Heteroskedasticita

Problém *heteroskedasticity* se vztahuje k nestejné velikosti diagonálních prvků kovarianční matice Σ vektoru náhodných složek ε jednorovnicového ekonometrického modelu. Matice Σ je v tomto případě diagonální, avšak její diagonální prvky (rozptyly náhodných složek v jednotlivých obdobích) nemají stejnou velikost, nelze tedy psát $\Sigma = \sigma^2.\boldsymbol{I_T}$, což by odpovídalo *homoskedasticitě* lineárního regresního modelu.

Negativní důsledky heteroskedasticity

V důsledku toho, že rozptyly náhodných složek nejsou stejně velké, nemá v této podobě zobecněného lineárního regresního modelu metoda OLS optimální vlastnosti – přesněji neposkytuje vydatné odhady, byť tyto zůstávají nestranné. Abychom získali vydatné odhady, je nutno použít **váženou metodu nejmenších čtverců WLS**, což je speciální (jednodušší) případ **zobecněné metody nejmenších čtverců GLS.**

Nejčastější příčiny heteroskedasticity

1) **Chybná specifikace modelu**, kdy vynecháme některou podstatnou vysvětlující proměnnou.

2) **Kumulace chyb měření proměnných při rostoucí hodnotě vysvětlované proměnné** mající za následek zvětšování rozptylu náhodných složek (následně i reziduí).

3) **Značná rozdílnost velikosti dat** v rámci jednoho náhodného výběru a odtud vyplývající závislost rozptylu vysvětlované proměnné (následně i rozptylu náhodných složek) na velikosti hodnot některé z vysvětlujících proměnných

4) **Použití** nikoliv původních pozorování, ale **skupinových průměrů** spočtených z nějakým způsobem setříděných údajů.

Ad 3) ***Heteroskedasticitu*** lze zaznamenat častěji u modelů založených na průřezových datech než u modelů využívajících časových řad. U modelů časových řad jsou totiž zařazené proměnné (vysvětlované i vysvětlující) zpravidla hodnotami poměrně blízké, zatímco srovnáváme-li průřezová data (např. firemní v rámci určitého odvětví) budou tato poznamenána (až řádově) rozdílnou intenzitou ekonomické činnosti podniku (počet zaměstnanců, objem tržeb, zásoby, hospodářský výsledek apod).

Vážená metoda nejmenších čtverců

Odhadová funkce pro vektor parametrů β je dána vztahem:

$$\hat{\beta} = (\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\Sigma^{-1}\boldsymbol{y}$$

kde jednotlivé prvky diagonální matice $\Sigma - 1$ jsou reciproké hodnoty původních rozptylů, tj. $[\sigma_t^{\,2}]^{-1} = \sigma^{tt} = 1/\sigma_t^{\,2}$. **Metoda poskytuje (v modelu zatíženém pouze heteroskedasticitou) nestranné a vydatné odhady parametrů**. V praktické situaci je ovšem nutno nahradit neznámé rozptyly $\sigma_t^{\,2}$ nějakými jejich (nestrannými) odhady $\boldsymbol{s}_t^{\,2}$.

Pokud bychom (výjimečně) znali veličiny $\sigma_t^{\,2}$, můžeme postupovat tak, že všechny proměnné modelu (tj. vysvětlovanou i k vysvětlujících) vydělíme směrodatnými odchylkami σ_t. Taková transformace modelových veličin povede k možnosti uplatnění prosté metody OLS na takto transformovaný model, přičemž se zachovají všechny příznivé statistické vlastnosti této metody. Pracujeme tedy s veličinami:

$$\boldsymbol{y}_t^{\,*} = \boldsymbol{y}_t/\sigma_t\ , \qquad \boldsymbol{x}_{tj}^{\,*} = \boldsymbol{x}_{tj}/\sigma_t \qquad \boldsymbol{j} = 1,2,\ldots,\boldsymbol{k};\ \boldsymbol{t} = 1,2,\ldots,\boldsymbol{T}\ \text{, přičemž}\ \boldsymbol{x}_{t1} = 1$$

Postupy navržené k indikaci heteroskedasticity v modelu

V předchozích desetiletích bylo vyvinuto několik testů, které umožňují indikovat *heteroskedasticitu*, případně odhadnout míru jejího vlivu. V prvních dvou testech se předpokládá, že

a) Kolísání náhodné složky je spojeno s proměnlivostí určité vysvětlující proměnné (a následně též s variabilitou vysvětlované proměnné) .

b) Náhodné složky mají normální rozdělení, aby bylo možno formulovat příslušné statistické testy.

V některých níže uvedených testovacích prostupech (jmenovitě u Goldfeld-Quandtova a Glejserova) se objevuje úvaha spočívající v tom, že variabilita náhodných složek je v nějaké formě „spřažena" s proměnlivostí hodnot některé z vysvětlujících proměnných. I když chování obou by – striktně vzato při správné specifikaci modelu – nemělo být nijak související, často se taková vázanost skutečně projevuje.

