

Heteroskedasticita

Problém *heteroskedasticity* se vztahuje k nesteržné 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 $\Sigma = \sigma^2 \cdot I_T$, 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 seříděných 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

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

kde jednotlivé prvky diagonální matice Σ^{-1} jsou reciproké hodnoty původních rozptylů, tj. $[\sigma_t^2]^{-1} = \sigma_t^{-2} = 1/\sigma_t^2$. 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:

$$\mathbf{y}_t^* = \mathbf{y}_t / \sigma_t, \quad \mathbf{x}_{tj}^* = \mathbf{x}_{tj} / \sigma_t \quad j = 1, 2, \dots, k; \quad t = 1, 2, \dots, T, \quad \text{přičemž } x_{t1} = 1$$

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 postupech (jmenovitě u Goldfeld-Quandtova a Glejserova) se objevuje úvaha spočívající v tom, že variabilita náhodných složek je v nějaké formě „spřažená“ 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 tisícovkami) 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 vyšší 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 nepřímou, z velikosti reziduí.

1) GOLDFELDův - QUANDTův test⁴

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 pozorování⁶, přičemž hodnota T_2 se volí tak, aby počet $T - T_2$ zbývajících aktivně uplatněných pozorování byl sudý. Formálně vyjádřeno

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

B) Zbývajících pozorování rozdělíme do dvou (stejně početných) skupin o $1/2 \cdot (T - T_2)$ 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) vysokými hodnotami $x_{.j}^*$. Schématicky vyjádřeno tedy:

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

„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 stejnohlé prvky vektorů 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 $T-T_2$.

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

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 $e_1 = \tilde{y}_1 - \tilde{X}_1 \cdot b_1$, resp. $e_2 = \tilde{y}_2 - \tilde{X}_2 \cdot b_2$. Spočteme součet čtverců reziduí v první skupině (označíme SSE_1) i ve druhé skupině (označíme SSE_2), tedy

$$SSE_1 = \sum_{i=1}^{(T-T_2)/2} e_{1,i}^2 \quad SSE_2 = \sum_{i=T-T_2}^T e_{2,i}^2$$

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

$$F = \frac{\frac{SSE_2}{(T-T_2-2k)/2}}{\frac{SSE_1}{(T-T_2-2k)/2}} = \frac{SSE_2}{SSE_1}$$

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.⁷

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 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 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. $T/6$).

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

2) GLEJSERŮV test⁸

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

$$\mathbf{y}_t = \beta_0 + \beta_1 \cdot \mathbf{x}_{t1} + \beta_2 \cdot \mathbf{x}_{t2} + \dots + \beta_j \cdot \mathbf{x}_{tj} + \dots + \beta_k \cdot \mathbf{x}_{tk} + \varepsilon_t$$

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

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

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

- 1) $|\mathbf{e}_t| = \alpha_1 + \gamma_1 \mathbf{x}_{tj^*}$
- 2) $|\mathbf{e}_t| = \alpha_2 + \gamma_2 \mathbf{x}_{tj^*}^{-1}$
- 3) $|\mathbf{e}_t| = \alpha_3 + \gamma_3 \cdot \ln \mathbf{x}_{tj^*}$
- 4) $|\mathbf{e}_t| = \alpha_1 + \gamma_1 \sqrt{\mathbf{x}_{tj^*}}$

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 $\mathbf{x}_{.j}$. 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

- 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.

3. SPEARMANŮV korelační koeficient (pořadové korelace)⁹

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_{it} podle velikosti (zpravidla od nejmenší po největší). Podle tohoto seřazení přeskupíme též hodnoty pozorování ostatních vysvětlujících veličin x_{jt} (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 :

$$y^p_t = \beta_1 + \beta_1 \cdot x_{t1}^p + \beta_2 \cdot x_{t2}^p + \dots + \beta_j \cdot x_{tj}^p + \dots + \beta_k \cdot x_{tk}^p + \varepsilon_t$$

a určíme (stejně jako v Goldfeld-Quandtově testu bez ohledu na znaménka) rezidua $|e_t|$.

