Kapitel 08:

Heteroskedastizität

Moodle

Lehrbuch

Das klären wir in diesem Kapitel:

Heteroskedastizität und Unkorreliertheit bei bekannter Matrix Ω

FGLS unter Heteroskedastizität und Unkorreliertheit

Tests für Heteroskedastizität

Anwendung: Nachfrage nach Zigaretten

Heteroskedastizität

und

Unkorreliertheit

bei bekannter Matrix Ω

Heteroskedastizität und Unkorreliertheit bei bekannter Matrix Ω

Wir betrachten den Fall reiner Heteroskedastizität, wie er in Querschnittsmodellen häufig ist.

In dem Modell

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + u_i$$

sei also

$$Var(u_i|X) = \sigma_i^2 > 0 \quad \text{und } Cov(u_i, u_j|X) = 0 \text{ für } i \neq j$$

neu!

$$\Sigma(u \mid x) = \begin{pmatrix} \underbrace{Var(u_1 \mid x)}_{\sigma_1^2} & \underbrace{Cov(u_1, u_2 \mid x)}_{=0} & \cdots & \\ \underbrace{Cov(u_2, u_1 \mid x)}_{=0} & \underbrace{Var(u_2 \mid x)}_{\sigma_2^2} & \cdots & \\ \vdots & \vdots & & \\ & & & \underbrace{Var(u_n \mid x)}_{\sigma_n^2} \end{pmatrix}$$

$$= \sigma^2 \underbrace{\boxed{\begin{pmatrix} \frac{\sigma_1^2}{\sigma^2} & 0 & \cdots & 0 \\ 0 & \frac{\sigma_2^2}{\sigma^2} & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \frac{\sigma_n^2}{\sigma^2} \end{pmatrix}}}_{\Omega}$$

Ω positiv definit symmetrisch

$\Rightarrow$ Wurzel Z von Ω mit $Z \cdot Z' = \Omega$

$$Z = \begin{pmatrix} \frac{\sqrt{\lambda_1}}{\sqrt{\ }} & 0 & \cdots & 0 \\ 0 & \frac{\sqrt{\lambda_2}}{\sqrt{\ }} & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \frac{\sqrt{\lambda_n}}{\sqrt{\ }} \end{pmatrix}$$

$\Rightarrow$ Wurzel von Ω^{-1} erfüllt $P = Z^{-1}$; $P'P = \Omega$

$$P = \begin{pmatrix} \frac{\sigma}{\sigma_1} & 0 & & 0 \\ 0 & \frac{\sigma}{\sigma_2} & & 0 \\ \vdots & & & \\ 0 & 0 & & \frac{\sigma}{\sigma_n} \end{pmatrix}$$

Transformierte Daten:

$$\tilde{y} = P \cdot y = \begin{pmatrix} \frac{\sigma}{\sigma_1} \\ 0 \\ 0 \\ \\ 0 \end{pmatrix} \cdot y_1 + \begin{pmatrix} 0 \\ \frac{\sigma}{\sigma_2} \\ 0 \\ \vdots \\ 0 \end{pmatrix} y_2 + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{\sigma}{\sigma_n} \end{pmatrix} y_n$$

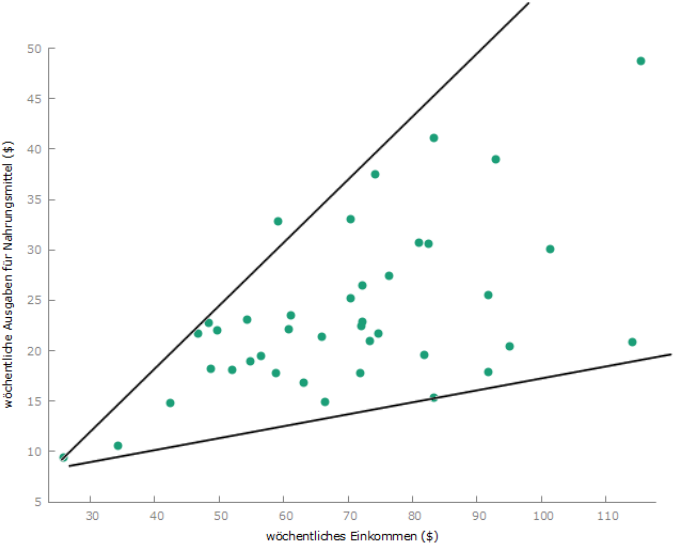
$$= \begin{pmatrix} \frac{\sigma}{\sigma_1} \cdot y_1 \\ 0 \\ \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{\sigma}{\sigma_2} \cdot y_2 \\ \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ 0 \\ \\ \frac{\sigma}{\sigma_n} \cdot y_n \end{pmatrix}$$

$$= \sigma \begin{pmatrix} \frac{y_1}{\sigma_1} \\ \frac{y_2}{\sigma_2} \\ \vdots \\ \frac{y_n}{\sigma_n} \end{pmatrix} = \tilde{y}$$

$$\tilde{X} = PX = \begin{pmatrix} \frac{1}{\sigma_1} & \frac{x_{11}}{\sigma_1} & \frac{x_{12}}{\sigma_1} & \cdots & \frac{x_{1k}}{\sigma_1} \\ \frac{1}{\sigma_2} & \frac{x_{21}}{\sigma_2} & \frac{x_{22}}{\sigma_2} & \cdots & \frac{x_{2k}}{\sigma_2} \\ \vdots & & & & \\ \frac{1}{\sigma_n} & \frac{x_{n1}}{\sigma_n} & \frac{x_{n2}}{\sigma_2} & \cdots & \frac{x_{nk}}{\sigma_n} \end{pmatrix}$$

