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 $\Sigma = \sigma^2 . I_T$, 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

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \quad \textbf{, kde}$$

jednotlivé prvky diagonální matice Σ^{-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 σ_t^2 **nějakými jejich (nestrannými) odhady s_t^2.**

Pokud bychom (výjimečně) znali veličiny σ_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:**

$$y_t^* = y_t / \sigma_t \,, \quad x_{tj}^* = x_{tj} / \sigma_t \quad j = 1,2,...,k; t = 1,2,...,T \text{ , \textbf{přičemž} } x_{t1} = 1$$

<u>Postupy navržené k indikaci heteroskedasticity v modelu</u>

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.**

1) <u>GOLDFELDův - QUANDTův test[1]</u>

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

[1] Goldfeld,S., Quandt,R. : Nonlinear Methods in Econometrics.

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

provedení testu:

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

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

B. Zbývající pozorování rozdělíme do dvou (stejně početných skupin o $1/2.(T - T_2)$ prvcích. První skupina obsahuje vzorek s nízkými hodnotami proměnné $x_{.j^*}$, druhá skupina vzorek s vysokými hodnotami x_{i^*}.

$$
y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_{T-T_2} \\ y_{T-T_2+1} \\ \dots \\ y_{T-T_2} \\ y_{T-T_2+1} \\ \dots \\ y_T \end{pmatrix}
\qquad
X = \begin{pmatrix}
x_{11} & x_{12} & y_{13} & & y_{1,k} \\
y_{21} & y_{22} & y_{23} & & y_{2,k} \\
\dots & \dots & \dots & & \dots \\
y_{T-T_2,1} & y_{T-T_2,2} & y_{T-T_2,3} & & y_{T-T_2,k} \\
y_{T-T_2+1,1} & y_{T-T_2+1,2} & y_{T-T_2+1,3} & \dots.. & y_{T-T_2+1,k} \\
\dots & \dots & \dots & & \dots \\
y_{T-T_2,1} & y_{T-T_2,2} & y_{T-T_2,3} & & y_{T-T_2,k} \\
y_{T-T_2+1,1} & y_{T-T_2+1,2} & y_{T-T_2+1,3} & & y_{T-T_2+1,k} \\
\dots & \dots & \dots & & \dots \\
y_{T,1} & y_{T,2} & y_{T,3} & & y_{T,k}
\end{pmatrix}
$$

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 ${}^1e = {}^1y - {}^1X.{}^1b$, resp. ${}^2e = {}^2y - {}^2X_1.{}^2b$. Spočteme součet čtverců reziduí v první skupině (označíme SSE_1) i ve druhé skupině (označíme SSE_2), tedy

$$SSE1 = \sum_{t=1}^{(T-T_2)/2} {}^1e^2 \qquad\qquad SSE1 = \sum_{t=T-T_2}^{T} {}^2e^2$$

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

$$F = \frac{\dfrac{SSE2}{(T-T_2-2k)/2}}{\dfrac{SSE1}{(T-T_2-2k)/2}} = \frac{SSE2}{SSE1}$$

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

Statistikou F lze testovat přítomnost, zčásti i stupeň heteroskedasticity: Pokud hodnota F (empirická) < F* (tabulková) akceptujeme nulovou hypotézu o homoskedasticitě (rozptyly ve skupinách se liší statisticky nevýznamně)

Pokud naopak hodnota F (empirická) > F* (tabulková) , zamítáme nulovou hypotézu ve prospěch alternativní H1 a vyvodíme odtud, že v modelu je přítomná znatelná heteroskedasticita. Míru její síly rámcově posoudíme rozdílem F - F*.

<u>poznámky a modifikace: </u>**Počet vynechávaných pozorování T_2 ovlivňuje průkaznost (sílu) testu. Je-li 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 T_2 velké , je rozdíl průkaznější, ale počet stupňů volnosti může být (při malém T) nízký a síla testu bude slabá. Proto se pro empirické úlohy běžného rozsahu (T= 30 až 60) doporučuje jako T_2 vzít něco mezi T/4 a T/8 (např. T/6).**

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

2) GLEJSERův test[3]

předpoklady:

a) 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

$$y_t = \beta_1 + \beta_1.x_{t1} + \beta_2.x_{t2} + \dots \beta_{j^*}.x_{tj^*} + \dots + \beta_k.x_{tk} + \varepsilon_t$$

B. Určíme odhad parametrů $_{OLS}\hat{\beta}$ **a následně vektor reziduí** $e = y - X\hat{\beta}$**.**

C. Formulujeme variantně regresní vztahy mezi vektorem absolutních hodnot reziduí a vektorem předpokládané ovlivňující vysvětlující proměnné $x_{.j^*}$ **Regresní závislosti mohou být např. těchto typů :**

