

Kapitel GLS:

Das verallgemeinerte Regressionsmodell



Moodle



Lehrbuch

Das klären wir in diesem Kapitel:

Verletzung von MLR 5: Konsequenzen für OLS-Schätzer

GLS: Generalized Least Squares

FGLS: Feasible Generalized Least Squares

Klassisches Regressionsmodell (OLS)

$$MLR\ 1: \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

$$MLR\ 3: \quad rk(\mathbf{X}) = K + 1$$

$$MLR\ 4: \quad E[\mathbf{u}|\mathbf{X}] = \mathbf{0}$$

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

MLR 5 definiert klassische (homoskedastische und seriell unkorrelierte) Fehlerterme.

Verallgemeinertes Regressionsmodell (GLS)

MLR 1, 3 & 4 wie OLS und

$$MLR\ 5': \quad \Sigma(\mathbf{u}|\mathbf{X}) = \sigma^2 \boldsymbol{\Omega}$$

wobei $\boldsymbol{\Omega}$ eine positiv definit symmetrische (pds) $n \times n$ – Matrix ist.

Ω
 $n \times n$ positiv definit, falls

$$x' \Omega x > 0 \quad \text{für alle Vektoren} \\ x \in \mathbb{R}^n, x \neq 0$$

Eigenschaften positiv definiter Matrizen

Sei die $n \times n$ -Matrix **A** positiv definit symmetrisch (pds), d.h. es gelte $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ für alle $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{0}$ und $\mathbf{A}' = \mathbf{A}$. 

Dann hat **A** vollen Rang und es existiert die Inverse $\mathbf{A}^{-1}$, welche ebenfalls pds ist.

Für eine pds Matrix **A** existiert eine pds Matrix **Z**, sodass $\mathbf{A} = \mathbf{Z}\mathbf{Z}$.

Da $\mathbf{A}^{-1}$ ebenfalls pds ist, existiert auch eine pds Matrix **P**, sodass:

$$\mathbf{A}^{-1} = \mathbf{P}\mathbf{P}$$

Es gilt dann $\mathbf{P} = \mathbf{Z}^{-1}$.

Spezialfälle

Bevor wir das zentrale Resultat, das

Gauß-Markov-Aitken-Theorem

für allgemeine pds-Matrizen Ω besprechen, widmen wir uns zwei *Spezialfällen*:

- ▶ heteroskedastische und seriell unkorrelierte Störterme *Kapitel 8*
- ▶ homoskedastische und seriell korrelierte Störterme *Kapitel 12*

Spezialfall: $\mathbf{u}$ heteroskedastisch, seriell unkorreliert

Die Störterme sind seriell unkorreliert ($\text{Cov}(u_i, u_j) = 0 \ \forall \ i \neq j$), die Varianzen variieren aber über die Beobachtungen ($\text{Var}(u_i) = \sigma_i^2$).

Wir schreiben dann

$$\mathbb{E}(\mathbf{u}|\mathbf{X}) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} = \sigma^2 \underbrace{\begin{pmatrix} \frac{\sigma_1^2}{\sigma^2} & 0 & \dots & 0 \\ 0 & \frac{\sigma_2^2}{\sigma^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma_n^2}{\sigma^2} \end{pmatrix}}_{\mathbf{\Omega}} = \sigma^2 \mathbf{\Omega}$$

mit $\sigma > 0$ und $\sigma_i > 0 \ \forall \ i = 1, \dots, n$

Dieser Fall ist häufig in Querschnittsmodellen.

$$\underbrace{\begin{pmatrix} \frac{\sigma_1^2}{\sigma^2} & 0 \\ 0 & \frac{\sigma_2^2}{\sigma^2} \end{pmatrix}}_{\Omega}^{-1} = \frac{1}{\frac{\sigma_1^2}{\sigma^2} \cdot \frac{\sigma_2^2}{\sigma^2}} \begin{pmatrix} \frac{\sigma_2^2}{\sigma^2} & 0 \\ 0 & \frac{\sigma_1^2}{\sigma^2} \end{pmatrix}$$

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

$$P = \frac{1}{\sigma_1} \cdot \frac{1}{\sigma_2} \begin{pmatrix} \frac{\sigma_2}{\sigma} & 0 \\ 0 & \frac{\sigma_1}{\sigma} \end{pmatrix}$$

Spezialfall heteroskedastisch, seriell unkorreliert

Inverse von Ω :

$$\Omega^{-1} = \begin{pmatrix} \frac{\sigma^2}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{\sigma^2}{\sigma_2^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma^2}{\sigma_n^2} \end{pmatrix}$$

Zerlegung von Ω^{-1} :

$$\begin{pmatrix} \frac{\sigma^2}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{\sigma^2}{\sigma_2^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma^2}{\sigma_n^2} \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{\sigma}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{\sigma}{\sigma_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma}{\sigma_n} \end{pmatrix}}_{P = \sqrt{\Omega^{-1}}} \underbrace{\begin{pmatrix} \frac{\sigma}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{\sigma}{\sigma_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma}{\sigma_n} \end{pmatrix}}_{P = \sqrt{\Omega^{-1}}}$$