Příklad:

Představme si prostorový statistický vzorek, jehož prvky tvoří soubor firem určitého ekonomického odvětví (např. stavebnictví) a v němž jsou zastoupeny jak drobné firmy (s desítkami) , tak mamutí firmy (s tisíci) zaměstnanců. Vezměme za vysvětlovanou proměnnou veličinu hospodářský výsledek (zisk, případně ztrátu – bráno před zdaněním). Tato závisle proměnná může být vysvětlována řadou „interních" (počet zaměstnanců, rozsah kapitálu: strojní vybavení: jeřáby, bagry, dopravní prostředky, zásoby, apod.) indikátorů, jakož i některých „externích" (bankovní úvěry, pracovníci subdodavatelů) indikátorů. Je očekávatelné, že variabilita zisku (či ztráty) bude souviset (a to i kauzálně) s některou z těchto vysvětlujících proměnných (např. počtem zaměstnanců)[1]. Přitom ani tak (z hlediska statistického) nerozhoduje, zda se u závisle proměnné pohybujeme v kladných či záporných hodnotách, důležitější je jejich absolutní velikost.[2]. V chování závisle proměnné se přirozeně musí projevit jak variabilita oné vysvětlující proměnné, tak variabilita náhodné složky.[3] Úvaha o souvislosti hodnot y_t, případně ε_t vůči, řekněme veličině x_{tj*} je proto plně na místě.

---

[1] Je nepravděpodobné, že by velká firma měla výši hospodářského výsledku v desetitisících korun, resp.je nemožné, aby měla drobná firma zisk či ztrátu ve stamilionech korun.

[2] Je současně zřejmé, že těchto indikátorů může být i několik, často se vzájemnou závislostí.

[3] Na velikost variability náhodné složky usuzujeme přirozeně nepřímo, z velikosti reziduí.

1) GOLDFELDův - QUANDTův test[4]

Předpoklady

a) Kolísání náhodné složky je spojeno s proměnlivostí jedné určité vysvětlující proměnné (které se promítá do variability vysvětlované proměnné)

b) Náhodné složky mají $T-$ rozměrné normální rozdělení

Předpoklad Uživatel musí provést úvahu, která vysvětlující proměnná je nejtěsněji svázána s variabilitou náhodných složek.

Provedení testu

A) Všech T pozorování se uspořádá podle velikosti domnělé vysvětlující proměnné $x_{.j^*}$.[5] Souběžně se ze vzorku vynechá určitý počet prostředních T_2 pozorování[6], přičemž hodnota T_2 se volí tak, aby počet $T-T_2$ zbývajících aktivně uplatněných pozorování byl sudý. Formálně vyjádřeno

$$y=\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ \underline{y_{(T-T_2)/2}} \\ y_{(T-T_2)/2+1} \\ \dots \\ \underline{y_{T-T_2}} \\ y_{T-T_2+1} \\ \dots \\ y_T \end{pmatrix}, X=\begin{pmatrix} x_{11} & x_{12} & x_{13}^* & & x_{1,k} \\ x_{21} & x_{22} & x_{23}^* & & x_{2,k} \\ \dots & \dots & \dots & & \dots \\ \underline{x_{(T-T_2)/2,1}} & \underline{x_{(T-T_2)/2,2}} & \underline{x_{(T-T_2)/2,3}}^* & & \underline{x_{(T-T_2)/2,k}} \\ x_{(T-T_2)/2+1,1} & x_{(T-T_2)/2+1,2} & x_{(T-T_2)/2+1,3}^* & \dots\dots & x_{(T-T_2)/2+1,k} \\ \dots & \dots & \dots & & \dots \\ \underline{x_{T-T_2,1}} & \underline{x_{T-T_2,2}} & \underline{x_{T-T_2,3}}^* & & \underline{x_{T-T_2,k}} \\ x_{T-T_2+1,1} & x_{T-T_2+1,2} & x_{T-T_2+1,3}^* & & x_{T-T_2+1,k} \\ \dots & \dots & \dots & & \dots \\ x_{T1} & x_{T2} & x_{T3}^* & & x_{Tk} \end{pmatrix}$$

B) Zbývající pozorování rozdělíme do dvou (stejně početných) skupin o $1/2.(T-T_2)$ prvcích. První skupina obsahuje vzorek (níže značený y_1, X_1) s nízkými hodnotami proměnné $x_{.j^*}$, druhá skupina vzorek (níže značený y_2, X_2) s vysokými hodnotami x_{tj^*}. Schématicky vyjádřeno tedy:

$$y=\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_v \\ \tilde{y}_2 \end{pmatrix} \qquad X=\begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_v \\ \tilde{X}_2 \end{pmatrix}, \text{ kde}$$

„ovlnkované“ symboly u složek vektoru y a submatic matice X označují již datové struktury po provedení výše uvedeného přeskupení řádků.