C) Vypočteme Spearmanův koeficient pořadové korelace podle vzorce

$$R_s = 1 - \frac{6 \cdot \sum_{t=1}^T d_t^2}{T(T^2 - 1)} \quad t = 1, 2, \dots, T$$

kde d_t jsou *diference* v pořadích odpovídajících si (tj. ke stejnému řádku matice X patřících) dvojic $|e_t|$ a x_{jt}^*

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

$$Q_s = R_s \cdot \frac{\sqrt{T - k}}{\sqrt{1 - R_s^2}}$$

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

⁹ Spearman, Ch.

Poznámka Často se v praxi zkoumá 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.

4. WHITEŮV obecný test [1980]¹⁰

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

$$\begin{aligned} H_0 : \sigma_t^2 &= \sigma^2 && \text{pro všechna } t && \text{proti alternativě} \\ H_1 : \sigma_t^2 &\neq \sigma^2 && \text{aspoň pro jedno } t \end{aligned}$$

Protože se však nacházíme v modelu, který má T obecně různých parametrů (rozuměno $\sigma_1^2, \sigma_2^2, \dots, \sigma_T^2$), 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 $\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)$ je vektor OLS reziduí a necht' OLS estimátor rozptylu je $\hat{\sigma}^2 = (\mathbf{T} - \mathbf{k})^{-1} \sum \mathbf{e}_t^2$. Při existenci homoskedasticity budou estimátory

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

konzistentními estimátory téže kovarianční matice Σ .

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

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

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

$$\hat{\text{Cov}}(\text{OLS}_{WH} \hat{\beta}) = (\mathbf{X}' \mathbf{X})^{-1} \left(\sum_{t=1}^T \mathbf{e}_t^2 \mathbf{x}_t \mathbf{x}_t' \right) (\mathbf{X}' \mathbf{X})^{-1}$$

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

$$\hat{\text{Cov}}(\text{OLS} \hat{\beta}) = \hat{\sigma}^2 (\mathbf{X}' \mathbf{X})^{-1} \quad \square.$$

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 \mathbf{X} (to je však dost neobvyklá situace). Druhý estimátor

¹⁰ White, H.: A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica* 48/1980b, s. 817-838.

(OLS) dává totiž konzistentní odhad $\mathbf{Cov}_{(OLS)}(\hat{\beta})$ 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.

Konkrétní provedení testu

Prostá operační verze je dána veličinou $\mathbf{T.R}^2$, kde \mathbf{R}^2 je čtverec koeficientu mnohonásobné korelace určený v regresi vektoru \mathbf{e}_t^2 na všechny proměnné v matici $\mathbf{x}_t \otimes \mathbf{x}_t'^{11}$.

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

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

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

Matici vysvětlujících proměnných však předtím případně upravíme tak, že

- vyškrtneme z ní všechny proměnné, které jsou redundantní (to jsou ty na „symetrických“ pozicích)
- zařadíme do ní jedničkový vektor $\mathbf{x}_{\cdot 1}$, jestliže tam původně nebyl

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

Regresní rovnice tedy vypadá takto:

$$\mathbf{e}_t^2 = \alpha_0 + \alpha_1 \mathbf{x}_{t1}^2 + \alpha_2 \mathbf{x}_{t2}^2 + \dots + \alpha_k \mathbf{x}_{tk}^2 + \alpha_{12} \mathbf{x}_1 \mathbf{x}_2 + \alpha_{13} \mathbf{x}_1 \mathbf{x}_3 + \alpha_{23} \mathbf{x}_2 \mathbf{x}_3 + \dots + \alpha_{k-1,k} \mathbf{x}_{k-1} \mathbf{x}_k + \zeta_t$$

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

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

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

¹¹ Jde o Kroneckerův součin matic (zde v kontextu vektorů)

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 $d = k \cdot (k + 1) / 2 - 1$.