Spezialfall homoskedastisch, seriell korreliert

Die Störterme haben identische Varianzen, sind aber untereinander korreliert:

$$\text{Var}(u_i|\mathbf{X}) = \sigma^2 > 0 \quad \forall i = 1, \dots, n \quad \& \quad \text{Cov}(u_s, u_t|\mathbf{X}) \neq 0 \quad \text{für } s \neq t$$

$$\begin{aligned} \Sigma(\mathbf{u}|\mathbf{X}) &= \begin{pmatrix} \sigma^2 & \text{Cov}(u_1, u_2|\mathbf{X}) & \dots & \text{Cov}(u_1, u_n|\mathbf{X}) \\ \text{Cov}(u_2, u_1|\mathbf{X}) & \sigma^2 & \dots & \text{Cov}(u_2, u_n|\mathbf{X}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(u_n, u_1|\mathbf{X}) & \text{Cov}(u_n, u_2|\mathbf{X}) & \dots & \sigma^2 \end{pmatrix} \\ &= \sigma^2 \underbrace{\begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{pmatrix}}_{\Omega} \end{aligned} \quad \rho_{ij} = \frac{\text{Cov}(u_i, u_j|\mathbf{x})}{\underbrace{\sqrt{\text{Var}(u_i|\mathbf{X})}}_{\sigma} \underbrace{\sqrt{\text{Var}(u_j|\mathbf{X})}}_{\sigma}}$$

Dieser Fall ist häufig in Zeitreihendaten.

Verletzung von MLR 5: Konsequenzen für OLS-Schätzer

Verletzung von MLR 5: Konsequenzen für OLS-Schätzer

Welche Konsequenzen hat die Verallgemeinerung von MLR 5:

$$\mathbb{E}(\mathbf{u}|\mathbf{X}) = \sigma^2 \mathbf{I} \text{ durch}$$

$$\text{MLR 5'} : \mathbb{E}(\mathbf{u}|\mathbf{X}) = \sigma^2 \mathbf{\Omega}$$

auf die OLS-Schätzer?

Wir schätzen den Parametervektor β zunächst mit OLS:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (\text{mit MLR 3})$$

$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) \quad (\text{mit MLR 1})$$

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

Mit MLR 4 $E[\mathbf{u}|\mathbf{X}] = \mathbf{0}$ gilt:

$$E[\hat{\beta}|\mathbf{X}] = E[\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} \checkmark$$

Der OLS-Schätzer ist also nach wie vor erwartungstreu!

$$\begin{aligned} E[\hat{\beta}] &= \beta, f.1 // 5 \\ \beta &= A \cdot y \\ A \cdot X &= \underline{I} \end{aligned}$$

$$A = (X'X)^{-1}X'$$

$$AX = (X'X)^{-1}X'X = I \checkmark$$

Varianz-Kovarianz-Matrix des OLS-Schätzers $\hat{\beta}$
bei Verletzung von MLRS:

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

Zum Vergleich:
MLRS

$$\begin{aligned} \varphi(\hat{\beta} | X) &= \varphi(\underbrace{\beta}_{\text{Konstante}} + (X'X)^{-1}X'u | X) \\ &= \varphi((X'X)^{-1}X'u | X) \end{aligned}$$

$$\varphi(\hat{\beta} | X) = \sigma^2 (X'X)^{-1}$$

$$= (X'X)^{-1}X' \underbrace{\varphi(u | X)}_{\sigma^2 \Omega} (X'X)^{-1}$$

$$= (X'X)^{-1}X' \sigma^2 \Omega (X')' ((X'X)^{-1})'$$

$$= \sigma^2 \underbrace{(X'X)^{-1}}_{\uparrow} X' \Omega X \underbrace{(X'X)^{-1}}_{\uparrow}$$

Konsequenzen für OLS

Für die Varianz-Kovarianz-Matrix von $\hat{\beta}$ gilt aber:

$$\begin{aligned}\mathbb{V}(\hat{\beta}|\mathbf{X}) &= \mathbb{V}(\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \mathbb{V}(\mathbf{u}|\mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Die übliche OLS-Varianz-Kovarianz-Matrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ ist also für $\Omega \neq I$ nicht mehr die korrekte Varianz-Kovarianz-Matrix der geschätzten Koeffizienten.

Konsequenzen für OLS

Es gelte *MLR 5'* : $\mathbb{E}(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{\Omega}$ mit $\mathbf{\Omega}$ pds.

- ▶ man kann mit OLS die Parameter unverzerrt schätzen,
- ▶ aber OLS ist im Allgemeinen nicht mehr der BLUE (es gibt Schätzer mit geringerer Varianz),
- ▶ und die OLS - Kovarianzmatrix ist keine gute Schätzung der tatsächlichen Kovarianzmatrix, daher sind darauf basierende Standardfehler, *t*-Tests und *F*-Tests falsch.