---

[4] **Goldfeld,S., Quandt,R. : Nonlinear Methods in Econometrics. 1972.**

[5] Znamená to, že **přeskupíme** příslušné **řádky matice X a** rovněž **stejnolehlé prvky vektoru *y* tak, aby v j*-tém sloupci matice *X* byly hodnoty x_j* uspořádány vzestupně.**

[6] Jde o řádky s pořadovými čísly $(T-T_2)/2+1$ až $T-T_2$.

$$\tilde{y} = \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \dots \\ \tilde{y}_{(T-T_2)/2} \\ \tilde{y}_{T-T_2+1} \\ \dots \\ \tilde{y}_{T-1} \\ \tilde{y}_T \end{pmatrix} \qquad \tilde{X} = \begin{pmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \tilde{x}_{13} & \dots & \tilde{x}_{1k} \\ \tilde{x}_{21} & \tilde{x}_{22} & \tilde{x}_{23} & \dots & \tilde{x}_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ \tilde{x}_{(T-T_2)/2,1} & \tilde{x}_{(T-T_2)/2,2} & \tilde{x}_{(T-T_2)/2,3} & \dots & \tilde{x}_{(T-T_2)/2,k} \\ \tilde{x}_{T-T_2+1,1} & \tilde{x}_{T-T_2+1,2} & \tilde{x}_{T-T_2+1,3} & \dots & \tilde{x}_{T-T_2+1,k} \\ \dots & \dots & \dots & \dots & \dots \\ \tilde{x}_{T-1,1} & \tilde{x}_{T-1,2} & \tilde{x}_{T-1,3} & \dots & \tilde{x}_{T-1,k} \\ x_{T,1} & x_{T,2} & x_{T,3} & \dots & x_{Tk} \end{pmatrix}$$

C) Pro každou z obou skupin zvlášť odhadneme (prostou metodou nejmenších čtverců OLS) vektory parametrů β_1, β_2 a následně spočteme příslušné vektory reziduí e_1, e_2 jako $e_1 = \tilde{y}_1 - \tilde{X}_1.b_1$, resp. $e_2 = \tilde{y}_2 - \tilde{X}_2.b_2$. Spočteme součet čtverců reziduí v první skupině (označíme SSE_1) i ve druhé skupině (označíme SSE_2), tedy

$$SSE_1 = \sum_{t=1}^{(T-T_2)/2} e_1^{\,2} \qquad SSE_2 = \sum_{t=T-T_2+1}^{T} e_2^{\,2}$$

D) Ze získaných hodnot vytvoříme statistiku (při vzestupném uspořádání)

$$\boldsymbol{F} = \frac{\dfrac{\boldsymbol{SSE}_2}{(\boldsymbol{T} - \boldsymbol{T}_2 - 2\boldsymbol{k})/2}}{\dfrac{\boldsymbol{SSE}_1}{(\boldsymbol{T} - \boldsymbol{T}_2 - 2\boldsymbol{k})/2}} = \frac{\boldsymbol{SSE}_2}{\boldsymbol{SSE}_1}$$

Tato **statistika má** při splnění homoskedasticity ***Fisher-Snedecorovo rozdělení F*** **o** $(T - T_2 - 2k)/2$ **a** opět o $(T - T_2 - 2k)/2$ **stupni volnosti**.[7]

Statistikou F lze testovat přítomnost, zčásti i stupeň **heteroskedasticity**. **Pokud hodnota** $F_{empirická} < F^*_{tabulková}$ **přijímáme nulovou hypotézu** o homoskedasticitě (rozptyly ve skupinách se liší statisticky nevýznamně)

Pokud naopak hodnota $F_{empirická} > F^*_{tabulková}$, **zamítáme nulovou hypotézu ve prospěch alternativní** $\boldsymbol{H}_1$ a vyvodíme odtud, že v modelu je přítomná znatelná heteroskedasticita. Míru její síly rámcově posoudíme rozdílem F - F*.

Poznámky a modifikace

Počet vynechávaných pozorování $\boldsymbol{T}_2$ **ovlivňuje** ***průkaznost (sílu) testu***. Je-li $\boldsymbol{T}_2$ malé, pak rozdíly součtu čtverců reziduí mezi horní a dolní částí vzorku nemusí být výrazné a test k indikaci heteroskedasticity nepovede. Naopak, je-li $\boldsymbol{T}_2$ velké, je rozdíl průkaznější, ale počet stupňů volnosti může být (při malém $\boldsymbol{T}$) nízký a síla testu bude slabá. Proto se pro empirické úlohy běžného rozsahu ($\boldsymbol{T}$ = 30 až 60) doporučuje jako $\boldsymbol{T}_2$ vzít něco mezi $\boldsymbol{T}$/4 a $\boldsymbol{T}$/8 (např. $\boldsymbol{T}$/6).

[7] Odtud je vidět výhodnost volby stejného počtu pozorování pro horní a dolní část vzorku: F-statistika nabývá nejjednoduššího možného tvaru.

2) GLEJSERův test[8]

Předpoklady

Kolísání náhodné složky je spojeno s proměnlivostí určité vysvětlující proměnné (a následně též s variabilitou vysvětlované proměnné

Motivace testu

Sílu regresní závislosti mezi absolutními hodnotami reziduí a potenciální vlivovou proměnnou ověříme pomocí t-statistiky v jednoduché regresi mezi vektorem absolutizovaných reziduí a touto proměnnou.

