Modul 6: Heteroskedastizität

PI 6250 – Ökonometrie I

Max Heinze (mheinze@wu.ac.at)
Department für Volkswirtschaftslehre, WU Wien

Basierend auf einem Foliensatz von Simon Heß

5. Juni 2025

Was ist Heteroskedastizität?

Robuste Standardfehler

Tests für Heteroskedastizität

Weighted Least Squares

Homoskedastizität und Heteroskedastizität

Wir erinnern uns an Annahme **MLR.5** zum Thema **Homoskedastizität**:

$$\mathrm{Var}(u_i \mid x_{i1}, \ldots, x_{iK}) = \mathrm{Var}(u_i) = \sigma^2,$$

- Wir haben bereits darüber gesprochen, dass diese Annahme **häufig verletzt** ist.
    - Personen mit mehr Bildung haben vermutlich eine höhere Varianz im Einkommen.
    - Personen mit mehr Einkommen haben vermutlich eine höhere Varianz darin, wie viel CO_2-Emissionen sie verursachen.
- Wir nennen den Fall, in dem MLR.5 verletzt ist, **Heteroskedastizität** (engl. **heteroskedasticity**).
    - Heteroskedastizität tritt dann auf, wenn **bestimmte Individuen** oder **Gruppen von Individuen** **mehr oder weniger unerklärte Variation** haben als der Rest.

Heteroskedastizität

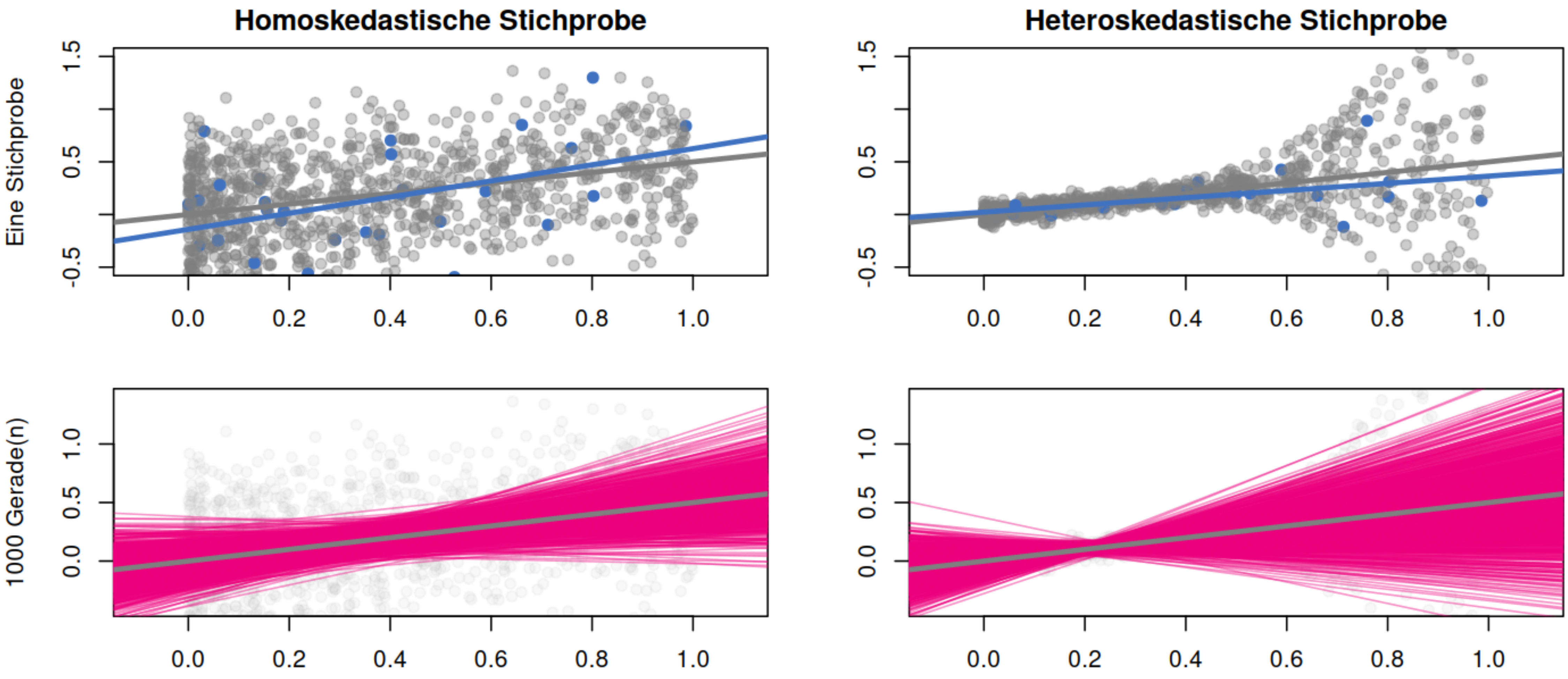
Wenn unser Fehlerterm **heteroskedastisch** ist, dann ist die Varianz abhängig von i:

$$\mathrm{Var}(u_i \mid x_{i1}, \ldots, x_{iK}) = \mathrm{E}(u_i^2 \mid x_{i1}, \ldots, x_{iK}) = \sigma_i^2 \qquad \neq \sigma^2,$$

$$\mathrm{Var}(\boldsymbol{u} \mid \boldsymbol{X}) = \mathrm{E}(\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}) = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2) \qquad \neq \sigma^2 \boldsymbol{I}.$$

- Der **OLS-Schätzer** ist in so einem Fall immer noch **unverzerrt** und **konsistent**, da diese beiden Eigenschaften nur MLR.1 bis MLR.4 voraussetzen.
- Allerdings stimmt die **Formel, mit der wir** $\mathrm{Var}(\hat{\beta})$ **und** $\mathrm{s.e.}(\hat{\beta})$ **berechnet haben**, nicht mehr, und OLS ist auch nicht mehr **effizient**.

Illustration



Was tun?

Wir fassen zusammen: Oft haben wir mit **heteroskedastischen Fehlern** zu tun. Das führt dazu, dass der OLS-Schätzer **nicht mehr effizient** ist und wir die Varianz von $\hat{\beta}$ nicht mehr wie zuvor berechnen können.

- Das verursacht eine **Reihe von Problemen**:
    - Unsere **Standardfehler** sind nicht mehr korrekt.
    - Also sind auch unsere **t-Statistiken**, **F-Statistiken**, etc. irreführend.
    - Ineffizienz bedeutet, dass es jetzt einen ***besseren*** **Schätzer** geben muss.
- Was können wir **dagegen tun**? **Nichts**. Aber wir können lernen, **damit umzugehen**.
    - In ausreichend großen Stichproben wird das Effizienzproblem kleiner.
    - Wir können den **ineffizienten OLS-Schätzer behalten**, aber die **Standardfehler ersetzen**.
    - Wir können **testen, ob Heteroskedastizität vorliegt**.
    - Wir können **einen anderen, effizienten Schätzer benutzen**.

Was ist Heteroskedastizität?

Robuste Standardfehler

Tests für Heteroskedastizität

Weighted Least Squares

Appendix

Varianz des OLS-Schätzers bei Heteroskedastizität

Ursprünglich haben wir angenommen, dass $\operatorname{Var}(\boldsymbol{u} \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}_N$. Unter dieser Annahme war die **Varianz des OLS-Schätzers**

$$\operatorname{Var}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

Wir treffen jetzt eine weniger restriktive Annahme:

$$\operatorname{Var}(\boldsymbol{u} \mid \boldsymbol{X}) = \operatorname{E}(\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}) = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_N^2) =: \boldsymbol{\Omega}$$

Unter **dieser Annahme** ist die **Varianz des OLS-Schätzers**

$$\operatorname{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{\Omega} \boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

Übungsaufgabe

Was passiert mit dieser Formel, wenn $\boldsymbol{\Omega} = \sigma^2 \boldsymbol{I}$?

• **Beweis**

Wir brauchen wieder einen Schätzer

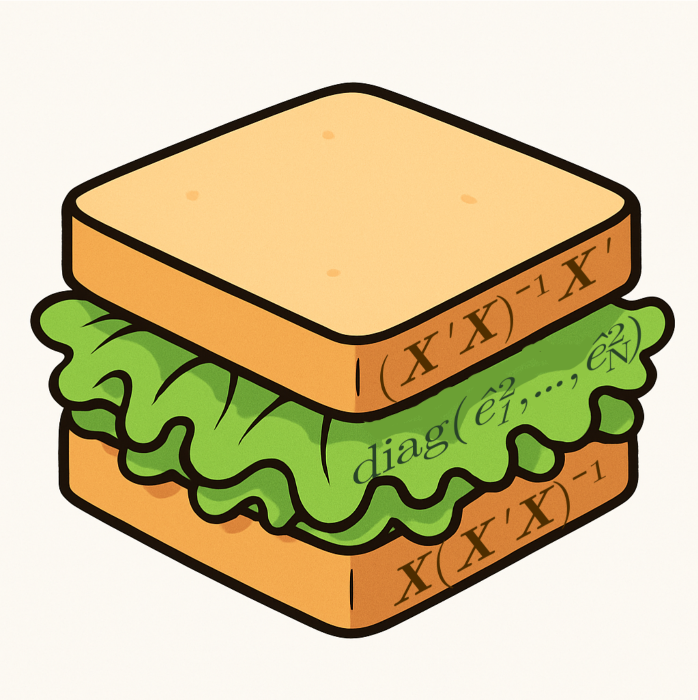
Wir haben ein Problem mit dieser Gleichung:

$$\operatorname{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

Wie **kennen $\boldsymbol{\Omega}$ nicht**. Allerdings ist $\operatorname{diag}(\hat{u}_1^2, \ldots, \hat{u}_N^2)$ ein **konsistenter Schätzer für $\boldsymbol{\Omega}$**.

Mit diesem Schätzer können wir folgenden **Schätzer für die Varianz von $\hat{\beta}$** konstruieren:

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\operatorname{diag}(\hat{u}_1^2, \ldots, \hat{u}_N^2)\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$



Dieser Schätzer wird manchmal **sandwich estimator** genannt (siehe rechts).

Robuste Standardfehler

- Mit diesem Schätzer berechnete Standardfehler nennen wir **gegen Heteroskedastizität robuste Standardfehler** (engl. **heteroskedasticity-robust standard errors**).
- Die daraus berechneten **t-Statistiken** und **F-Statistiken** nennen wir auch **robust**.
- **Robuste Standardfehler** sind sowohl unter Heteroskedastizität als auch unter Homoskedastizität valide.
- **Nicht-robuste Standardfehler** sind nur unter Homoskedastizität valide.
- **t-Statistiken**, die mit **robusten** Standardfehlern berechnet werden, sind nur in **großen Stichprioben** annähernd t-verteilt. In kleinen Stichproben kann die Verteilung sich stark unterscheiden.
- **t-Statistiken** mit **nicht-robusten** Standardfehlern berechnet werden, sind auch in **kleinen Stichproben** exakt t-verteilt, **wenn die Fehler homoskedastisch sind**.

Wir schätzen wieder Modelle mit Schulen 🧑‍🏫

R Code | Start Over | Run Code

```r
# Pakete laden
library(AER) # Enthält unseren Datensatz
library(dplyr)
library(sandwich) # Robuste Standardfehler
libtaty(lmtest) # Robuste Tests

# Daten laden
data("CASchools")

# Variablen berechnen mit mutate()
CASchools <- CASchools |>
  mutate(student_teacher_ratio = students / teachers,
         test_score = (read + math)/2)
```

Gewöhnlicher Output (und gewöhnliche Standardfehler 🚨)

R Code Start Over Run Code

```r
model_1 <- lm(test_score ~ student_teacher_ratio + income + I(income^2) + lunch + english, data = CAS
summary(model_1)
```

Robuste Standardfehler 🥳

R Code

```r
robust_se <- vcovHC(model_1, type = "HC1")
coeftest(model_1, vcov. = robust_se)
```

Was ist Heteroskedastizität?

Robuste Standardfehler

Tests für Heteroskedastizität

Weighted Least Squares

Appendix

Breusch-Pagan-Test

Wir können **testen**, ob bestimmte Formen von **Heteroskedastizität** vorliegen (wir kennen die genaue Form der vorliegenden Heteroskedastizität allerdings für gewöhnlich nicht).

Die erste Herangehensweise, die wir besprechen, ist der **Breusch-Pagan-Test** von Breusch & Pagan (1979). Mit diesem **LM-Test** überprüfen wir, ob σ_i^2 linear von den Regressoren abhängt:

$$\sigma_i^2 = \delta_0 + \delta_1 x_{i1} + \cdots + \delta_K x_{iK} + \text{Fehler}.$$

Die **Nullhypothese** des Tests ist:

$$H_0 : \delta_1 = \cdots = \delta_K = 0$$

In großen Stichproben ist die **LM-Statistik** dieses Tests unter der Nullhypothese χ^2-verteilt mit K Freiheitsgraden.

Breusch-Pagan-Test

Wir führen den **Breusch-Pagan-Test** wie folgt durch:

(1) Wir schätzen die Hauptregression $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$ mit OLS und behalten die Residuen $\hat{u}_i$.

