Heteroskedasticitet & heterogene effekter

Økonometri A

Bertel Schjerning

Program

Motivation (W8.1)

Heteroskedastisk-robust inferens (W8.2)

Tests for heteroskedasticitet (W8.3)

- Breusch-Pagen testet
- White's testet

Weighted least squares estimation (W8.4-W8.5)

- Weighted Least Squares og den lineære sandsynlighedsmodel
- Feasible Generalized Least Squares

Heterogene effekter (W9.3)

Motivation

Motivation

Den lineære regressionsmodel:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$$

Indtil videre har vi antaget:

MLR.5: Variansen af fejlleddet er konstant (homoskedasticitet)

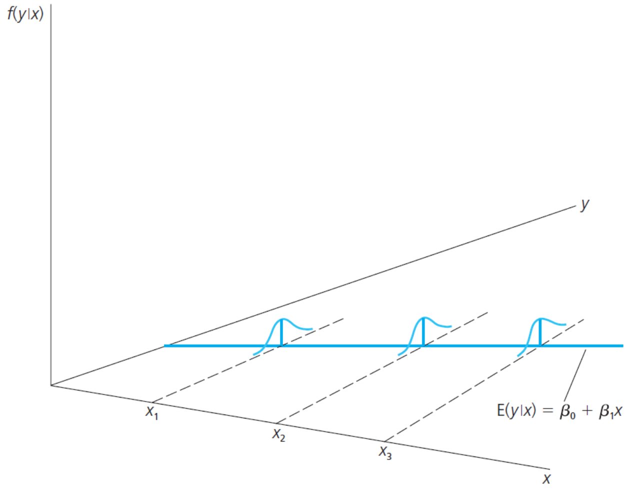
$$Var(u|x) = \sigma^2 \tag{1}$$

Denne antagelse er restriktiv.

Er MLR.5 *ikke* opfyldt, har modellen **heteroskedastiske fejlled**.

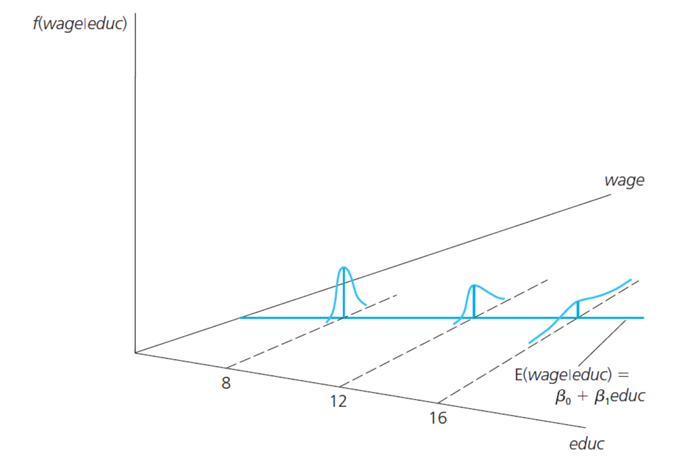
Homoskedasticitet

FIGURE 2.8 The simple regression model under homoskedasticity.



Heteroskedasticitet

FIGURE 2.9 Var(*wage*|*educ*) increasing with *educ*.



Definition og konsekvenser

Definition af heteroskedasticitet:

$$Var(u_i|x) = \sigma_i^2$$

Fejlledsvariansen kan variere på tværs af individer.

Når MLR.5 ikke er opfyldt (Heteroskedasticitet):

- Variansformlen for OLS estimatoren er ikke gyldig.
- Konfidensintervaller og t- og F-test er ikke gyldige.
- OLS forbliver middelret og konsistent, men er ikke længere BLUE eller asymptotisk efficient.

Mulige løsninger:

- **Robust standardfejl** (for at korrigere teststørrelser).
- **WLS** og **FGLS** (for at forbedre efficiens).

Hvornår opstår der heteroskedasticitet?

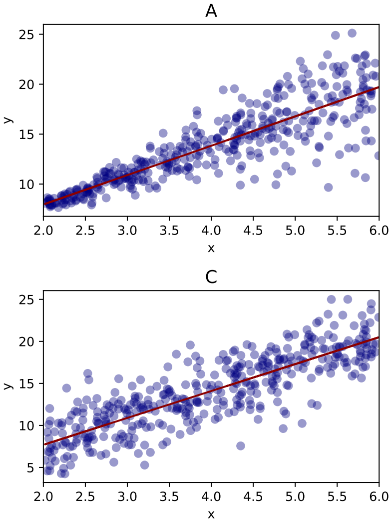
Heteroskedasticitet opstår ofte når:

- Observationer er meget forskellige i størrelse (fx lande, kommuner, virksomheder).
- Observationer er gennemsnit over varierende antal enheder (fx per capita eller karaktergennemsnit).
- Forkert funktionel form: y kan være heteroskedastisk, mens $\log(y)$ ikke er.
- Lineære sandsynlighedsmodel.