Provedení testu

A) Formulujeme výchozí regresní závislost mezi vysvětlovanou proměnnou a maticí vysvětlujících proměnných

$$\boldsymbol{y}_t = \beta_1 + \beta_1.\boldsymbol{x}_{t1} + \beta_2.\boldsymbol{x}_{t2} + \ldots\beta_{j^*}.\boldsymbol{x}_{tj^*} + \ldots + \beta_k.\boldsymbol{x}_{tk} + \varepsilon_t$$

B) Určíme odhad parametrů $_{OLS}\hat{\beta}$ a následně vektor reziduí

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\hat{\beta}$$

C) Formulujeme variantně regresní vztahy mezi vektorem absolutních hodnot reziduí a vektorem předpokládané ovlivňující vysvětlující proměnné $\boldsymbol{x}_{.j^*}$. Regresní závislosti mohu být např. těchto typů

1)	$\lvert\boldsymbol{e}_t\rvert = \alpha_1 + \gamma_1\boldsymbol{x}_{tj^*}$	lineární
2)	$\lvert\boldsymbol{e}_t\rvert = \alpha_2 + \gamma_2\boldsymbol{x}_{tj^*}^{-1}$	reciproká
3)	$\lvert\boldsymbol{e}_t\rvert = \alpha_3 + \gamma_3.\ln\boldsymbol{x}_{tj^*}$	logaritmická
4)	$\lvert e_t\rvert = \alpha_4 + \gamma_4\sqrt{x_{tj*}}$	odmocninná

D) Spočtou se hodnoty regresních koeficientů v těchto regresích α_i, γ_i a zejména hodnoty t-statistik příslušných těmto parametrům $\boldsymbol{t}_{\alpha_i}, \boldsymbol{t}_{\gamma_i}$. Pokud je některá z hodnot $\boldsymbol{t}_{\gamma_i}$ statisticky významná, pak je to indikací příslušné (lineární nebo nelineární) korelovanosti vektoru absolutních hodnot reziduí s veličinou $\boldsymbol{x}_{.j^*}$. Nastane-li to u více formulovaných závislostí, pak vybereme tu závislost, kde jsou regresní parametry α_i, γ_i „nejvíce“ statisticky významné. Ta pak dává podnět pro konkrétní podobu transformace modelových veličin.[9]

Poznámky a modifikace

a) Podle toho, zda jsou statisticky významné oba regresní koeficienty α_i, γ_i nebo jen jeden γ_i, rozlišujeme heteroskedasticitu smíšenou nebo čistou.

b) ***Glejserův test*** **má zpravidla vyšší sílu ve srovnání s testem** ***Goldfelda a Quandta.***

---

[8] **Glejser, H.: A New test for Heteroscedasticity**. *Journal of the American Statistical Association* 64/1969 s.316-323.

[9] Často ale bývají rozdíly ve statistických významnostech parametrů u „podobných“ typů neliinearit velmi malé, takže rozlišení „charakteru heteroskedasticity“ je dost problematické.

3. SPEARMANův korelační koeficient (pořadové korelace)[10]

Předpoklady

a) Kolísání náhodné složky je spojeno s proměnlivostí určité vysvětlující proměnné (a následně též s variabilitou vysvětlované proměnné)

b) Náhodné složky mají $T-$ rozměrné normální rozdělení

Motivace testu

předpoklad a) uživatel by měl rozhodnout o tom, která proměnná je nejvíce svázána s variabilitou náhodných složek.

Provedení testu

A) Seřadíme hodnoty domněle ovlivňující nezávisle proměnné $\boldsymbol{x}_{i^*}$ podle velikosti (zpravidla od nejmenší po největší). Podle tohoto seřazení přeskupíme též hodnoty pozorování ostatních vysvětlujících veličin $\boldsymbol{x}_{i^*}$ (permutacemi řádků matice $\boldsymbol{X}$) a též hodnoty vysvětlované proměnné ve vektoru $\boldsymbol{y}_t$.

B) Formulujeme regresní závislost $\boldsymbol{y}_t$ na (přeskupené) vysvětlující proměnné $\boldsymbol{y}^p, \boldsymbol{x}_{.j}{}^p$:

$$\boldsymbol{y}^p{}_t = \beta_1 + \beta_1.\boldsymbol{x}_{t1}{}^p + \beta_2.\boldsymbol{x}_{t2}{}^p + \ldots \beta_{j^*}.\boldsymbol{x}_{tj}{}^p + \ldots + \beta_k.\boldsymbol{x}_{tk}{}^p + \varepsilon_t$$

a určíme *(stejně jako v Goldfeld-Quandtově testu bez ohledu na znaménka)* rezidua $|\boldsymbol{e}_t|$.

C) Vypočteme ***Spearmanův koeficient pořadové korelace*** podle vzorce

$$\boldsymbol{R}_S = 1 - \frac{6.\sum_{t=1}^{T} \boldsymbol{d}_t{}^2}{\boldsymbol{T}(\boldsymbol{T}^2 - 1)} \qquad \boldsymbol{t} = 1,2,\ldots,\boldsymbol{T}$$

kde $\boldsymbol{d}_t$ jsou *diference* v pořadích odpovídajících si (tj. ke stejnému řádku matice $\boldsymbol{X}$ patřících) dvojic $|\boldsymbol{e}_t|$ a $\boldsymbol{x}_{tj}$ *

Hodnoty *Spearmanova korelačního koeficientu* mají obdobnou interpretaci jako u klasického párového korelačního koeficientu. **Hodnoty blízké 0 naznačují nekorelovanost, hodnoty blízké krajním bodům intervalu přípustných hodnot** $< -1,1 >$ **pak udávají silnou zápornou, resp. kladnou korelovanost.** V tomto druhém případě je patrné, že v modelu je přítomná zřetelná heteroskedasticita. U veličiny $\boldsymbol{R}_s$ lze testovat, zda je hodnota P_s v základním souboru rovna nule.