(2) Als nächstes rechnen wir folgende Hilfsregression:

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \cdots + x_{iK} + \text{Fehler},$$

und behalten das R^2 dieser Regression.

(3) Die Statistik NR^2 ist die ungefähre LM-Statistik und ist in großen Stichproben χ^2_K-verteilt.

White-Test

Der **White-Test** von White (1980) ist eine Variante des Breusch-Pagan-Tests mit einer flexibleren Spezifikation: er berücksichtigt auch **alle möglichen quadrierten Terme und Interaktionen** der Regressoren. Wir führen ihn wie folgt durch:

(1) Wir schätzen die Hauptregression $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$ mit OLS und behalten die Residuen $\hat{u}_i$.

(2) Als nächstes rechnen wir folgende Hilfsregression:

$$\begin{aligned}\hat{u}_i^2 = \delta_0 + &\delta_1 x_{i1} + \cdots + x_{iK} + \\ &\delta_{K+1} x_{i1}^2 + \cdots + \delta_{2K} x_{iK}^2 + \\ &\delta_{2K+1} x_{i1} x_{i2} + \cdots + \delta_{(K(K+3)2)} x_{i,K-1} x_{iK} + \text{Fehler},\end{aligned}$$

und behalten das R^2 dieser Regression.

(3) Die Statistik NR^2 ist die ungefähre LM-Statistik und ist in großen Stichproben χ^2_K-verteilt.

White-Test

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \cdots + x_{iK} + \\ \delta_{K+1} x_{i1}^2 + \cdots + \delta_{2K} x_{iK}^2 + \\ \delta_{2K+1} x_{i1} x_{i2} + \cdots + \delta_{(K(K+3)2)} x_{i,K-1} x_{iK} + \text{Fehler},$$

Diese Regression hat $(K(K+3)2)$ Regressoren. Das sind sehr viele Regressoren. Wenn K groß und N klein ist, sind es vielleicht sogar zu viele.

Eine **alternative Version** des White-Tests ist

$$\hat{u}_i^2 = \delta_0 + \delta_1 \hat{y}_i + \delta_2 \hat{y}_i^2 + \text{Fehler}.$$

- Wir **regressieren** also $\hat{u}_i^2$ auf die **angepassten Werte** aus der Regression von Schritt (1).
- Da die $\hat{y}_i$ eine lineare Funktion der erklärenden Variablen sind, ist $\hat{y}_i^2$ eine bestimmte **Funktion der Quadrate und Kreuzprodukte** der erklärenden Variablen.

Tests für Heteroskedastizität in R

R Code Start Over Run Code

```r
# Breusch-Pagan-Test
cat("Breusch-Pagan-Test:")
bptest(model_1)

# White-Test (alternative Form)
cat("White-Test:")
bptest(model_1, ~ fitted(model_1) + I(fitted(model_1)^2), data = CASchools)
```

Was ist Heteroskedastizität?

Robuste Standardfehler

Tests für Heteroskedastizität

Weighted Least Squares

Appendix

Wie finden wir einen effizienten Schätzer?

Angenommen, wir haben **Heteroskedastizität**, aber wir **kennen die** σ_i^2. Wir wollen folgende Regression schätzen:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + u_i,$$

wissen aber, dass **OLS ineffizient** ist.

Mit den **Fehlervarianzen** σ_i^2 können wir aber einen **effizienten Schätzer** konstruieren.

Dazu **dividieren wir die Regression durch** $\sigma_i = \sqrt{\sigma_i^2}$:

$$\frac{y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{x_{i1}}{\sigma_i} + \cdots + \beta_K \frac{x_{iK}}{\sigma_i} + \frac{u_i}{\sigma_i}$$

Warum tun wir das? **Weil wir so die Varianz so skalieren können, dass sie für alle i gleich ist.**

- Wenn $\mathrm{Var}(u_i) = \sigma_i^2$, dann ist $\mathrm{Var}(u_i/\sigma_i) = 1$. MLR.5 ist also erfüllt.

WLS-Schätzer

Wir **gewichten** (engl. **weight**) also Beobachtungen mit größerer Varianz weniger stark als solche mit geringerer Varianz – daher der Name **weighted least squares**. In Matrixschreibweise:

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta}_{\text{WLS}} + \tilde{\boldsymbol{u}},$$

wobei $\tilde{\boldsymbol{y}} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{y}$, $\tilde{\boldsymbol{X}} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{X}$ und $\tilde{\boldsymbol{u}} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{u}$; $\boldsymbol{\Omega} = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$.