Heteroskedasticitet er specifikt for den enkelte model og population.

Heteroskedasticitet

Quiz: I hvilke af disse eksempler er der heteroskedasticitet?



Heteroskedastisk-robust inferens

Variansen af OLS estimatoren

I lektion 2 viste vi, at variansen på $\hat{\beta}_1$ i SLR er givet ved:

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

Hvordan gjorde vi det?

Variansen af OLS-estimatoren: Bevis (recap)

Ingredienser til beviset:

$$\text{OLS-estimatoren: } \hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n u_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Antagelser:

- SLR.1: Model: $y_i = \beta_0 + \beta_1 x_i + u_i$
- SLR.2: Observationerne er uafhængige.
- SLR.3: $\sum_{i=1}^n (x_i - \bar{x})^2 \neq 0$
- SLR.4: $E(u_i \mid x_i) = 0$
- SLR.5: $Var(u_i \mid x_i) = \sigma^2$

Regneregl:

- $Var(a(x)y + b(x) \mid x) = a(x)^2 Var(y \mid x)$

Variansen af OLS-estimatoren: Bevis (recap)

$$
\begin{aligned}
Var(\hat{\beta}_1 \mid X) &= Var\left(\frac{\sum_{i=1}^n u_i(x_i - \bar{x})}{SST_x} \middle| X\right) \\
&= \frac{1}{SST_x^2} Var\left(\sum_{i=1}^n u_i(x_i - \bar{x}) \middle| X\right) \quad \text{(brug SLR.2)} \\
&= \frac{1}{SST_x^2} \sum_{i=1}^n Var(u_i \mid X)(x_i - \bar{x})^2 \\
&= \frac{1}{SST_x^2} \sum_{i=1}^n \sigma^2 (x_i - \bar{x})^2 \quad \text{(brug SLR.5)} \\
&= \frac{\sigma^2}{SST_x^2} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}
\end{aligned}
$$

Variansen af OLS-estimatoren

Da σ^2 er ukendt, erstatter vi den med estimatet:

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2 \Longrightarrow \widehat{var}(\hat{\beta}_1|X) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

hvor $\hat{u}_i$ er OLS residualerne.

Bemærk: $\hat{\sigma}^2$ er ikke længere gyldig under heteroskedasticitet.

Variansen af OLS-estimatoren

Vi kan godt udregne variansen af OLS selvom $var(u_i|x) = \sigma_i^2$:

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

- SST_x kan ikke længere forkortes ud.
- Observationerne vægtes forskelligt: Større vægt til observationer med x langt fra $\bar{x}$.
- Men vi kender ikke σ_i^2.
- Derfor har vi brug for en ny estimator for $var(\hat{\beta}_1|x)$.

Variansen af OLS-estimatoren: White

Uden MLR.5 foreslog White at erstatte σ_i^2 med $\hat{u}_i^2$ i ligning (2).

White's heteroskedasticitisk-robust variansestimator i SLR:

$$\widehat{var}(\hat{\beta}_1|X) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$$

White har vist, at:

- Estimatoren giver en konsistent variansestimator ...
  uanset formen på heteroskedasticiteten.
  (asymptotisk resultat)

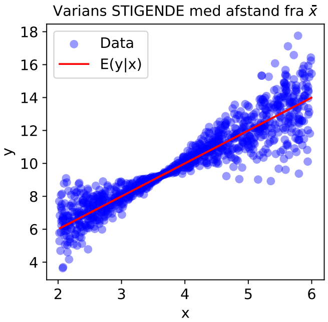
De fleste moderne økonometriske analyser bruger som udgangspunkt heteroskedasticitisk-robuste variansestimater.

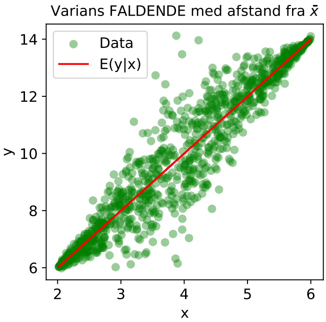
Hvad betyder heteroskedasticitet for variansen af OLS?

- Hvis σ_i^2 er størst for x_i langt fra gennemsnittet $\bar{x}$, vil dette øge variansen af $\hat{\beta}_1$.

$$\widehat{var}(\hat{\beta}_1|X) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$$

- Når observationer langt fra $\bar{x}$ har større $\hat{u}_i^2$, vægter de tungere i summen, hvilket øger variansen.