Test významnosti veličiny $\boldsymbol{R}_s$ je založen na statistice

$$\boldsymbol{Q}_S = \boldsymbol{R}_s.\frac{\sqrt{\boldsymbol{T}-\boldsymbol{k}}}{\sqrt{1-\boldsymbol{R}_s{}^2}}$$

kde $\boldsymbol{T}-\boldsymbol{k}$ **je počet stupňů volnosti Studentova t-rozdělení**, kterou statistika $\boldsymbol{Q}_s$ má při nulové hypotéze $\boldsymbol{H}_0$, která odpovídá absenci heteroskedasticity.

---

[10] **Spearman, Ch.**

Poznámka 1 Často se v praxi zkusmo pro potlačení heteroskedasticity používá *logaritmická transformace* dat (ať už „mírnější“ s přirozeným nebo „ostřejší“ s dekadickým logaritmem. Stejnou úlohu může splnit také např. *odmocninná transformace*. Postup je obhajitelný, pokud to není v rozporu s poznatky ekonomické teorie charakterizujícími povahu závislostí veličin v uvažovaném regresním vztahu.

Poznámka 2 Spearmanův koeficient pořadové korelace není typickým ekonometrickým testem heteroskedasticity, protože vztahy mezi na jedné straně absolutizovanými reziduy a na druhé straně seřazenými hodnotami vybrané vysvětlující proměnné posuzuje „ordinálně“: tzn. závisí pouze na uspořádání prvků obou vektorů, nikoliv už na individuální rozdílech mezi nimi.

Poznámka 3

Je zřejmé, že minimální hodnotu nabude koeficient R_s tehdy, jestliže si “stejnolehlá pořadí“ v obou vektorech budou odpovídat a všechny diference budou nulové. Pak bude výraz $\sum_{t=1}^{T} d_t^{\,2}$ nulový a hodnota koeficientu R_s bude -1.

Na druhé straně maximální rozdíl mezi reziduy může být T-1, druhý největší T-3, další T-5 , až „minimální“ 1-T (ve čtvercích se absolutní hodnoty setřou). Půjde tedy o součet

$2.\left[(T-1)^2+(T-3)^2+(T-5)^2+(T-7)^2+.....+(T-2.T/2)^2\right]$ pro sudá T.

$2.\left[(T-1)^2+(T-3)^2+(T-5)^2+(T-7)^2+.....+(T-2.(T-1)/2)^2\right]$ pro lichá T.

$$2.\left[(T-1)^2+(T-3)^2+(T-5)^2+(T-7)^2+.....+(T-2.T/2)^2\right]=$$
$$2.\left[(T^2-2T+1)+(T^2-6T+9)+(T^2-10T+25)+(T^2-14T+49)+(T^2-18T+81)\right]=$$
$$2.\left[(T^2-1)-2T+2+(T^2-1)-6T+10+(T^2-1)-10T+26+(T^2-1)-14T+50+(T^2\right.$$

pro sudá T.

Součet T/2 členů (T^2-1) dává $T.(T^2-1)/2$

Součet T/2 členů aritmetické posloupnosti $-2T,-6T,-10T,-14T$ resp. tvaru $-\left[4(k-1)+2\right]$ k=1,2,,,,,T/2 dává $-\frac{T}{2}(2+4(\frac{T}{2}-1)+2)=-\frac{T}{2}(2T)=-T^2$

Konečně součet T/2 členů posloupnosti $2+10+26+50+$ tedy $1+(2k-1)^2$ dává...

Dokončit !

4. WHITEův obecný test [1980][11]

Abychom mohli formulovat příhodnější (a obecněji uplatnitelné) testy, je nezbytné specifikovat, přinejmenším v hrubé podobě, povahu heteroskedasticity.

Nejlepší situace by byla, pokud bychom mohli testovat obecnou hypotézu ve tvaru

$$\boldsymbol{H}_0 : \sigma_t^{\,2} = \sigma^2 \qquad \text{pro všechna } t \qquad \text{proti alternativě}$$
$$\boldsymbol{H}_1 : \sigma_t^{\,2} \neq \sigma^2 \qquad \text{aspoň pro jedno } t$$

Protože se však nacházíme v modelu, který má $\boldsymbol{T}$ obecně různých parametrů (rozuměno $\sigma_1^{\,2},\sigma_2^{\,2},....,\sigma_T^{\,2}$), je takovýto cíl obecně nedosažitelný.

Nicméně, Australan **Halbert WHITE** [1980] navrhl jisté řešení v podobě obecného testu. Jím navržený **test je založen na skutečnosti, že *OLS estimátor kovarianční matice* Σ_b je v případě výskytu heteroskedasticity obecně nekonzistentní**.

