 \\ \ldots \\ \gamma_k \end{pmatrix} + w = \begin{pmatrix} w_2 \\ w_3 \\ \ldots \\ w_T \end{pmatrix}$$

Z této regrese se použije pouze odhad parametru ρ_1, zatímco ostatní se neuplatní; ty se potom získávají až následně krokem ***C2)*** z (5) s již dosazeným $\hat{\rho}_1$

Poznámka 4

Nevýhodou tohoto postupu je však zřetelně **zvýšený počet odhadovaných parametrů**, který dosáhne **počtu 2k-1** - ke každému původnímu β_j (j=2,.., k) (tedy až na β_1) přísluší nyní dvojice parametrů β_j, γ_j z nichž jeden je „původní“ a druhý γ_j vznikne násobením β_j hodnotou $-\rho_1$. V případě relativně malého počtu pozorování T ve srovnání s počtem vysvětlujících proměnných regresní rovnice k není tedy tento postup příliš vhodný.

---

[4] **Durbin,J.: Testing for Serial Correlation in Least-Squares Regression when Some of the Regressors are Lagged Dependent Variables. Econometrica 38/1970 str. 410-421.**

[5] Koeficient autoregrese ρ_1 se v tomto případě bere již jako známý.

Nevýhodu spojenou s oblastmi nerozhodnutelnosti testování závisejícími na d_D, d_H u Durbin-Watsonova koeficientu odstraňuje podobně konstruovaná míra známá jako

D) von Neumannův koeficient (podíl)[6] autokorelace reziduí

Tato míra je definována vztahem

(7)
$$vN = \frac{T}{T-1} \cdot \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Lze ukázat, že ***jsou-li náhodné složky*** ε_t ***, a tedy i rezidua*** e_t ***normálně rozdělena, pak pro dost velký počet pozorování T má statistika vN také přibližně normální rozdělení.*** Její střední hodnota a rozptyl jsou dány výrazy

(7A)
$$E(vN) = \frac{2T}{T-1} \qquad D(vN) = \frac{4T^2(T-2)}{(T+1)(T-1)^3}$$

Kritické hodnoty ***vN***-podílu jsou pro různá T a obvykle používané hladiny významnosti tabelovány.

V situacích, kdy se mezi vysvětlujícími proměnnými objevují též zpožděné endogenní proměnné, není použití *Durbin-Watsonova koeficientu* vhodné. Rezidua e_t získaná metodou OLS nejsou v tomto případě nezávisle rozdělena, dokonce ani tehdy ne, jsou-li nezávisle rozděleny náhodné složky ε_t. To snižuje přínos této statistiky při aplikaci v ekonometrických modelech .

Příčinou toho, že DW-koeficient nedává při přítomnosti vysvětlujících zpožděných endogenních proměnných objektivní závěry, je skutečnost, že DW-koeficient se v tomto případě blíží ke 2 v důsledku výskytu právě těchto proměnných, nejen v důsledku přítomných (případně však i neautokorelovaných) náhodných složek.

Poznámka 5

Z výrazu (7) je zřejmé, že mezi oběma charakteristikami platí $VN = \frac{T}{T-1}.DW$

Účinnějším indikátorem autokorelovanosti reziduí je v některých situacích

E) Durbinova h-statistika autokorelace reziduí definována následovně

(8)
$$h = (1 - DW/2)\left[\frac{T}{1 - T.\mathrm{var}(b_{y_{t-1}}).}\right]^{1/2}$$

kde $\mathrm{var}(b_{y_{t-1}})$ **je odhad výběrového rozptylu odhadnutého regresního koeficientu u zpožděné endogenní proměnné** y_{t-1}. Při nulové hypotéze o sériové nezávislosti náhodných složek je ***h-statistika*** **asymptoticky normálně rozdělena (s nulovou střední hodnotou a jedničkovým rozptylem)** Lze ji testovat jako normální směrodatnou odchylku (alternativní hypotézou je přítomnost autokorelace 1. řádu). Omezenost jejího použití vyplývá z podmínky kladného jmenovatele $1 - \mathrm{var}(b_{y_{t-1}})$. Zde je nutno uplatnit alternativní testovací postupy.

---

[6] **von Neuman, John: Distribution of the Ratio of the Mean Square Successive Difference to the Variance. Annals of Mathematical Statistics 1941 s. 367-295**

Poznámka 5 V případě, že h-statistika není definována, doporučuje se (následně po provedení OLS-regrese) např. definovat regresní rovnici ve tvaru

(9) $$\boldsymbol{e_t} = \alpha_1.\boldsymbol{e_{t-1}} + \alpha_2.\boldsymbol{y_{t-1}} + \boldsymbol{x_t} + \eta_t$$

Testování hypotézy $\rho = 0$ se převede na testování statistické významnosti koeficientu α_1 v této regresi.