Beispiel: Ausgaben für Lebensmittel und Einkommen

Daten entnommen aus Griffiths, Hill, Judge (1993) Table 5.2, Seite 182



GLS unter Heteroskedastizität und Unkorreliertheit

MLR: $\Omega = I_n$

Für $\Sigma(\mathbf{u}|X) = \sigma^2\Omega$ gilt:

$$\Omega = \begin{pmatrix} \frac{\sigma_1^2}{\sigma^2} & 0 & \dots & 0 \\ 0 & \frac{\sigma_2^2}{\sigma^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma_n^2}{\sigma^2} \end{pmatrix} \text{ und } P = \begin{pmatrix} \frac{\sigma}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{\sigma}{\sigma_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma}{\sigma_n} \end{pmatrix}$$

Da Ω die relativen Varianzen $\frac{\sigma_i^2}{\sigma^2}$ der Störterme enthält, hat P auf der Diagonalen die Kehrwerte der relativen Standardabweichungen.

Der Parameter σ^2 dient hier nur der Skalierung und wird nicht geschätzt. $tr(\Omega) = \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma^2} \overset{!}{=} n \Leftrightarrow \sum_{i=1}^{n} \sigma_i^2 = n \cdot \sigma^2$

Üblich: $tr(\Omega) = n$, also $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2$;

Die einzelnen Einträge $\frac{\sigma_i^2}{\sigma^2}$ sind dann im Durchschnitt gleich eins.

GLS Schätzer unter Heteroskedastizität

$\tilde{X} = PX$
$\tilde{y} = Py$

wobei
$P' \cdot P = \Omega^{-1}$

Der GLS Schätzer $\tilde{\boldsymbol{\beta}}$ ist definiert durch

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}\mathbf{y} \\ &\dots \\ &= \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{\mathbf{y}}\end{aligned}$$

mit den transformierten Daten

1. Zeile von X

2. Zeile von X

$$PX = \tilde{X} = \begin{pmatrix} \frac{\sigma}{\sigma_1}\mathbf{x}_{1\bullet} \\ \frac{\sigma}{\sigma_2}\mathbf{x}_{2\bullet} \\ \vdots \\ \frac{\sigma}{\sigma_n}\mathbf{x}_{n\bullet} \end{pmatrix} \text{ und } P\mathbf{y} = \tilde{\mathbf{y}} = \begin{pmatrix} \frac{\sigma}{\sigma_1}y_1 \\ \frac{\sigma}{\sigma_2}y_2 \\ \vdots \\ \frac{\sigma}{\sigma_n}y_n \end{pmatrix}$$

n. Zeile von X

$$\left(\sigma^2\tilde{X}'\tilde{X}\right)^{-1}\sigma^2\tilde{X}'\tilde{y}$$

$$\sigma^{-2}\cdot\sigma^2 = 1$$

$\tilde{\beta}$ ist invariant gegenüber σ^2

(hängt nicht von σ^2 ab)

$\tilde{\beta}$ invariant in Bezug auf σ^2

σ^2 ist lediglich ein Skalierungsparameter, er ist für die Berechnung von $\tilde{\beta}$ unerheblich.

Es gilt:

$$PX = \tilde{X} = \begin{pmatrix} \frac{\sigma}{\sigma_1}\mathbf{x}_{1\bullet} \\ \frac{\sigma}{\sigma_2}\mathbf{x}_{2\bullet} \\ \vdots \\ \frac{\sigma}{\sigma_n}\mathbf{x}_{n\bullet} \end{pmatrix} = \sigma \begin{pmatrix} \mathbf{x}_{1\bullet}/\sigma_1 \\ \mathbf{x}_{2\bullet}/\sigma_2 \\ \vdots \\ \mathbf{x}_{n\bullet}/\sigma_n \end{pmatrix}$$

und entsprechend

$$P\mathbf{y} = \tilde{\mathbf{y}} = \begin{pmatrix} \frac{\sigma}{\sigma_1}y_1 \\ \frac{\sigma}{\sigma_2}y_2 \\ \vdots \\ \frac{\sigma}{\sigma_n}y_n \end{pmatrix} = \sigma \begin{pmatrix} y_1/\sigma_1 \\ y_2/\sigma_2 \\ \vdots \\ y_n/\sigma_n \end{pmatrix}$$

Da $\left(\tilde{X}'\tilde{X}\right)^{-1}$ mit $\tilde{X}'\tilde{\mathbf{y}}$ multipliziert wird, kürzt sich σ^2 weg.