Nechť $e = (e_1, e_2, ..., e_T)$ je vektor OLS-reziduí a nechť OLS-estimátor rozptylu náhodných složek je $\hat{\sigma}^2 = (\boldsymbol{T} - \boldsymbol{k})^{-1} \sum \mathbf{e}_t^{\,2}$. **Při existenci homoskedasticity budou** oba následující estimátory

$$\tilde{\Sigma} = \frac{\Sigma \mathbf{e}_t^{\,2} \mathbf{x}_t . \mathbf{x}_t'}{\boldsymbol{T}} \quad \text{a} \quad \hat{\Sigma} = \frac{\hat{\sigma}^2 . \boldsymbol{X}' \boldsymbol{X}}{\boldsymbol{T}}$$

konzistentními estimátory téže kovarianční matice Σ.

Poznámka Přesná kovarianční matice estimátoru OLS (prosté metody nejmenších čtverců) v modelu zatíženého (jen) heteroskedasticitou, má tvar

$$\boldsymbol{Cov}(_{OLS}\hat{\beta}) = \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} (\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}) (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

přičemž **její** (konzistentní) **odhad lze pořídit** pomocí výrazu představovaného tzv.

WHITEovým estimátorem.

$$\hat{\boldsymbol{C}}\hat{\boldsymbol{o}}\hat{\boldsymbol{v}}(_{OLSWh}\hat{\beta}) = (\boldsymbol{X}'\boldsymbol{X})^{-1} (\sum_{t=1}^{T} \mathbf{e}_t^{\,2} \boldsymbol{x}_t \boldsymbol{x}_t^{\,\prime}) (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

Konvenční (a běžně užívaný) **OLS-estimátor** (v tomtéž modelu) **má** - jak známo - tvar

$$\hat{\boldsymbol{C}}\hat{\boldsymbol{o}}\hat{\boldsymbol{v}}(_{OLS}\hat{\beta}) = \boldsymbol{s}^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} \quad .$$

Za přítomnosti heteroskedasticity budou mít oba tyto estimátory tendenci se rozcházet vždy, kromě jediné zvláštní situace, takové, kdy by heteroskedasticita nebyla nijak závislá na obsahu matice $\boldsymbol{X}$ (to je však dost neobvyklá situace). Druhý estimátor (OLS) dává totiž konzistentní odhad $\boldsymbol{Cov}(_{OLS}\hat{\beta})$ jen tehdy, když heteroskedasticita v modelu přítomna není (je-li přítomna, pak je nekonzistentní).

Bude tedy možné založit test na rozdílu obou těchto odhadů kovarianční matice tím, že se bude testovat, zda je rozdíl mezi oběma estimátory statisticky významný.

---

[11] **White,H.: A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica* 48/1980b, s. 817-838.**

Konkrétní provedení testu

Prostá operační verze je dána veličinou $T.R^2$, kde R^2 **je čtverec koeficientu mnohonásobné korelace určený v regresi vektoru** e_t^2 **na všechny proměnné v matici** $\mathbf{x}_t \otimes \mathbf{x}_t'$[12].

Zapíšeme-li matici $\mathbf{x}_t \otimes \mathbf{x}_t'$ (rozměrů [Txk,k]) strukturně, dostaneme:

$$\mathbf{x}_t \otimes \mathbf{x}_t' = \begin{pmatrix} x_{t1}^2 & x_{t1}x_{t2} & x_{t1}x_{t3} & \dots & x_{t1}x_{tk} \\ x_{t1}x_{t2} & x_{t2}^2 & x_{t2}x_{t3} & \dots & x_{t2}x_{tk} \\ x_{t1}x_{t3} & x_{t2}x_{t3} & x_{t3}^2 & \dots & x_{t3}x_{tk} \\ \dots & \dots & \dots & \dots & \dots\dots \\ x_{t1}x_{tk} & x_{t2}x_{tk} & x_{t3}x_{tk} & \dots & x_{tk}^2 \end{pmatrix}$$

(Každý z „prvků“ této matice je T-členný vektor).

Matici vysvětlujících proměnných však předtím případně upravíme tak, že

a) vyškrtneme z ní všechny proměnné, které jsou redundantní (to jsou ty na „symetrických“ pozicích)

b) zařadíme do ní jedničkový vektor $\boldsymbol{x}_{.1}$, jestliže tam původně nebyl

$$\breve{\mathbf{x}}_t' \otimes \breve{\mathbf{x}}_t' = \begin{pmatrix} x_{t1}^2 & \times & \times & \dots & \times \\ x_{t1}x_{t2} & x_{t2}^2 & \times & \dots & \times \\ x_{t1}x_{t3} & x_{t2}x_{t3} & x_{t3}^2 & \dots & \times \\ \dots & \dots & \dots & \dots & \dots\dots \\ x_{t1}x_{tk} & x_{t2}x_{tk} & x_{t3}x_{tk} & \dots & x_{tk}^2 \end{pmatrix}$$

Regresní rovnice tedy vypadá takto:

$$e_t^2 = \alpha_0 + \alpha_1 x_{t1}^2 + \alpha_2 x_{t2}^2 + \dots + \alpha_k x_{tk}^2 + \alpha_{12} x_1 x_2 + \alpha_{13} x_1 x_3 + \alpha_{23} x_2 x_3 + \dots + \alpha_{k-1,k} x_{k-1} x_k + \zeta_t$$

Z ní spočteme koeficient determinace $\tilde{R}^2$ obvyklým způsobem:

$$\tilde{R}^2 = 1 - \frac{\sum \zeta_t^2}{a' Z' Z a}$$

Za nulové hypotézy (což je však širší situace než heteroskedasticita) **má statistika** $T.R^2$ **asymptoticky normální** χ^2 **- rozdělení o tolika stupních volnosti, kolik je počet vysvětlujících proměnných v pomocné regresi** (avšak po ubrání konstanty).