Heteroskedasticitisk-robust variansestimator: MLR

Regressionsmodel (MLR.1):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$

Vi antager, at variansmatricen for $\mathbf{u}$ er (her benyttes MLR.2):

$$\begin{aligned}\text{Var}(\mathbf{u}|\mathbf{X}) = \boldsymbol{\Omega} &= \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{pmatrix} \\ &= \text{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2)\end{aligned}$$

Heteroskedasticitisk-robust variansestimator i MLR: bevis

Ingredienser til bevis

1. Regneregler ($\mathbf{A}$ og $\mathbf{B}$ matricer, $\mathbf{x}$ stok. vektor, c skalar)
   1.1 $c\mathbf{AB} = \mathbf{A}c\mathbf{B} = \mathbf{AB}c$
   1.2 $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$ og $(\mathbf{A}')' = \mathbf{A}$
   1.3 $\mathbf{AA}^{-\mathbf{1}} = \mathbf{I}$ og $\mathbf{A}^{-\mathbf{1}}\mathbf{A} = \mathbf{I}$
   1.4 Hvis $\mathbf{A}$ er symmetrisk ($\mathbf{A} = \mathbf{A}'$), så er $\mathbf{A}^{-\mathbf{1}}$ (hvis den eksisterer) også symmetrisk.
   1.5 $\text{Var}(\mathbf{A}x) = \mathbf{A}\text{Var}(x)\mathbf{A}'$
2. Antagelser
   2.1 MLR.1: Model $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$
   2.2 MLR.3: $\mathbf{X}$ fuld rang
   2.3 MLR.4 og MLR.2: $E(\mathbf{u}|\mathbf{X}) = \mathbf{0}$
   2.4 MLR.2: $\text{Var}(\mathbf{u}|\mathbf{X}) = \mathbf{\Omega} = \text{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2)$
3. Estimator: $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{y}$

Variansen af OLS-estimatoren: bevis

1. **Indsæt sandt** y **i** $\hat{\beta}$ (MLR.1):

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

2. **Tag variansen af** $\hat{\beta}$ **givet** $\mathbf{X}$:

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}|\mathbf{X}) &= \mathrm{Var}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \mid \mathbf{X}\right) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathrm{Var}(\mathbf{u}|\mathbf{X})\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbf{\Omega}\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= n(\mathbf{X}'\mathbf{X})^{-1}\frac{1}{n}\mathbf{X}'\,\mathbf{\Omega}\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{\Sigma}(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

hvor

$$\mathbf{\Sigma} = \frac{1}{n}\mathbf{X}'\mathbf{\Omega}\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\mathbf{x_i}'\mathbf{x_i}.$$

Asymptotisk fordeling af OLS med robust varians

Under MLR.1-MLR.4 er OLS-estimatoren $\hat{\boldsymbol{\beta}}$ asymptotisk normalfordelt:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{a}{\sim} N\left(0, \boldsymbol{A}^{-1}\boldsymbol{\Sigma}\boldsymbol{A}^{-1}\right),$$

hvor

$$\boldsymbol{\Sigma} = \frac{1}{n}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\mathbf{x_i}'\mathbf{x_i},$$

$$\mathbf{A} = p\lim\frac{1}{n}(\mathbf{X}'\mathbf{X}) = p\lim\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x_i}'\mathbf{x_i}) = E(\mathbf{x_i}'\mathbf{x_i}),$$

som er ikke-singulære matricer (MLR.3), der *ikke afhænger af n*.

Den asymptotiske varians for $\hat{\boldsymbol{\beta}}$ går mod 0 med raten $\frac{1}{n}$:

$$\text{Var}(\hat{\beta}) = \frac{1}{n}\boldsymbol{A}^{-1}\boldsymbol{\Sigma}\boldsymbol{A}^{-1}.$$

Heteroskedasticitisk-robust variansestimator: MLR

White har vist, at

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 \mathbf{x_i}'\mathbf{x_i}$$

er en konsistent estimator for $\mathbf{\Sigma}$

$$plim\hat{\mathbf{\Sigma}} = plim\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 \mathbf{x_i}'\mathbf{x_i} = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 \mathbf{x_i}'\mathbf{x_i} = E(\sigma_i^2 \mathbf{x_i}'\mathbf{x_i}) = \mathbf{\Sigma}.$$

Robust varians-covarians estimator for OLS (justeret for frihedsgrader)

$$\widehat{\text{Var}}(\hat{\beta}|\mathbf{X}) = \frac{1}{n-k-1}((\mathbf{X}'\mathbf{X})/n)^{-1}\hat{\mathbf{\Sigma}}(\mathbf{X}'\mathbf{X}/n)^{-1}.$$

Med tilhørende heteroskedasticitisk-konsistente standardfejl (HCSE):

$$\widehat{se_{robust}}(\hat{\beta}|\mathbf{X}) = \sqrt{\text{diag}(\widehat{\text{Var}})}.$$

Heteroskedastiske-robuste teststørrelser

Hypoteser med én restriktion:

$$H_0 : \beta_j = a \text{ mod alternativet } H_1 : \beta_j \neq a.$$

Robust t-test:

$$t = \frac{\hat{\beta}_j - a}{HCSE_j},$$

hvor $HCSE_j$ er de heteroskedastisk-robuste standardfejl for $\hat{\beta}_j$.