(1) $\quad |e_t| = \alpha_1 + \gamma_1 x_{tj*}$

(2) $\quad |e_t| = \alpha_2 + \gamma_2 x_{tj*}^{-1}$

(3) $\quad |e_t| = \alpha_3 + \gamma_3 . \ln x_{tj*}$

(4) $\quad |e_t| = \alpha_1 + \gamma_1 \sqrt{x_{tj*}}$

D. Spočtou se hodnoty regresních koeficientů v těchto regresích α_i, γ_i **a zejména hodnoty t-statistik příslušných těmto parametrům** $t_{\alpha_i}, t_{\gamma_i}$**. Pokud je některá z hodnot** t_{γ_i} **statisticky významná, pak je to indikací příslušné (lineární nebo nelineární) korelovanosti vektoru absolutních hodnot reziduí s veličinou** $x_{.j^*}$**.**

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.**

poznámky a modifikace:

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

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

---

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

3. SPEARMANův korelační koeficient (pořadové korelace)[4].

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** x_{i^*} **podle velikosti** (zpravidla od nejmenší po největší). **Podle tohoto seřazení přeskupíme též hodnoty pozorování ostatních vysvětlujících veličin** x_{i^*} **(permutacemi řádků matice** X **) a též hodnoty vysvětlované proměnné ve vektoru** y_t **.**

B. **Formulujeme regresní závislost** y_t **na** (přeskupené) **vysvětlující proměnné** $y^p, x_{.j}{}^p$**:** $\quad y^p{}_t = \beta_1 + \beta_1.x_{t1}{}^p + \beta_2.x_{t2}{}^p + \ldots \beta_{j^*}.x_{tj}{}^p + \ldots + \beta_k.x_{tk}{}^p + \varepsilon_t$ **a určíme** (stejně jako v Goldfeld-Quandtově testu bez ohledu na znaménka) **rezidua** $|e_t|$**.**

C. **Vypočteme Spearmanův koeficient pořadové korelace podle vzorce :**

$$R_S = 1 - \frac{6.\sum_{t=1}^{T} d_t^2}{T(T^2-1)} \qquad \text{, t = 1,2,…,T , kde}$$

d_t **jsou diference v pořadích odpovídajících si** (tj. ke stejnému řádku matice X patřících) **dvojic** $|e_t|$ **a** 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 R_s lze testovat, zda je hodnota P_s v základním souboru rovna nule.

Test významnosti veličiny R_s **je založen na statistice**

$$Q_S = R_s.\frac{\sqrt{T-k}}{\sqrt{1-R_s^2}}$$

kde $T-k$ **je počet stupňů volnosti Studentova t-rozdělení, kterou statistika** Q_s **má při nulové hypotéze** H_0 **, která odpovídá absenci heteroskedasticity.**

<u>Poznámka 3</u> : Často se v praxi zkusmo pro potlačení heteroskedasticity používá logaritmická transformace dat (ať už „mírnější“ s přirozeným nebo

[4] Spearman, Ch.

„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.

4. WHITEův obecný test [1980][5]

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

Nejlepší by bylo, pokud bychom mohli testovat obecnou hypotézu ve tvaru

$H_0 : \sigma_t^2 = \sigma^2$ **pro všechna t** **proti alternativě**

$H_1 : \sigma_t^2 \neq \sigma^2$ **aspoň pro jedno t**

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

Nicméně , 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 nekonzistentní.**

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

$$\tilde{\Sigma} = \frac{\Sigma e_t^2 \boldsymbol{x_t}.\boldsymbol{x_t}'}{T} \quad \text{a} \quad \hat{\Sigma} = \frac{\hat{\sigma}^2 . X'X}{T}$$

konzistentními estimátory téže kovarianční matice Σ**.**

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

$$Cov(_{OLS}\hat{\beta}) = \sigma^2 .(X'X)^{-1}(X'VX)(X'X)^{-1} \quad ,$$

přičemž jejíž (konzistentní) odhad lze pořídit pomocí výrazu představujícího tzv.

WHITEův estimátor

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

Konvenční (běžně užívaný) OLS-estimátor (v tomtéž modelu) však má tvar

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

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

Za přítomnosti heteroskedasticity budou mít oba tyto estimátory tendenci se rozcházet vždy, kromě jediné zvláštní situace, kdy by heteroskedasticita nebyla nijak závislá na obsahu matice X (to je však dost neobvyklá situace). Druhý estimátor (OLS) dává totiž konzistentní odhad $Cov(_{OLS}\hat{\beta})$ 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.