Pokud nejsou žádné redundance v $\mathbf{x}_t' \otimes \mathbf{x}_t'$ a $\mathbf{x}_t$ obsahuje konstantu, pak je počet stupňů volnosti d roven $d = k.(k+1)/2 - 1$.

[12] Jde o Kroneckerův součin matic (zde v kontextu vektorů)

Vlastnosti testu

Whiteův test je mimořádně obecný. Abychom jej mohli provést, nepotřebujeme činit žádné speciální předpoklady o tvaru heteroskedasticity. Byť je toto zřejmě předností, je to potenciálně i úskalí.

Test totiž **může odkrýt heteroskedasticitu, ale může také** pouze **odkrýt určitou specifikační chybu** (např. vynechání vysvětlující veličiny $\boldsymbol{x}_{.2}$ v běžné regresi). Může však také indikovat jiné chybné specifikace jako chybnou specifikaci funkce $\boldsymbol{E}[\boldsymbol{y}_t] = \boldsymbol{x}_t'.\beta$ nebo korelaci mezi $\boldsymbol{X}$ a $\boldsymbol{e}$ přítomnou v modelu se stochastickými regresory.

Sílu testu též **nelze přesvědčivě vyhodnotit** - vůči některým alternativám může být slabá. Slabinou je tedy *nekonstruktivnost testu* - zamítneme-li homoskedasticitu, nedávají výsledky testu návod, co učinit dále.

Avšak, pokud je výzkumník dost sebejistý, že se v modelu tyto problémy nevyskytují, může být test dostatečně účinný pro detekci heteroskedasticity. Test je podobný jiným testům jako LM-testu.

Modifikace

Hsieh [1983] modifikoval tento test pro testy *heteroskedasticity* a „*heterošpičatosti*“ a vyšetřoval jeho sílu při malých výběrech pomocí *Monte Carlo* experimentů.

5. BREUSCH-PAGANův (též GODFREYho) TEST [1979][13]

Goldfeld–Quandtův test lze pokládat za přiměřeně silný, pokud jsme schopni identifikovat proměnnou, podle které lze provést rozdělení datového vzorku (na ony dvě části, které pak slouží jako základ testu). Tento aspekt je však poněkud limitující. **V některých situacích je** totiž **proměnlivost disturbancí vázána ke skupině více, nejen k jediné vysvětlující proměnné.**

T.Breusch a ***A.Pagan*** navrhli test založený na principu Lagrangeových multiplikátorů, který testuje hypotézu

$$\sigma_t^2 = \sigma^2 \boldsymbol{f}(\alpha_0 + \alpha' \boldsymbol{z}_t) ,$$

kde $\boldsymbol{z}_t$ je vektor k-1 nezávisle proměnných (regresorů), $\boldsymbol{z}_t = (\boldsymbol{x}_{t2}, \boldsymbol{x}_{t3}, ..., \boldsymbol{x}_{tk})$.
Model je zřejmě homoskedastický, jestliže platí $\alpha = 0$.

Test lze provést jednoduchou **regresí** se statistikou **Lagrangeových multiplikátorů**[14]

$\boldsymbol{LM} = \frac{1}{2}$. vysvětlený součet čtverců v regresi $\frac{\boldsymbol{e}_t^2}{\boldsymbol{e}'\boldsymbol{e}/\boldsymbol{T}}$ na $\boldsymbol{z}_t$.

Provedení testu

Pro výpočetní účely vezmeme $\mathbf{Z}$ jako matici [T x (k+1)] pozorování proměnných $(1, \boldsymbol{z}_t)$ a nechť g je (sloupcový) vektor hodnot $\boldsymbol{g}_t = \frac{\boldsymbol{e}_t^2}{\boldsymbol{e}'\boldsymbol{e}/\boldsymbol{T}}$[15] . Testová statistika má tvar

$$BP = \frac{g'\left[Z(Z'Z)^{-1}Z'\right]g}{2}$$ maticově $$\mathbf{BP} = \frac{\mathbf{g}'_{[1,T]}\left[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\right]_{[T,T]}\mathbf{g}_{[T,1]}}{\mathbf{2}}$$

$$BP = \frac{1}{2}\cdot\left(\frac{Te_1^2}{e'e} \quad \frac{Te_2^2}{e'e} \quad .... \quad \frac{Te_T^2}{e'e}\right)\begin{pmatrix} \zeta_{11} & \zeta_{12} & .... & \zeta_{1T} \\ \zeta_{12} & \zeta_{22} & .... & \zeta_{2T} \\ .... & .... & .... & .... \\ \zeta_{1T} & \zeta_{2T} & .... & \zeta_{TT} \end{pmatrix}\cdot\begin{pmatrix} \frac{Te_1^2}{e'e} \\ \frac{Te_2^2}{e'e} \\ \frac{Te_3^2}{e'e} \\ ... \\ \frac{Te_T^2}{e'e} \end{pmatrix}$$

Lze ukázat, že **za platnosti nulové hypotézy (tj. při dodržení homoskedasticity), je veličina LM asymptoticky rozdělena jako χ^2 - rozdělení o *k* stupních volnosti**.