- t-testet er asymptotisk normalfordelt under heteroskedasticitet.
- Ved små stikprøver kan teststørrelsen afvige fra normalfordelingen.

Forskellen er, at vi nu bruger robuste standardfejl i stedet for de sædvanlige OLS-standardfejl.

Let's Code!

- **Jupyter Notebook:** `09_heteroscedasticity_examples.ipynb`
- **Part 1:** Lønreggresion med robuste standardfejl

Heteroskedastiske-robuste teststørrelser

Hypotesetest med flere restriktioner:

$$H_0 : \mathbf{R}\beta = \mathbf{r} \quad \text{mod} \quad H_1 : \mathbf{R}\beta \neq \mathbf{r}.$$

Hvor:

- $\mathbf{R}$ er en $q \times (k+1)$ restriktionsmatrice,
- $\mathbf{r}$ er en $q \times 1$ vektor.

Teststørrelser forenkles under MLR.5:

- F-test (baseret på SSR eller R^2)
- Chow-testet
- LM-testet

Ikke gyldige under heteroskedasticitet.

Heteroskedastiske-robuste teststørrelser

Wald-teststørrelsen (Se Wooldridge E-4a):

$$W = (\mathbf{R}\hat{\beta} - \mathbf{r})' \left[\mathbf{R}\widehat{var}(\hat{\beta}|\mathbf{X})\mathbf{R}'\right]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}),$$

hvor:

- $\widehat{var}(\hat{\beta}|\mathbf{X})$ er den robuste varians for $\hat{\beta}$,
- $\mathbf{R}$ er en $q \times (k+1)$ restriktionsmatrice,
- $\mathbf{r}$ er en $q \times 1$ vektor.

Wald-testet:

- W er asymptotisk χ^2-fordelt med q frihedsgrader.
- $q \leq k+1$ (vi kan ikke lave flere restriktioner end parametre).

Heteroskedastiske-robuste teststørrelser

Hypotese med én restriktion på flere parametre:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \quad \text{mod} \quad H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}.$$

Eksempel med $q = 1$ restriktion på 2 parametre:

$$H_0 : \beta_1 = \beta_2 \quad \text{mod} \quad H_1 : \beta_1 \neq \beta_2$$

giver:

$$\mathbf{R} = (0\ 1\ \ -1)\,, \quad \boldsymbol{\beta} = (\beta_0\ \beta_1\ \beta_2)'\,, \quad \mathbf{r} = 0$$

Udregning af $\mathbf{R}\boldsymbol{\beta} - \mathbf{r}$:

$$\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \beta_1 - \beta_2$$

Heteroskedastiske-robuste teststørrelser

Hypotese med flere restriktioner:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \quad \text{mod} \quad H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}.$$

Eksempel med $q = 2$ restriktioner på 2 parametre:

$$H_0 : \beta_1 = 1 \text{ og } \beta_2 = 2 \quad \text{mod} \quad H_1 : \beta_1 \neq 1 \text{ eller } \beta_2 \neq 2$$

giver:

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \quad \mathbf{r} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \begin{pmatrix} \beta_1 - 1 \\ \beta_2 - 2 \end{pmatrix}$$

Let's Code!

- **Jupyter Notebook:**
  `09_heteroscedasticity_examples.ipynb`
- **Part 2:** Wald test med (og uden) robuste standardfejl

Tests for heteroskedasticitet

Test for heteroskedasticitet

Vi har følgende regressionsmodel:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u.$$

Vi antager, at MLR.1-MLR.4 er opfyldt.

Hypotese:

$$H_0 : Var(u|x) = E(u^2|x) = \sigma^2 \quad \text{(konstant)}.$$

Under H_0 vil $Var(u|x)$ være uafhængig af x.

Problem: Vi observerer ikke σ^2 direkte.

Løsning: Vi approximerer σ^2 med kvadrede OLS residualer $\hat{u}_i^2 \approx \sigma_i^2$

Grafiske test for heteroskedasticitet

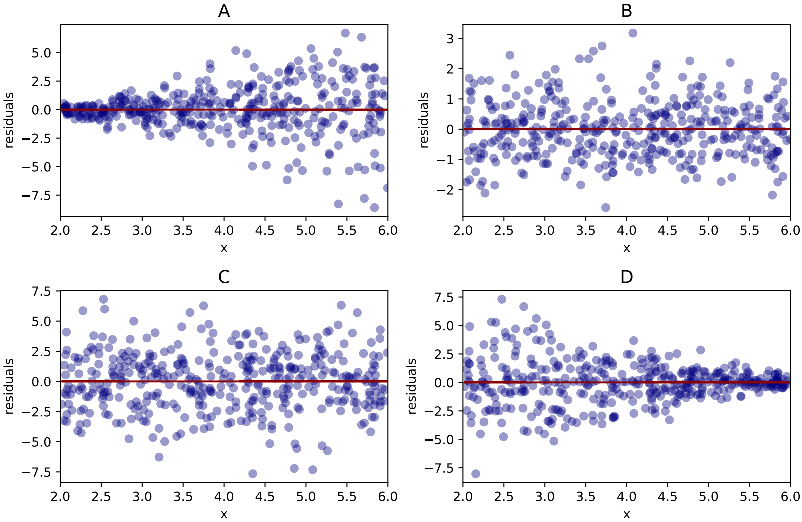
Procedure:

1. Estimer modellen med OLS og gem residualerne $\hat{u}$.
2. Plot residualerne $\hat{u}$ eller de kvadrerede residualer $\hat{u}^2$ mod x-variablene eller den prædikterede værdi $(\hat{y})$.
3. Se efter systematisk variation i residualerne:
   - Stigende varians af residualerne mod x-variable eller $\hat{y}$
   - Aftagende varians af residualerne mod x-variable eller $\hat{y}$
   - Andet mønster i variansen af residualerne

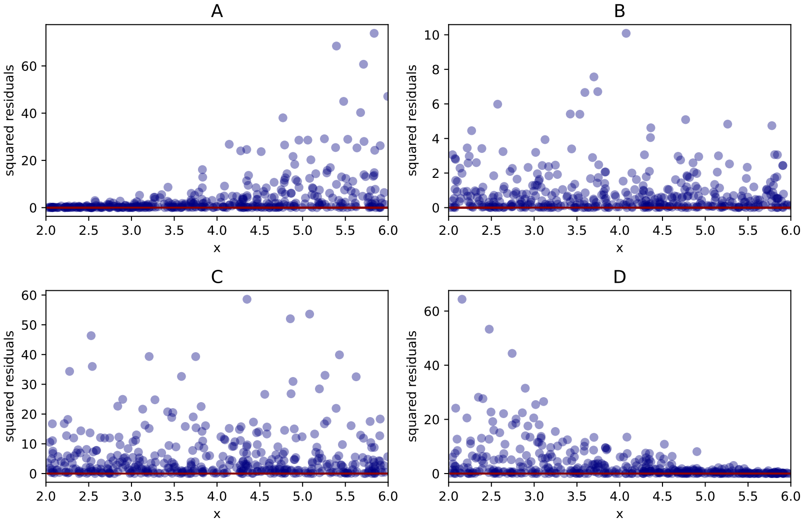
Bemærk: Kun muligt at plotte $\hat{u}$ ved små datasæt.

Ved store datasæt kan et bin-scatterplot anvendes (virker dog bedst for $\hat{u}^2$).

Grafiske test: Residual plot



Grafiske test: Kvadreret residual plot



Breusch-Pagan testet

Vi undersøger, om der er en lineær sammenhæng mellem variansen af fejlleddet (σ^2) og x-variablene:

$$E(\sigma^2|x) = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k.$$

Nulhypotesen for homoskedasticitet er:

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_k = 0.$$

Alternativhypotesen (heteroskedasticitet) er, at mindst ét $\delta_j \neq 0$ for $j = 1, \ldots, k$.

Ideen bag testet:

Erstat σ^2 med $\hat{u}^2$, og udfør testet som et F-test (eller LM-test) .

Breusch-Pagan testet

Manuel procedure:

1. Estimer OLS og gem residualerne $\hat{u}$.
2. Konstruer de kvadrerede residualer $\hat{u}^2$.
3. Estimer OLS på:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v,$$

4. Beregn LM- eller F-testet på baggrund af $R^2_{\hat{u}^2}$:

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

$$LM = R^2_{\hat{u}^2} n \overset{a}{\sim} \chi^2_{(k)}$$

Bemærk: F- og LM-tests er valide under H_0 om homoskedasticitet.

White's test

White foreslog en alternativ test for heteroskedasticitet.

Testen svarer til Breusch-Pagen testet, hvor vi tilføjer alle x'erne kvadreret samt alle krydsprodukterne

Dvs. med $k = 3$ estimerer vi følgende hjælpeligning

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

Generelt kræver White's test således estimation af en hjælpeligning med $1 + k^2$ variable.

White's test tester for ikke-lineær heteroskedasticitet. Fx hvis variansen er større både i for lave og høje x.

White's test

For at undgå for mange variable i hjælperegressionen findes der en simplere version af White's test:

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

Hypotese

$$H_0 : \delta_1 = \delta_2 = 0.$$

Teststørrelse: F- eller LM-testet.