Gewichteter kleinste Quadrate Schätzer (WLS)

Wenn wir das Minimierungsproblem der gewöhnlichen kleinsten Quadrate auf die transformierten Daten anwenden, erhalten wir den gewichteten kleinste Quadrate Schätzer:

$$\min_{\mathbf{b}\in\mathbb{R}^{k+1}} \sigma^2 \sum_{i=1}^{n} \left(\underset{\text{gewichtete Residuum } \tilde{u}_i(b)}{\boxed{\frac{\overbrace{\boxed{y_i - \mathbf{x}_{i\bullet}\mathbf{b}}}^{\text{Residuum } \hat{u}_i(b)}}{\sigma_i}}} \right)^2$$

$$\Rightarrow \tilde{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n} \frac{\mathbf{x}_{i\bullet}\mathbf{x}'_{i\bullet}}{\sigma_i^2} \right]^{-1} \left[\sum_{i=1}^{n} \frac{\mathbf{x}_{i\bullet}\mathbf{y}}{\sigma_i^2} \right]$$

Beobachtungen mit großen Varianzen σ_i^2 erhalten in den Berechnungen ein kleineres Gewicht und haben daher einen kleineren Einfluss in den Schätzungen.

Der Parameter σ^2 hat keinen Einfluss auf die Minimum-Stelle $\tilde{\boldsymbol{\beta}}$.

Beispiel: Sparen und Einkommen

Originalmodell:

$$sav_i = \beta_0 + \beta_1 inc_i + u_i, \;\; Var(u_i|inc) = \sigma^2 inc_i$$

Transformiertes Modell:

$$\frac{sav_i}{\sqrt{inc_i}} = \beta_0 \frac{1}{\sqrt{inc_i}} + \beta_1 \frac{inc_i}{\sqrt{inc_i}} + \frac{u_i}{\sqrt{inc_i}}$$

Das transformierte Modell hat hier keine Konstante!

Es gilt:

$$E\left[\left(\frac{u_i}{\sqrt{inc_i}}\right)^2 \middle| inc\right] = \frac{1}{inc_i} E[u_i^2|inc] = \frac{1}{inc_i}\sigma^2 inc_i = \sigma^2$$

$$\Sigma(u \mid X) = \begin{pmatrix} \sigma^2 \, inc_1 & 0 & \cdots & 0 \\ 0 & \sigma^2 \, inc_2 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \sigma^2 \, inc_n \end{pmatrix}$$

$$= \sigma^2 \underbrace{\begin{pmatrix} inc_1 & 0 & \cdots & 0 \\ 0 & inc_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & inc_n \end{pmatrix}}_{\Omega}$$

Wurzel Z von Ω:

$$Z = \begin{pmatrix} \sqrt{\ln c_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\ln c_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\ln c_n} \end{pmatrix}$$

Inverse P von Z:

$$P = \begin{pmatrix} \frac{1}{\sqrt{\ln c_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\ln c_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\ln c_n}} \end{pmatrix}$$

$$\tilde{y} = P \cdot y = \begin{pmatrix} \frac{y_1}{\sqrt{inc_1}} \\ \frac{y_2}{\sqrt{inc_2}} \\ \vdots \\ \frac{y_n}{\sqrt{inc_n}} \end{pmatrix} \quad \tilde{y}_1$$

$$X = \begin{pmatrix} 1 & inc_1 \\ 1 & inc_2 \\ \vdots & \\ 1 & inc_n \end{pmatrix}$$

$$\tilde{X} = P \cdot X = \begin{pmatrix} \frac{1}{\sqrt{inc_1}} & \frac{inc_1}{\sqrt{inc_1}} \\ \frac{1}{\sqrt{inc_2}} & \frac{inc_2}{\sqrt{inc_2}} \\ \vdots & \vdots \\ \frac{1}{\sqrt{inc_n}} & \frac{inc_n}{\sqrt{inc_n}} \end{pmatrix} = \begin{array}{cc} \tilde{x}_{10} & \tilde{x}_{11} \\ \left(\begin{array}{c} \frac{1}{\sqrt{inc_1}} \\ \frac{1}{\sqrt{inc_2}} \\ \\ \frac{1}{\sqrt{inc_n}} \end{array} \right. & \left. \begin{array}{c} \sqrt{inc_1} \\ \sqrt{inc_2} \\ \\ \sqrt{inc_n} \end{array} \right) \end{array}$$

$$\tilde{y}_i = \beta_0 \cdot \tilde{x}_{i0} + \beta_1 \tilde{x}_{i1} + \tilde{u}_i$$

$$\boxed{\tilde{y}_i = \beta_0 \cdot \frac{1}{\sqrt{inc_i}} + \beta_1 \sqrt{inc_i} + \frac{u_i}{\sqrt{inc_i}}}$$

Heteroskedastizität bei Durchschnitten

Handelt es sich bei den einzelnen Beobachtungen um Durchschnitte, sollten sie mit der Wurzel ihrer jeweiligen Gruppengröße gewichtet werden.

Beispiel:

$$\overline{contrib_i} = \beta_0 + \beta_1 \overline{earns_i} + \beta_2 \overline{age_i} + \bar{u}_i$$

wobei i eine Gruppe (Firma, Stadt, etc.) der Größe m_i und $\bar{u}_i = \frac{1}{m_i} \sum_{t=1}^{m_i} u_{it}$.