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 sice totiž může odkrýt heteroskedasticitu, ale může také pouze odkrýt určitou specifikační chybu (např. vynechání vysvětlující veličiny x_{i2} v běžné regresi. Může však také indikovat jiné chybné specifikace jako chybná specifikace funkce $E[\mathbf{y}_t] = \mathbf{x}_t' \cdot \beta$ nebo korelaci mezi \mathbf{X} a \mathbf{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]¹²

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

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

kde \mathbf{z}_t je vektor $k-1$ nezávisle proměnných (regresorů), $\mathbf{z}_t = (\mathbf{x}_{t2}, \mathbf{x}_{t3}, \dots, \mathbf{x}_{tk})$.

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

Test lze provést jednoduchou regresí se statistikou Lagrangeových multiplikátorů¹³

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

Provedení testu

Pro výpočetní účely vezmeme \mathbf{Z} jako matici $[T \times k]$ pozorování proměnných $(1, \mathbf{z}_t)$

a necht' \mathbf{g} je (sloupcový) vektor hodnot $\mathbf{g}_t = \frac{\mathbf{e}_t^2}{\mathbf{e}'\mathbf{e}/T}$ ¹⁴. Testová statistika má tvar

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

$$KB = \frac{1}{2} \cdot \begin{pmatrix} \frac{\mathbf{T}\mathbf{e}_1^2}{\mathbf{e}'\mathbf{e}} \\ \frac{\mathbf{T}\mathbf{e}_2^2}{\mathbf{e}'\mathbf{e}} \\ \frac{\mathbf{T}\mathbf{e}_3^2}{\mathbf{e}'\mathbf{e}} \\ \dots \\ \frac{\mathbf{T}\mathbf{e}_T^2}{\mathbf{e}'\mathbf{e}} \end{pmatrix} \begin{pmatrix} \zeta_{11} & \zeta_{12} & \dots & \zeta_{1T} \\ \zeta_{12} & \zeta_{22} & \dots & \zeta_{2T} \\ \dots & \dots & \dots & \dots \\ \zeta_{1T} & \zeta_{2T} & \dots & \zeta_{TT} \end{pmatrix} \begin{pmatrix} \frac{\mathbf{T}\mathbf{e}_1^2}{\mathbf{e}'\mathbf{e}} & \frac{\mathbf{T}\mathbf{e}_2^2}{\mathbf{e}'\mathbf{e}} & \dots & \frac{\mathbf{T}\mathbf{e}_T^2}{\mathbf{e}'\mathbf{e}} \end{pmatrix}$$

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

¹² Breusch, T., Pagan A.: A simple test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* 47/1979 s.1287-1294

¹³ Breusch, T., Pagan A.: The LM Test and Its Applications to Model Specification in Econometrics. *Review of Economic Studies* 47/1980 s.239-254

¹⁴ V podstatě jde o čtverec konkrétního rezidua od čtverce „průměrného rezidua“.

6. KOENKER-BASSETŮV test

Proto *Koenker* [1981]¹⁵ a *Koenker/Basset* [1982]¹⁶ navrhli, aby výpočet veličiny **LM** byl založen na robustnějším estimátoru $\hat{\sigma}^2$ rozptylu σ^2 náhodných složek než je SSE/T , jmenovitě na veličině¹⁷

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

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

$$KB = \frac{(\mathcal{g} - \bar{\mathcal{g}}\iota)' [\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] (\mathcal{g} - \bar{\mathcal{g}}\iota)}{V}$$

neboli

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

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 \mathbf{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*.

¹⁵ Koenker, R.: A Note on Studentizing a Test of Heteroscedasticity. *Journal of Econometrics* 17/1981 s. 107-112

¹⁶ Koenker, R. Basset, G.: Robust Test for Heteroscedasticity Based on Regression Quantiles. *Econometrica* 50/1982 s. 43-61.

¹⁷ Jde o (výběrovou) střední kvadratickou odchylku veličin e_t^2