Stata kan lave White's test ved post estimation:

estat hettest $\hat{y}$ $\hat{y}^2$, iid

Breusch-Pagan og White's test: Python eksempel

```
White's test for A: LM-stat = 70.91, p-værdi = 0.0000
White's test for B: LM-stat = 1.34, p-værdi = 0.5108
White's test for C: LM-stat = 0.59, p-værdi = 0.7463
White's test for D: LM-stat = 63.14, p-værdi = 0.0000

Breusch-Pagan test for A: LM-stat = 65.81, p-værdi = 0.0000
Breusch-Pagan test for B: LM-stat = 1.25, p-værdi = 0.9396
Breusch-Pagan test for C: LM-stat = 0.01, p-værdi = 1.0000
Breusch-Pagan test for D: LM-stat = 62.46, p-værdi = 0.0000
```

Let's Code!

- **Jupyter Notebook:**
  `09_heteroscedasticity_examples.ipynb`
- **Part 3:** Breusch-Pagan og White tests for heteroscedasticitet

Opsummering af test for heteroskedasticitet

- Hvis MLR.1-MLR.4 er opfyldt, kan man teste MLR.5 ved Breusch-Pagen test eller White's test.
- Hvis MLR.4 ikke er opfyldt, er Breusch–Pagen testet og White's test ikke gyldige.
- Tegn på heteroskedasticitet kan også forekomme, hvis MLR.4 ikke er opfyldt (selvom MLR.5 er opfyldt). Fx hvis vi anvender forkert funktionel form (mere om det senere)
- Generelt anvender vi som udgangspunkt heteroskedasticitet-robust inferens.

Weighted least squares estimation

Heteroskedasticitet og efficiens

Hvis MLR.5 ikke er opfyldt (men MLR.1-MLR.4 er opfyldt):

- OLS er **ikke** BLUE
- OLS er **ikke** asymptotisk efficient

Hvorfor er det sådan:

- OLS er konstrueret således, at summen af de kvadrerede residualer minimeres:
$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x ... - \hat{\beta}_k x_k)^2$$
- Alle observationerne får samme vægt.
- Når vi har heteroskedasticitet, er observationerne trukket fra fordelinger med forskellig varians.
- Observationer med mindre varians indeholder mere information om den underliggende sammenhæng mellem x og y.

Weighted Least Squares (WLS)

Hvorfor svarer OLS til at minimere de kvadrerede residualer?

$$\min_{\beta_j} R = \sum_{i=1}^{n} \hat{u}_i^2 =$$

Weighted Least Squares (WLS)

Da observationer med mindre varians indeholder mere information om den underliggende sammenhæng mellem x og y, er det optimalt at:

- Vægte hver observation med den inverse varians af fejlleddet $\frac{1}{\sigma_i^2}$.

Dvs. i stedet for at minimere de uvægtede kvadrade afvigelser ønsker vi at minimere

$$\sum_{i=1}^{n} \hat{u}_i^2 \frac{1}{\sigma_i^2} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x ... - \hat{\beta}_k x_k)^2 \frac{1}{\sigma_i^2}$$

svarende til moment betingelsen

$$\sum_{i=1}^{n} x_{ij} u_i \frac{1}{\sigma_i^2} = 0$$

Hvorfor løser det problemet med heteroskedasticitet?

Weighted Least Squares (WLS)

Antag at vi kender formen af heteroskedasticiteten

$$var(u_i|x) = \sigma^2 h(x)$$

- σ^2 er ukendt.
- $h(x)$ er en kendt funktion og $h(x) > 0$ for alle x.

Weighted Least Squares (WLS)

Betragt fx følgende model:

$$
\begin{aligned}
sav_i &= \beta_0 + \beta_1 inc_i + u_i \\
var(u_i|inc) &= \sigma^2 inc_i
\end{aligned}
.
$$

I dette tilfælde er $h(x) = inc_i$:

- Variansen af fejllleddet er proportional med indkomsten.
- Variansen er positiv (så længe indkomsten er positiv).
- Standardafvigelsen på fejlleddet er $\sigma\sqrt{inc_i}$.

Her vil vi bruge $1/h(x) = 1/inc_i$ som vægt.

Weighted Least Squares (WLS)

Vægte som løsning på problemet med heteroskedasticitet.

$$\begin{aligned}\min \sum_{i=1}^{n} \hat{u}_i^2 \frac{1}{h(x)} &= \sum_{i=1}^{n} (sav_i - \hat{\beta}_0 - \hat{\beta}_1 inc_i)^2 \frac{1}{inc_i} \\ &= \sum_{i=1}^{n} \left(\frac{sav_i}{\sqrt{inc_i}} - \hat{\beta}_0 \frac{1}{\sqrt{inc_i}} - \hat{\beta}_1 \frac{inc_i}{\sqrt{inc_i}} \right)^2 \qquad (3)\end{aligned}$$

Bemærk nu at minimering af (3) svarer til at vi estimerer

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

Dvs. den oprindelige model divideret igennem med $\sqrt{inc_i}$

Weighted Least Squares (WLS)

Den transformerede regressionsmodel:

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

- har to "nye" variable og intet konstantled.
- Samme parametre: β_0 og β_1.
- Nyt fejlled $\frac{u_i}{\sqrt{inc_i}}$.