Falls u_{it} iid mit $Var(u_{it}) = \sigma^2$, dann gilt $Var(\bar{u}_i) = \frac{1}{m_i}\sigma^2$.

Transformierte Daten:

$$\overline{contrib_i}\sqrt{m_i} = \beta_0\sqrt{m_i} + \beta_1 \overline{earns_i}\sqrt{m_i} + \beta_2 \overline{age_i}\sqrt{m_i} + \bar{u}_i\sqrt{m_i}$$

$$\Rightarrow Var(\bar{u}_i\sqrt{m_i}) = (\sqrt{m_i})^2 \frac{1}{m_i}\sigma^2 = \sigma^2$$

$$Var(\bar{v}_i) = \frac{1}{m_i}\sigma^2$$

$$\Sigma(v) = \begin{pmatrix} \frac{\sigma^2}{m_1} & 0 & 0 \\ 0 & \frac{\sigma^2}{m_2} & 0 \\ & & \\ 0 & 0 & \frac{\sigma^2}{m_n} \end{pmatrix}$$

$$\Rightarrow \Omega = \begin{pmatrix} \frac{1}{m_1} & 0 & 0 \\ 0 & \frac{1}{m_2} & 0 \\ & & \\ 0 & 0 & \frac{1}{m_n} \end{pmatrix}$$

$$P = \begin{pmatrix} \sqrt{m_1} & 0 & & 0 \\ 0 & \sqrt{m_2} & & 0 \\ & & & \\ 0 & 0 & & \sqrt{m_n} \end{pmatrix}$$

$$\tilde{y} = Py = \begin{pmatrix} \sqrt{m_1} \cdot y_1 \\ \sqrt{m_2} \cdot y_2 \\ \vdots \\ \sqrt{m_n} \cdot y_n \end{pmatrix}$$

$\tilde{X} = PX$ entsprechend

$$\tilde{x}_{i\cdot} = \left(\sqrt{m_i} \cdot 1 \,,\, \sqrt{m_i} \cdot x_{i1} \,,\, \dots \,,\, \sqrt{m_i} \cdot x_{ik}\right)$$

i. - Zeile von $\tilde{X}$

FGLS
unter
Heteroskedastizität
und
Unkorreliertheit

FGLS unter Heteroskedastizität und Unkorreliertheit

Falls σ_i erwartungstreu durch $\hat{\sigma}_i$ geschätzt werden kann, dann hat das transformierte Modell

$$\tilde{\boldsymbol{y}} = \tilde{X}\boldsymbol{\beta} + \tilde{\boldsymbol{u}} \text{ mit } \tilde{\boldsymbol{y}} = P\mathbf{y},\ \tilde{X} = PX,\ \tilde{\boldsymbol{u}} = P\mathbf{u}$$

eine einfache Struktur: Wir teilen jede Beobachtung i durch $\hat{\sigma}_i$.

$$\hat{P} = \begin{pmatrix} \frac{1}{\hat\sigma_1} & 0 & 0 \\ 0 & \frac{1}{\hat\sigma_2} & 0 \\ 0 & 0 & \frac{1}{\hat\sigma_n} \end{pmatrix}$$

$$\tilde{y}_i = \frac{y_i}{\hat{\sigma}_i},\ \tilde{x}_{ij} = \frac{x_{ij}}{\hat{\sigma}_i}$$

Schätzer für σ_i

Wir gewichten also jede Beobachtung umgekehrt proportional zu der Unsicherheit, mit der sie behaftet ist.

Falls also die Standardabweichungen σ_i erwartungstreu geschätzt werden können, ist GLS durchführbar (also FGLS).

Schätzung von σ_i für $i = 1, \ldots, n$?

Um alle n σ_i zu schätzen, benötigten wir aber n weitere Gleichungen, dies sind zu viele.

Deswegen stellen wir weitere Annahmen an die funktionale Form der Heteroskedastizität.

Zum Beispiel ist es plausibel anzunehmen, dass die Fehlervarianz in einfacher Weise von den Regressoren abhängt.

Für die Spezifikation einer Funktion $\sigma_i^2 = f(x_{i1}, \ldots, x_{ik})$ sind die Parameter dieser Funktion zu schätzen. Hieraus ergeben sich Schätzer für σ_i.

zum Beispiel $\sigma_i^2 = \sigma^2 \cdot inc_i$

Eigenschaften des GLS- und FGLS-Schätzers

- Der GLS-Schätzer ist BLUE (und damit effizient)
- Der FGLS-Schätzer basiert auf geschätzten Größen und ist daher nur asymptotisch effizienter als OLS.
- In kleinen Stichproben muss FGLS nicht zwangsläufig besser sein als OLS.
- Die Eigenschaften hängen davon ab, dass die korrekte Form für die Heteroskedastizität (in der Hilfsregression) unterstellt wird.