<u>Konkrétní provedení testu:</u>

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 $\boldsymbol{x}_t \otimes \boldsymbol{x}_t'$[6].

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

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

(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 $x_{.1}$, jestliže tam původně nebyl.**

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

Regresní rovnice tedy vypadá takto:

$$e_t^2 = \alpha_0 + \alpha_1 x_{t1}^2 + \alpha_2 x_{t2}^2 + \ldots + \alpha_k x_{tk}^2 + \alpha_{12}x_1x_2 + \alpha_{13}x_1x_3 + \alpha_{23}x_2x_3 + \ldots + \alpha_{k-1,k}x_{k-1}x_k + \varsigma_t$$

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

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

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

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 $\boldsymbol{x}_t' \otimes \boldsymbol{x}_t'$ **a** $\boldsymbol{x}_t$ **obsahuje konstantu, pak je počet stupňů volnosti** d **roven** $d = k.(k+1)/2 - 1$.

<u>Vlastnosti testu:</u>

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 sice totiž může odkrýt heteroskedasticitu, ale může také pouze odkrýt určitou specifikační chybu (např. vynechání vysvětlující veličiny $x_{.2}$ **v běžné regresi. M*ůže však také indikovat jiné chybné specifikace jako chybná specifikace funkce*** $E[y_t] = x_t'.\beta$ ***nebo korelaci mezi*** X ***a*** 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.

<u>Modifikace:</u> 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][7]

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.

Breusch a Pagan navrhli test založený na principu Lagrangeových multiplikátorů, který testuje hypotézu

$$\sigma_t^2 = \sigma^2 f(\alpha_0 + \alpha' z_t) \ ,$$

kde z_t je vektor k-1 nezávisle proměnných (regresorů), $z_t = (x_{t2}, x_{t3}, ..., x_{tk})$.

Model je zřejmě ***homoskedastický***, ***jestliže platí*** $\alpha = 0$.

Test lze provést jednoduchou regresí se statistikou Lagrangeových multiplikátorů[8]:

$$\textbf{\textit{LM = 1/2. vysvětlený součet čtverců v regresi}} \quad \frac{e_t^2}{e'e/T} \quad \textbf{\textit{na}} \quad z_t \ .$$

Provedení testu:

Pro výpočetní účely vezmeme $\boldsymbol{Z}$ jako matici [T x k] pozorování proměnných $(1, z_t)$

a nechť $\boldsymbol{g}$ je (sloupcový) vektor hodnot $g_t = \dfrac{e_t^2}{e'e/T}$ [9]. Testová statistika má tvar

$$LM = \frac{g' . \left[Z(Z'Z)^{-1} Z' \right] . g}{2}$$

$$KB = \frac{1}{2} . \begin{pmatrix} \frac{Te_1^2}{e'e} \\ \frac{Te_2^2}{e'e} \\ \frac{Te_3^2}{e'e} \\ \cdots \\ \frac{Te_T^2}{e'e} \end{pmatrix} \begin{pmatrix} \zeta_{11} & \zeta_{12} & \cdots & \zeta_{1T} \\ \zeta_{12} & \zeta_{22} & \cdots & \zeta_{2T} \\ \cdots & \cdots & \cdots & \cdots \\ \zeta_{1T} & \zeta_{2T} & \cdots & \zeta_{TT} \end{pmatrix} . \begin{pmatrix} \frac{Te_1^2}{e'e} & \frac{Te_2^2}{e'e} & \cdots & \frac{Te_T^2}{e'e} \end{pmatrix}$$

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

Poznámka: Bylo namítáno, že ***Breusch-Paganův test je*** příliš ***citlivý na dodržení předpokladu o normalitě*** *náhodných složek*

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

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

[9] V podstatě jde o čtverec konkrétního rezidua od čtverce.“ průměrného rezidua“.

KOENKER-BASSETův test

Proto **Koenker [1981]**[10] a **Koenker/Basset [1982]**[11] navrhli, aby výpočet veličiny *LM* byl založen na robustnějším estimátoru $\hat{\sigma}^2$ rozptylu σ^2 náhodných složek než je SSE/T, jmenovitě na veličině[12]

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

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

$$KB = \frac{(\vartheta - \overline{\vartheta}\,\iota)'.\left[\boldsymbol{Z(Z'Z)^{-1}Z'}\right].(\vartheta - \overline{\vartheta}\,\iota)}{V} \qquad \text{neboli}$$

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

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. **Waldman [1983]** ukázal, že pokud naplníme všechny sloupce v matici $\boldsymbol{Z}$ 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.

---

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

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

12 Jde o (výběrovou) střední kvadratickou odchylku veličin $e_t^{\,2}$