Variansen af det nye fejlled

$$var\left(\frac{u_i}{\sqrt{inc_i}}|inc\right) = \left(\frac{1}{\sqrt{inc_i}}\right)^2 var(u_i|inc) = \frac{1}{inc_i}\sigma^2 inc = \sigma^2$$

Variansen af det nye fejlled er konstant (homoskedastisk).

Weighted Least Squares (WLS)

Generel procedure Hvis vi kender formen på heteroskedasticiteten, kan vi lave en transformation af modellen, hvor det nye fejlled er homoskedastisk.

Model:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$$
$$Var(u_i|x) = \sigma^2 h(x_{i1}, x_{i2}, .., x_{ik}) = \sigma^2 h_i$$

- Da h er kendt, kan vi beregne h_i for alle observationer.
- Den transformerede model fås ved at dividere med $\sqrt{h_i}$.

OBS Parameterestimaterne fra den transformerede model skal fortolkes som i den oprindelige model.

Weighted Least Squares (WLS)

Den transformerede model

$$\frac{y_i}{\sqrt{h_i}} = \beta_0 \frac{1}{\sqrt{h_i}} + \beta_1 \frac{x_{i1}}{\sqrt{h_i}} + ... + \beta_k \frac{x_{ik}}{\sqrt{h_i}} + \frac{u_i}{\sqrt{h_i}}$$

Det nye fejlled er $\frac{u}{\sqrt{h_i}}$ og har følgende egenskaber:

- Betinget middelværdi

$$E\left(\frac{u_i}{\sqrt{h_i}}|x\right) = \frac{1}{\sqrt{h_i}} E(u_i|x) = 0.$$

  under antagelse af at MLR.4 er opfyldt.

- Betinget varians:

$$Var\left(\frac{u_i}{\sqrt{h_i}}|x\right) = \left(\frac{1}{\sqrt{h_i}}\right)^2 Var(u_i|x) = \left(\frac{1}{\sqrt{h_i}}\right)^2 \sigma^2 h_i = \sigma^2.$$

  Altså homoskedastisk fejlled.

Weighted Least Squares (WLS)

Den transformerede model

$$\frac{y_i}{\sqrt{h_i}} = \beta_0 \frac{1}{\sqrt{h_i}} + \beta_1 \frac{x_{i1}}{\sqrt{h_i}} + \ldots + \beta_k \frac{x_{ik}}{\sqrt{h_i}} + \frac{u_i}{\sqrt{h_i}}$$

- Ingen konstantled (i stedet en ny variabel $\frac{1}{\sqrt{h_i}}$).
- De nye variable har sjældent en fortolkning.
- De samme parametre som i den oprindelige model.
- Det er nu muligt at lave efficient estimation i den transformerede model (da MLR.1-MLR.5 gælder).

I praksis transformer vi "aldrig" modellen manuelt:

- WLS kan laves i STATA ved at anvende vægtoptionen:
- regress y x [aw=$\frac{1}{h_i}$]

Weighted Least Squares og den lineære sandsynlighedsmodel

Regressionsmodellen er

$$\Pr(y = 1|x) = p(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + .. + \beta_k x_k$$
$$var(u|x) = p(x) \cdot (1 - p(x)).$$

$h(x)$ er (delvist) altså kendt.

WLS procedure

1. Estimer LPM vha. standard OLS.
2. Konstruer $h_i = p(x_i) \cdot (1 - p(x_i)) = \hat{y}_i(1 - \hat{y}_i)$
3. Dette kan være problematisk hvis $\hat{y}_i < 0$ eller $\hat{y}_i > 1$.
   Ad hoc korrektion:
   - $\hat{y}_i = 0.01$ hvis $\hat{y}_i \leq 0$
   - $\hat{y}_i = 0.99$ hvis $\hat{y}_i \geq 1$
4. Estimer LPM vha. WLS (OLS med vægte $1/h_i$).

Feasible Generalized Least Squares (FGLS)

Hvad hvis vi ikke kender formen på heteroskedasticiteten?

h er ukendt, men vi kan estimere den og benytter $\hat{h}$ i stedet for.

Der er mange måder at modellere h på:

- En måde er at antage

$$var(u_i|x) = \sigma^2 \exp(\delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} .. + \delta_k x_{ik}).$$

- Variansen er proportional med $\exp(\delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} .. + \delta_k x_{ik})$
- Denne funktionelle form sikrer at variansen altid er positiv.
- Hvis vi kan estimere $\delta_0, \delta_1, .., \delta_k$, kan vi beregne $\hat{h}_i$ for alle observationer.