$\widehat{Var}(\tilde{\beta}_j)$ unter GLS

Betrachte zunächst den OLS-Schätzer $\hat{\boldsymbol{\beta}}$ der Einfachregression

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$\Rightarrow \hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}} = \beta_1 + \frac{\overline{xu} - \bar{x}\bar{u}}{\overline{xx} - \bar{x}\bar{x}} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Mit $Var(u_i) = \sigma_i^2$ und $Cov(u_i, u_s) = 0$ gilt:

$$Var(\hat{\beta}_1|x) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma_i^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$$

White (1980) zeigte, dass

$$plim\ \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\hat{u}_i^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})\right)^2} = \frac{E[(x_i - E[x])^2u_i^2]}{E[(x_i - E[x])^2]^2},$$

wobei $\hat{u}_i^2$ die OLS-Residuen sind.

Heteroskedastizitäts-robuste Standardfehler für $\hat{\beta}_j$

Im multiplen linearen Regressionsmodell $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$ zeigte White (1980), dass

$$\widehat{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{\left(\sum_{i=1}^{n} \hat{r}_{ij}^2\right)^2}$$

ein konsistenter Schätzer für die Varianz von $\hat{\beta}_j$ ist.

$\hat{r}_{ij}^2$: i-tes Residuum der Regression von $\mathbf{x}_{\bullet j}$ auf alle anderen Regressoren.

Die Wurzel dieses Schätzers heißt „heteroskedasticity-robust standard error" für $\hat{\beta}_j$.

Mit diesem Standardfehler ist der t-Test für große Stichproben valide.

Beispiel: Robuste Standardfehler in *wage1.xls*

$$
\begin{aligned}
\widehat{\text{lwage}} = \underset{\substack{(0{,}100009)\\ [0{,}109469]}}{0{,}321378} &+ \underset{\substack{(0{,}0553572)\\ [0{,}0571419]}}{0{,}212676}\,\text{marrmale} - \underset{\substack{(0{,}0578355)\\ [0{,}0587700]}}{0{,}198268}\,\text{marrfem} \\
&- \underset{\substack{(0{,}0557421)\\ [0{,}0571163]}}{0{,}110350}\,\text{singfem} + \underset{\substack{(0{,}00669450)\\ [0{,}00741467]}}{0{,}0789103}\,\text{educ} + \underset{\substack{(0{,}00524285)\\ [0{,}00513912]}}{0{,}0268006}\,\text{exper} \\
&- \underset{\substack{(0{,}000110426)\\ [0{,}000106340]}}{0{,}000535245}\,\text{exper}^2 + \underset{\substack{(0{,}00676200)\\ [0{,}00694092]}}{0{,}0290875}\,\text{tenure} - \underset{\substack{(0{,}000231243)\\ [0{,}000243685]}}{0{,}000533142}\,\text{tenure}^2
\end{aligned}
$$

(Standardfehler), [robuster Standardfehler], $n = 526$, $R^2 = 0{,}461$

Die robusten Standardfehler sind (meistens) geringfügig größer als die „normalen" Standardfehler.

Tests für Heteroskedastizität

Breusch-Pagan-Test für Heteroskedastizität

Nullhypothese Homoskedastizität:

$$\Sigma(\mathbf{u}|X) = \sigma^2 I \qquad \text{(MLR 5)}$$

Für die Varianz von u_i gilt (unter H_0:)

$$Var(u_i|X) = E[u_i^2|X] - \underbrace{E[u_i|X]^2}_{=0\ (MLR\ 4)} = E[u_i^2|X]$$

Die Nullhypothese impliziert also

$$E[u_i^2|X] = E[u_i^2] = \sigma^2 \quad \text{für alle } i = 1, \dots, n$$

Regression der quadrierten Residuen auf X:

$\hat{u}_i^2$: Schätzer für u_i^2

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \ldots + \delta_k x_{ik} + \epsilon_i$$

$$H_0 : \delta_1 = \delta_2 = \ldots = \delta_k = 0$$

Breusch-Pagan-Test für Heteroskedastizität

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \ldots + \delta_k x_{ik} + \epsilon_i$$

$$H_0 : \delta_1 = \delta_2 = \ldots = \delta_k = 0$$

Der Breusch-Pagan-Test ist ein globaler F-Test der Regression der quadrierten Residuen auf die Regressoren!

Sei $R^2_{\hat{u}^2}$ das Bestimmtheitsmaß dieser Regression.

$$F = \frac{R^2_{\hat{u}^2}/k}{(1 - R^2_{\hat{u}^2})/(n-k-1)}$$

Falls $\hat{u}^2$ gut durch die Regressoren erklärt wird, falls also $R^2_{\hat{u}^2}$ groß ist, so ist F groß und die Nullhypothese muss abgelehnt werden.

Alternativ: Lagrange-Multiplikator Statistik

$$LM = n \cdot R^2_{\hat{u}^2} \sim \chi^2_k$$

Zusammenfassung Breusch-Pagan Test

1. Schätze das Regressionsmodell $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$ per OLS und speichere die quadrierten Residuen $\hat{u}_i^2$, $i = 1, \ldots, n$.

2. Schätze das Regressionsmodell $\hat{\boldsymbol{u}}^2 = X\boldsymbol{\delta} + \boldsymbol{\epsilon}$ per OLS und speichere das Bestimmtheitsmaß $R^2_{\hat{u}^2}$.