Der **WLS-Schätzer** ist in diesem Fall:

$$\hat{\boldsymbol{\beta}}_{\text{WLS}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{y}} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}.$$

Dieser **WLS-Schätzer** ist ein Spezialfall des **Generalized-Least-Squares-Schätzers** (GLS). GLS funktioniert mit jeder Varianz-Kovarianz-Matrix $\boldsymbol{\Omega}$, nicht nur mit der oben genannten diagonalen.

Varianz des WLS-Schätzers

Die Varianz des **WLS-Schätzers** ist:

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}_{\text{WLS}} \mid \boldsymbol{X}) = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}$$

Wir können diese Varianz **mithilfe von $\hat{\boldsymbol{\Omega}}$ schätzen.** Somit können wir auch **Standardfehler** für Tests erhalten. Die Varianz des **WLS-Schätzers** ist **geringer als die des OLS-Schätzers** (was wir nicht beweisen):

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}_{\text{OLS}} \mid \boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

Feasible Generalized Least Squares

Das **Problem** ist: Wir können das **nicht schätzen**. **GLS (WLS) setzt voraus, dass wir die σ_i^2 kennen**, das tun wir aber nicht.

- Wir können die σ_i^2 aber **schätzen**.
- Wir können zum Beispiel **annehmen**, dass

$$\sigma_i^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_i + \cdots + \delta_K x_K),$$

  wobei wir die Exponentialfunktion verwenden, um negative Werte zu vermeiden.
- Wir logarithmieren und setzen $\hat{u}_i^2$ für σ_i^2 ein:

$$\log(\hat{u}_i^2) = \alpha_0 + \delta_1 x_i + \cdots + \delta_K x_K + \text{Fehler}, \qquad \alpha_0 = \log(\sigma^2) + \delta_0$$

- Wir nennen die **angepassten Werte** aus dieser Regression $\hat{g}_i$ und benutzen $\hat{\sigma}_i = \sqrt{\exp(\hat{g}_i)}$ als **Gewichte**. Den Schätzer, den wir so erhalten, nennen wir **Feasible Generalized Least Squares** (**fGLS**).

Wie implementieren wir fGLS?

Wenn wir **Feasible Generalized Least Squares** anwenden wollen, können wir so vorgehen:

(1) Wir regressieren y mit OLS auf $x_1, \ldots, x_K$ und behalten die Residuen $\hat{u}$.

(2) Wir berechnen $\log(\hat{u}^2)$ mit diesen Residuen.

(3) Wir regressieren $\log(\hat{u}^2)$ mit OLS auf $x_1, \ldots, x_K$ und behalten die angepassten Werte $\hat{g}$.

(4) Um Schätzungen für die Varianz zu bekommen, berechnen wir $\hat{\sigma}_i^2 = \exp(\hat{g}_i)$.

(5) Als Letztes regressieren wir y mit WLS auf $x_1, \ldots, x_K$, wobei wir $1/\sqrt{\hat{\sigma}_i{}^2}$ als Gewichte verwenden.

Ein **Problem** bleibt: Wir kennen die „wahre" **funktionale Form der Heteroskedastizität nicht**, wir haben nur eine mögliche Form angewendet.

- WLS ist nur dann garantiert **effizient**, wenn diese Form **korrekt spezifiziert** ist.
- Ist das **nicht der Fall**, ist **fGLS** in **großen Stichproben** aber trotzdem **effizienter als OLS**.
- Außerdem ist **fGLS** **konsistent**, wenn auch **nicht unverzerrt**.

Literatur

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 1287. https://doi.org/10.2307/1911963

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817. https://doi.org/10.2307/1912934

Wooldridge, J. M. (2020). *Introductory econometrics : a modern approach* (Seventh edition, S. xxii, 826 Seiten). Cengage. https://permalink.obvsg.at/wuw/AC15200792

Robuste Standardfehler

Tests für Heteroskedastizität

Weighted Least Squares

Appendix

Varianz des OLS-Schätzers im Allgemeinen Fall

$$\begin{aligned}
\operatorname{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) &= \operatorname{Var}\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \mid \boldsymbol{X}\right) \\
&= \operatorname{Var}\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) \mid \boldsymbol{X}\right) \\
&= \operatorname{Var}\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \mid \boldsymbol{X}\right) \\
&= \operatorname{Var}\left(\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \mid \boldsymbol{X}\right) \\
&= \operatorname{Var}\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \mid \boldsymbol{X}\right) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\operatorname{Var}(\boldsymbol{u} \mid \boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}$$

• Zurück