Feasible Generalized Least Squares (FGLS)

Procedure

1. Estimer modellen med (standard) OLS.
2. Bestem residualerne $\hat{u}_i$ og beregn $\log(\hat{u}_i^2)$.
3. Estimer følgende model med (standard) OLS

$$\log(\hat{u}_i^2) = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} .. + \delta_k x_{ik} + e.$$

4. Beregn de predikterede værdier $\hat{g}_i = \widehat{log(\hat{u}_i^2)}$.
5. Udregn $\hat{h}_i = \exp(\hat{g}_i)$.
6. Estimer modellen

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$$

med WLS hvor vægtene er $1/\hat{h}_i$.

Opsummering af weighted least squares

Egenskaber ved WLS/FGLS:

- WLS svarer til OLS på den transformerede model.
- WLS kaldes "vægtet", fordi de enkelte observationer vægtes. WLS minimerer den vægtede sum af kvadrerede residualer.
- WLS er BLUE.
- WLS er konsistent og asymptotisk mere efficient end OLS (hvis der er heteroskedasticitet).
- t- og F-test er asymptotisk t- og F-fordelte.
- R^2 i den transformerede model giver sjældent mening.
- WLS er et specialtilfælde af GLS (Generalized Least Squares).
- Hvis h er ukendt, kan vi estimerer den vha. FGLS.

Heterogene effekter

Heterogene effekter

Indtil nu har vi undersøgt betydningen af heterogenitet i $var(u|x)$.

Hvad hvis der er heterogenitet i β_j (random slopes)?

Simpel model som opfylder MLR.1-MLR.4:

$$y = a_i + b_i x_i + u_i$$

hvor

$$E(a_i) = \alpha \Rightarrow a_i = \alpha + c_i$$
$$E(b_i) = \beta \Rightarrow b_i = \beta + d_i$$

Indsæt i den oprindelige model:

Heterogene effekter

Simpel model skrevet ud:

$$y = \alpha + c_i + (\beta + d_i)x_i + u_i \Leftrightarrow y = \alpha + \beta x_i + \underbrace{c_i + d_i x_i + u_i}_{v_i}$$

Dvs. modellen opfylder MLR.4, hvis

$$E(v_i|x) = E(c_i + d_i x_i + u_i|x) = E(c_i|x) + E(d_i|x)x_i = 0$$

To antagelser:

- $E(c_i|x) = 0$: Uobserveret heterogenitet i a_i skal være ukorreleret med x_i (svarer til standard MLR.4).
- $E(d_i|x) = 0$: Uobserveret heterogenitet i effekten af x skal være ukorreleret med x (no selection on gains).

Heterogene effekter

Givet de to antagelser estimerer OLS $\hat{\beta}_1 = \hat{E}(b_i) = \beta$ også kaldet:

- **APE**: Average Partial Effect
- **AME**: Average Marginal Effect
- **ATE**: Average Treatment Effect

"Kært barn har mange navne"

Hvis $E(d_i|x) \neq 0$ er OLS inkonsistent.

Dog med nogle interessante undtagelser.

Heterogene effekter

Undtagelse:

Hvis x er en dummy variable, gælder at

- $\hat{\beta}_0 = E(y_i|x = 0)$
- $\hat{\beta}_1 = E(y_i|x = 1) - E(y_i|x = 0)$

Indsæt modellen for y_i:

$$\hat{\beta}_1 = E(\alpha + c_i + (\beta + d_i)x_i + u_i|x = 1) - E(\alpha + c_i + u_i|x = 0)$$

$$= E(\beta + d_i|x = 1) + E(c_i + u_i|x = 1) - E(c_i + u_i|x = 0)$$

Dvs. OLS estimater den gns. effekt af x for dem som har valgt $x = 1$ kaldet Average Treatment Effect on the Treated (**ATT**).

Opsummering

Opsummering

Antagelsen om variansen af fejlleddet $var(u|x)$ er vigtig for variansen af estimatoren: $var(\hat{\beta}_j|x)$.

Variansen er vigtige for t-, F-teststørrelser og konfidensintervaler.

Man kan bruge Breusch-Pagen eller White's test til at teste for heteroskedasticitet.

Hvis der er heteroskedasticitet kan man bruge en af følgende:

- Anvend OLS med robuste standardfejl (det gør vi typisk)
- Antag en specifik form for heteroskedasticitet og transformer modellen, så der ikke længere er heteroskedasticitet (WLS/FGLS)

WLS/FGLS gør estimatoren mere efficient. Med store dataasæt er efficiens sjældent en problem.

Opsummering

Classical regression assumptions are helpful for the derivation of regression standard errors. They simplify the math and the resulting formula reveals the features of the data that determine statistical precision. However, we don't dwell on statistical tests for the validity of classical assumptions or on generalized least squares fix-ups for their failures. It seems to us that most of what is usually taught on inference in an introductory undergraduate class can be replaced with the phrase "use robust standard errors."

From *Undergraduate Econometrics Instruction: Through Our Classes, Darkly* by Angrist & Pischke

`https://www.aeaweb.org/articles?id=10.1257/jep.31.2.125`