Poznámka Námitka proti testu spočívá v tom, že ***Breusch-Paganův test*** **je** příliš **citlivý na dodržení předpokladu o normalitě náhodných složek.**

---

[13] **Breusch,T., Pagan A.: A simple test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* 47/1979 s.1287-1294**

[14] **Breusch,T., Pagan A.: The LM Test and Its Applications to Model Specification in Econometrics. Review of Economic Studies 47/1980 s.239-254**

[15] **V interpretaci jde o podíl čtverce konkrétního rezidua a čtverce.“ průměrného rezidua“.**

6. KOENKER-BASSETův test

Proto ***R.Koenker*** **[1981][16] a** ***R.Koenker/G.Basset*** **[1982][17] navrhli, aby výpočet veličiny** ***LM*** **byl založen na robustnějším estimátoru** $\hat{\sigma}^2$ **rozptylu** σ^2 **náhodných složek než je** ***SSE*** **/** ***T*** **,** jmenovitě na (skalární) veličině[18]

$$V = \frac{1}{T}.\sum_{t=1}^{T}\left[e_t^{\,2} - \frac{e'e}{T}\right]^2$$

Není-li vektor náhodných složek ε rozdělený normálně, nebude rozptyl $\varepsilon_t^{\,2}$ roven $2\sigma^4$. Vezměme tedy sloupcový vektor $\vartheta = (e_1^{\,2}, e_2^{\,2}, ..., e_T^{\,2})'$ a nechť $\mathbf{\iota} = (1,1,....,1)$ je Tx1 vektor jedniček. Označme $\bar{\vartheta} = \frac{e'e}{T} = \frac{SSE}{T}$. Po této změně bude výpočet statistiky ***LM*** založen na statistice

$$KB = \frac{(\vartheta - \bar{\vartheta}\iota)'\left[Z(Z'Z)^{-1}Z'\right](\vartheta - \bar{\vartheta}\iota)}{V} \qquad \mathrm{KB} = \frac{(\vartheta - \bar{\vartheta}\iota)'_{[1,T]}\left[Z(Z'Z)^{-1}Z'\right]_{[T,T]}(\vartheta - \bar{\vartheta}\iota)_{[T,1]}}{V}$$

$$KB = \frac{1}{V}.\begin{pmatrix} e_1^{\,2} - \frac{e'e}{T} & e_2^{\,2} - \frac{e'e}{T} & .... & e_T^{\,2} - \frac{e'e}{T} \end{pmatrix}\begin{pmatrix} \zeta_{11} & \zeta_{12} & .... & \zeta_{1T} \\ \zeta_{12} & \zeta_{22} & .... & \zeta_{2T} \\ .... & .... & .... & .... \\ \zeta_{1T} & \zeta_{2T} & .... & \zeta_{TT} \end{pmatrix}.\begin{pmatrix} e_1^{\,2} - \frac{e'e}{T} \\ e_2^{\,2} - \frac{e'e}{T} \\ ... \\ e_T^{\,2} - \frac{e'e}{T} \end{pmatrix}$$

Za předpokladu normality bude takto modifikovaná statistika mít totéž asymptotické rozdělení jako ***Breusch-Paganova*** **statistika**. Při absenci normality jsou náznaky toho, že tento bude test silnější.

Poznámka: Rozdíl mezi tímto a Breusch –Paganovým testem je patrný v přístupu k vyhodnocování čtverců reziduí: Zatímco zde má vektor

θ složky rovny *rozdílům* čtverce t-tého rezidua oproti průměrnému čtverci rezidua

je u BP-testu vektor

ς složen z *podílů* čtverce t-tého rezidua vůči průměrnému čtverci rezidua.

Waldman [1983] ukázal, že pokud naplníme všechny sloupce v matici Z stejnými regresory jako v případě *White-ova testu*, budou oba testy obsahem výpočtu shodné.

Jak je patrné, **ani** ***Whiteův*****, ani** ***Breusch-Paganův*** **ani** ***Koenker-Bassetův*** **test nevyžadují specifikaci proměnné, na níž domněle závisí variabilita náhodných složek.** V tomto směru jsou zřetelně obecnější než testy *Goldfeldův-Quandův* a *Glejserův*.

---

[16] **Koenker,R.: A Note on Studentizing a Test of Heteroscedasticity.** ***Journal of Econometrics*** **17/1981, s. 107-112**

[17] **Koenker,R.Basset, G.: Robust Test for Heteroscedasticity Based on Regression Quantiles.** ***Econometrica*** **50/1982 s.43-61.**

[18] **Jde o** (výběrovou) **střední kvadratickou odchylku veličin** e_t^2