3. Berechne entweder ...

   die F-Statistik mit $R^2_{\hat{u}^2}$ und berechne den entsprechenden p-Wert anhand der F-Verteilung mit k und $n - k - 1$ Freiheitsgraden

   ... oder ...

   die LM-Statistik $n \cdot R^2_{\hat{u}^2}$ und berechne den entsprechenden p-Wert anhand der χ^2-Verteilung mit k Freiheitsgraden.

4. Lehne die Nullhypothese der Homoskedastizität ab, falls der p-Wert kleiner ist als das Signifikanzniveau α.

Beispiel: Heteroskedastizität in der Immobilienpreise-Regression (*hprice*1)

$$\widehat{\text{price}} = \underset{(29{,}475)}{-21,7703} + \underset{(0{,}00064213)}{0,00206771}\,\text{lotsize} + \underset{(0{,}013237)}{0,122778}\,\text{sqrft} + \underset{(9{,}0101)}{13,8525}\,\text{bdrms}$$

$\hat{\delta}_1$

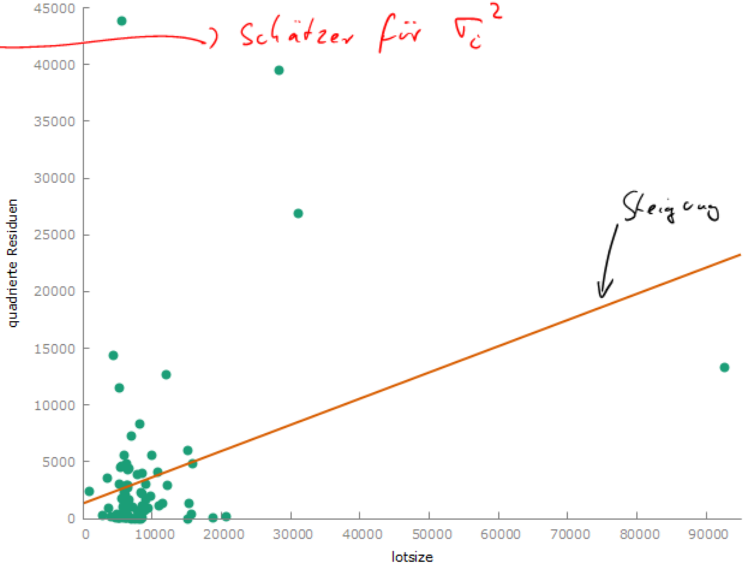
$$\widehat{\hat{u}^2} = \underset{(3259{,}5)}{-5522,79} + \underset{(0{,}071009)}{0,201521}\,\text{lotsize} + \underset{(1{,}4639)}{1,69104}\,\text{sqrft} + \underset{(996{,}38)}{1041,76}\,\text{bdrms}$$

$$n = 88 \quad R^2_{\hat{u}^2} = 0,1601 \quad F(3,84) = 5,3389 \quad p = 0,0020$$

Im vorliegenden Modell kann die Nullhypothese homoskedastischer Fehlerterme zu einem Signifikanzniveau $\alpha = 1\%$ mit einem Breusch-Pagan-Test zurückgewiesen werden.

$\hat{u}_i^2$ gegen $lotsize_i$ im Level-Level-Modell

$\Rightarrow$ Schätzer für σ_i^2



Steigung $\hat{\delta}_1 > 0$

Beispiel: Heteroskedastizität in der Immobilienpreise-Regression (*hprice*1) mit logarithmierten Daten

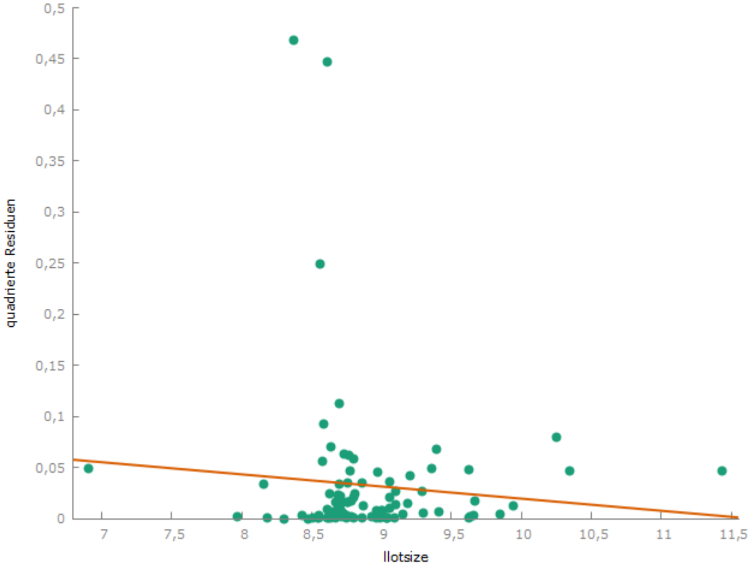
$$\widehat{\text{lprice}} = \underset{(0{,}65128)}{-1{,}29704} + \underset{(0{,}038281)}{0{,}167967}\,\text{llotsize} + \underset{(0{,}092865)}{0{,}700232}\,\text{lsqrft} + \underset{(0{,}027531)}{0{,}036958}\,\text{bdrms}$$

$$\widehat{\hat{u}^2} = \underset{(0{,}25786)}{0{,}509993} - \underset{(0{,}015156)}{0{,}007016}\,\text{llotsize} - \underset{(0{,}036767)}{0{,}062737}\,\text{lsqrft} + \underset{(0{,}010900)}{0{,}016841}\,\text{bdrms}$$

$$n = 88 \quad R^2_{\hat{u}^2} = 0{,}0480 \quad F(3, 84) = 1{,}4115 \quad p = 0{,}2451$$

In der logarithmischen Spezifikation kann die Nullhypothese homoskedastischer Fehlerterme nicht zurückgewiesen werden.