F) Berenblut-Webbův test[7] je založen na statistice

(10) $$\boldsymbol{BW} = \frac{\sum_{t=2}^{T} \boldsymbol{u_t}^2}{\sum_{t=1}^{T} \boldsymbol{e_t}^2}$$

kde $\boldsymbol{u_t}$ **jsou rezidua z regrese prvních diferencí** $\boldsymbol{y_t}$ **na první diference vysvětlujících proměnných** (bez konstanty) tj. z regrese

(11) $$y_t - y_{t-1} = \beta_1(x_{t1} - x_{t-1,1}) + \beta_2(x_{t2} - x_{t-1,2}) + .. + \beta_k(x_{tk} - x_{t-1,k}) + \varsigma_t$$

Jmenovatel **(10)** je obvyklý **SSE**, do kterého vstupují "původní rezidua" z OLS-regrese

(12) $$\boldsymbol{y_t} = \beta_1.1 + \beta_2\boldsymbol{x_{t2}} + \beta_3\boldsymbol{x_{t3}} + ... + \beta_k\boldsymbol{x_{tk}} + \varepsilon_t$$

Poznámka 6 Jen pro upřesnění značení: $\boldsymbol{e_t} = \hat{\varepsilon}_t$ a $\boldsymbol{u_t} = \hat{\varsigma}_t$.

Jestliže původní rovnice obsahuje konstantu, můžeme užít tabulky pro D-W testovou statistiku pro posouzení hodnot B-W-statistiky . BW-statistika je navíc uplatnitelná, i když se vyskytne situace, kdy $|\rho_1| \geq 1$.[8]

[7] **Berenblut, I.,I., Webb, G.,I.: A New Test for Autocorrelated Errors in the Linear Regression Model. Journal of the Royal Statistical Society Vol.35/1973 s. 33-50.**

[8] Stacionaritu procesu, k jejímuž zajištění je tato podmínka nutná, zajišťuje zde uplatnění "diferencovaných" pravostranných proměnných.

Jestliže je model homoskedastický a jsou-li náhodné složky regresní rovnice generovány autoregresním schématem 1.řádu, lze zapsat kovarianční matici náhodných složek v následující podobě:

$$\Sigma = \boldsymbol{E}(\varepsilon, \varepsilon') = \sigma^2 . \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^2 \\ \dots & \dots & \dots & \dots & \rho \\ \rho^{T-1} & \rho^{T-2} & \rho^2 & \rho & 1 \end{pmatrix}$$

V této matici, jak patrno, jsou na hlavní diagonále samé jedničky a na „rovnoběžkách“ s touto hlavní diagonálou vždy příslušné mocniny ρ^s, kde ***s*** je rovno rozdílu indexů příslušného prvku od součtu indexů diagonálního prvku (ležícího na stejném řádku, resp. sloupci).

Při takovémto schématu **lze uplatnit zobecněnou metodu nejmenších čtverců GLS** tak, že se transformace původních pozorování provede tak, že se k této transformaci použije matice ***R*** ve tvaru

$$\boldsymbol{R} = \frac{1}{\sqrt{1-\rho^2}} . \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & -\rho & 1 \end{pmatrix}$$

Tato matice má nenulové prvky jen ve dvou řadách. jednou je hlavní diagonála, která má všechny prvky rovny 1 až na první prvek, jehož hodnota je $\sqrt{1-\rho^2}$, zatímco druhou nenulovou řadou je řada ležící bezprostředně pod hlavní diagonálou, která je obsazena prvky s hodnotami rovnými $-\rho$.

Příslušná transformace se pak projeví tím způsobem, že pozorování jsou upravena do této podoby

$$\boldsymbol{y}^* = \begin{pmatrix} \boldsymbol{y}_1 . \sqrt{1-\rho^2} \\ \boldsymbol{y}_2 - \rho . \boldsymbol{y}_1 \\ .... \\ \boldsymbol{y}_T - \rho . \boldsymbol{y}_{T-1} \end{pmatrix} \qquad \boldsymbol{X}_{j^*} = \begin{pmatrix} \boldsymbol{x}_{j1} . \sqrt{1-\rho^2} \\ \boldsymbol{x}_{2j} - \rho . \boldsymbol{x}_{1j} \\ ..... \\ \boldsymbol{x}_{Tj} - \rho . \boldsymbol{x}_{T-1,j} \end{pmatrix}$$

tzn. že j-tý sloupec matice ***X*** je obsazen (vždy až na první prvek) **„zobecněnými diferencemi“**[9]

[9] Někdy se tento postup nazývá **Prais-Winstenovou transformací**.