$\hat{u}_i^2$ gegen *llotsize*$_i$ im log-log-Modell



White-Test für Heteroskedastizität

Ähnlich zum Breusch-Pagan Test werden die quadrierten Residuen auf die unabhängigen Variablen regressiert. Hier werden jedoch zusätzlich alle quadrierten Variablen und alle Kreuzprodukte hinzugezogen:

$$\begin{aligned}\hat{u}_i^2 &= \delta_0 + & \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 \\ &+ & \delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 \\ &+ & \delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + \epsilon_i\end{aligned}$$

$$H_0 : \delta_1 = \delta_2 = \ldots = \delta_9 = 0$$

Vorteil:

Der White-Test erkennt mehr Abweichungen der Homoskedastizität als der Breusch-Pagan-Test

Nachteil:

Bei k Regressoren im Original-Modell existieren $\frac{k(k+3)}{2}$ Restriktionen, welche die Freiheitsgrade entsprechend reduzieren.

White-Test für Heteroskedastizität

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

$$\hat{y}^2 = \hat{\beta}_0 \cdot \hat{y} + \hat{\beta}_1 x_1 \hat{y} + \hat{\beta}_2 x_2 \hat{y} + \hat{\beta}_3 x_3 \hat{y}$$

Alternative Form des White-Tests:

$$\hat{\boldsymbol{u}}^2 = \delta_0 + \delta_1 \hat{\boldsymbol{y}} + \delta_2 \hat{\boldsymbol{y}}^2 + \boldsymbol{\epsilon}$$

Diese Regression testet die Abhängigkeit der quadrierten Residuen von den Regressoren, deren Quadraten und Kreuzprodukten indirekt, da die Prognose $\hat{\boldsymbol{y}}$ von $\boldsymbol{y}$ diese Ausdrücke alle implizit enthält.

Homoskedastizität $H_0 \cdot \delta_1 = 0 \,,\ \delta_2 = 0$

Alternativer White-Test für Heteroskedastizität in *hprice*1

$$\widehat{\text{price}} = \underset{(29{,}475)}{-21,7703} + \underset{(0{,}00064213)}{0,00206771}\,\text{lotsize} + \underset{(0{,}013237)}{0,122778}\,\text{sqrft} + \underset{(9{,}0101)}{13,8525}\,\text{bdrms}$$

$$\widehat{\hat{u}^2} = \underset{(8876{,}2)}{19071,6} - \underset{(53{,}317)}{119,655}\,\widehat{\text{price}} + \underset{(0{,}074596)}{0,208947}\,\widehat{\text{price}}^2$$

$$n = 88 \quad R^2 = 0,1849 \quad F(2,85) = 9,6388 \quad p = 0,0002$$

Die Nullhypothese kann zurückgewiesen werden.

$$\widehat{\text{lprice}} = \underset{(0{,}65128)}{-1,29704} + \underset{(0{,}038281)}{0,167967}\,\text{llotsize} + \underset{(0{,}092865)}{0,700232}\,\text{lsqrft} + \underset{(0{,}027531)}{0,036958}\,\text{bdrms}$$

$$\widehat{\hat{u}^2} = \underset{(3{,}3450)}{5,04683} - \underset{(1{,}1633)}{1,70922}\,\widehat{\text{lprice}} + \underset{(0{,}10099)}{0,145135}\,\widehat{\text{lprice}}^2$$

$$n = 88 \quad R^2 = 0,0392 \quad F(2,85) = 1,7328 \quad p = 0,1830$$

Die Nullhypothese kann nicht zurückgewiesen werden.

Multiplikative Heteroskedastizität

Unter multiplikativer Heteroskedastizität wird angenommen, dass

$$\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

$$\sigma_i^2 = e^{\delta_0+\delta_1 x_{i1}+\ldots+\delta_k x_{ik}} \; ,$$

also

$$\ln\left(\sigma_i^2\right) = \delta_0 + \delta_1 x_{i1} + \ldots + \delta_k x_{ik}$$

$\rightsquigarrow \hat{\delta} \rightsquigarrow \hat{\sigma}_i^2$

Wir schätzen $\boldsymbol{\delta}$ um eine Schätzung $\hat{\sigma}_i^2$ zu erhalten.

Problem: Wir kennen σ_i^2 nicht.

Lösung: Wir benutzen $\ln\left(\hat{u}_i^2\right)$ als Schätzer für $\ln\left(\sigma_i^2\right)$

Schätze Hilfsregression (mit OLS):

$$\ln\left(\hat{u}_i^2\right) = \delta_0 + \delta_1 x_{i1} + \ldots + \delta_k x_{ik}$$

um mit dieser Regression die Prognosen $\widehat{\ln\left(\hat{u}_i^2\right)}$ zu berechnen.

Multiplikative Heteroskedastizität

Diese Prognosen liefern uns eine Schätzung für σ_i^2:

$$\hat{\sigma}_i^2 = e^{\widehat{\ln\left(\hat{u}_i^2\right)}}$$

Transformation der Daten:

$$\tilde{x}_{ji} = \frac{x_{ji}}{\hat{\sigma}_i},\ \tilde{y}_i = \frac{y_i}{\hat{\sigma}_i}$$

Schätze nun das Modell

$$\tilde{y}_i = \beta_0 \frac{1}{\hat{\sigma}_i} + \beta_1 \tilde{x}_{i1} + \ldots + \beta_k \tilde{x}_{ik} + \tilde{u}_i$$

Das Resultat ist dann ein FGLS-Schätzer.

Anwendung: Nachfrage nach Zigaretten

Nachfrage nach Zigaretten (*smoke.xls*)

Abhängige Variable: cigs, n=807, $R^2 = 0,052737$

	Koeffizient	Std. Fehler	t-Quotient	p-Wert
const	−3,63987	24,0787	−0,1512	0,8799
lincome	0,880269	0,727784	1,210	0,2268
lcigpric	−0,750854	5,77334	−0,1301	0,8966
educ	−0,501498	0,167077	−3,002	0,0028
age	0,770694	0,160122	4,813	0,0000
agesq	−0,00902280	0,00174303	−5,176	0,0000
restaurn	−2,82508	1,11179	−2,541	0,0112

$\rightarrow$ speichere die quadrierten Residuen $\hat{u}_i^2$!

Nachfrage nach Zigaretten (*smoke.xls*)

Regressiere die quadrierten Residuen auf die Regressoren:

Abhängige Variable: $\hat{u}_i^2$, $n = 807$, $R^2_{\hat{u}^2} = 0,039973$

	Koeffizient	Std. Fehler	t-Quotient	p-Wert
const	−636,304	652,495	−0,9752	0,3298
lincome	24,6385	19,7218	1,249	0,2119
lcigpric	60,9768	156,449	0,3898	0,6968
educ	−2,38423	4,52753	−0,5266	0,5986
age	19,4175	4,33907	4,475	0,0000
agesq	−0,214790	0,0472335	−4,547	0,0000
restaurn	−71,1814	30,1279	−2,363	0,0184

$F(6, 800) = 5,551687$ p-Wert(F) = 0,000012

$\rightarrow$ der Breusch-Pagan-Test verwirft die Nullhypothese der Homoskedastizität.

Nachfrage nach Zigaretten (*smoke.xls*)

Nehme multiplikative Heteroskedastizität an und regressiere $\ln(\hat{u}_i^2)$ auf die Regressoren:

Abhängige Variable: $\ln(\hat{u}_i^2)$, $n = 807$

	Koeffizient	Std. Fehler	t-Quotient	p-Wert
const	−1,92071	2,56304	−0,7494	0,4538
lincome	0,291540	0,0774685	3,763	0,0002
lcigpric	0,195423	0,614540	0,3180	0,7506
educ	−0,0797036	0,0177844	−4,482	0,0000
age	0,204006	0,0170441	11,97	0,0000
agesq	−0,00239214	0,000185536	−12,89	0,0000
restaurn	−0,627013	0,118344	−5,298	0,0000

$\rightarrow$ speichere die prognostizierten Logarithmen der quadrierten Residuen $\widehat{\ln(\hat{u}_i^2)}$ und die Werte der Exponentialfunktion an den entsprechenden Stellen: $\hat{\sigma}_i^2 = exp(\widehat{\ln(\hat{u}_i^2)})$.

Nachfrage nach Zigaretten (*smoke.xls*) FGLS

Datentransformation: Teile alle Daten durch $\sqrt{\hat{\sigma}_i^2}$ und führe die OLS-Schätzung mit den transformierten Daten durch:

Abhängige Variable: tilde_cigs, n=807,

	Koeffizient	Std. Fehler	t-Quotient	p-Wert
tilde_const	5,63534	17,8031	0,3165	0,7517
tilde_lincome	1,29524	0,437012	2,964	0,0031
tilde_lcigpric	−2,94028	4,46014	−0,6592	0,5099
tilde_educ	−0,463446	0,120159	−3,857	0,0001
tilde_age	0,481947	0,0968082	4,978	0,0000
tilde_agesq	−0,00562721	0,000939479	−5,990	0,0000
tilde_restaurn	−3,46107	0,795505	−4,351	0,0000

Im Vergleich zur OLS-Schätzung liefert die FGLS-Schätzung kleinere Standardfehler; *lincome* ist nun auch signifikant!

Zusammenfassung

- Heteroskedastizität-robuste Standardfehler
- Breusch-Pagan Test für Heteroskedastizität
- White Test für Heteroskedastizität
- Alternativer White Test für Heteroskedastizität
- FGLS bei Heteroskedastizität