Fazit: OLS liefert unverzerrte Parameter und falsche Standardfehler und Tests. Ob man also OLS anwenden sollte, hängt von der Anwendung ab (ist man an Tests interessiert oder nur an Parametern?)

GLS: Generalized Least Squares

GLS: Generalized Least Squares

Angenommen, die Matrix Ω in $\Sigma(\mathbf{u}|\mathbf{X}) = \sigma^2\Omega$ wäre bekannt und positiv definit symmetrisch.

Dann existiert eine pds $n \times n$ -Matrix $\mathbf{P}$ mit:

$$\mathbf{P}\mathbf{P} = \Omega^{-1} \text{ bzw. } \mathbf{P}^{-1}\mathbf{P}^{-1} = \Omega \text{ bzw. } \mathbf{P}\Omega\mathbf{P} = \mathbf{I}$$

Wir benutzen nun $\mathbf{P}$, um die Daten $\mathbf{X}$ und $\mathbf{y}$ so zu transformieren, dass wir die bisherigen Techniken valide nutzen können!

Transformation:

$$\tilde{\mathbf{X}} = \mathbf{P}\mathbf{X}$$

$$\tilde{\mathbf{y}} = \mathbf{P}\mathbf{y}$$

$$\tilde{\mathbf{u}} = \mathbf{P}\mathbf{u}$$

$$\mathbf{X} = \mathbf{P}^{-1}\tilde{\mathbf{X}}$$

GLS: Generalized Least Squares

$$\begin{aligned} \tilde{y} &= P(y) = P(x\beta + u) \\ \tilde{y} &= \underbrace{Px}_{\tilde{x}} + \underbrace{Pu}_{\tilde{u}} \end{aligned}$$

Lineare Transformation der Daten:

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

Schätze das Modell

$$\tilde{y} = \tilde{X}\beta + \tilde{u} \quad (\text{mit } \tilde{u} = Pu)$$

per OLS.

Beachte: β wird nicht transformiert!

GLS: Generalized Least Squares

Prüfe die klassischen Annahmen für das transformierte Modell:

MLR 1:

Prämultiplikation von $\mathbf{P}$ ist eine lineare Transformation ✓

MLR 3:

$$rk(\tilde{\mathbf{X}}) = rk(\mathbf{P}\mathbf{X}) = \min\{\underbrace{rk(\mathbf{P})}_n, \underbrace{rk(\mathbf{X})}_{K+1}\} = rk(\mathbf{X}) = \underline{K+1} \checkmark$$

MLR 4:

$$E[\tilde{\mathbf{u}}|\tilde{\mathbf{X}}] = \underbrace{E[\mathbf{P}\mathbf{u}|\mathbf{X}]}_{\substack{\text{MR4} \\ = 0}} = \mathbf{P}E[\mathbf{u}|\mathbf{X}] = \mathbf{P}\mathbf{0} \checkmark = \mathbf{0} \checkmark$$

$\Omega = \mathbf{P}^{-1} \cdot \mathbf{P}^{-1}$

MLR 5:

$$\Sigma(\tilde{\mathbf{u}}|\tilde{\mathbf{X}}) = \Sigma(\mathbf{P}\mathbf{u}|\mathbf{X}) = \mathbf{P} \underbrace{\Sigma(\mathbf{u}|\mathbf{X})}_{\sigma^2 \Omega} \mathbf{P}' = \sigma^2 \mathbf{P}\Omega\mathbf{P}' = \sigma^2 \mathbf{I} \checkmark$$

Demnach ist der OLS-Schätzer für das transformierte Modell BLUE!

GLS: Generalized Least Squares

Der OLS-Schätzer der Transformierten Daten $\tilde{\beta}$ lautet demnach

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

Wir nennen den **GLS-Schätzer** $\tilde{\beta}$ auch **Aitken-Schätzer**.

$$\begin{aligned}
 \hat{\sigma}^2(\tilde{\beta} | x) &= \sigma^2 (\tilde{x}' \tilde{x})^{-1} \\
 &= \sigma^2 ((Px)' Px)^{-1} \\
 &= \sigma^2 (x' P Px)^{-1} \\
 &= \sigma^2 (x' \Omega^{-1} x)^{-1}
 \end{aligned}$$

$$\hat{\sigma}^2(\hat{\beta} | x) = \sigma^2 (x' x)^{-1}$$

Die Varianz des Aitken-Schätzers $\tilde{\beta}$

Mit $\tilde{\beta} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}$ gilt:

Die Varianz des Aitken-Schätzers $\tilde{\beta}$

Mit $\tilde{\beta} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1}\mathbf{y}$ gilt:

$$\begin{aligned}\mathbb{V}(\tilde{\beta}|\mathbf{X}) &= \mathbb{V}((\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1}\mathbf{y}|\mathbf{X}) \\&= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1} \mathbb{V}(\mathbf{u}|\mathbf{X}) \left((\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1} \right)' \\&= \sigma^2 (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1}\Omega\Omega^{-1}\mathbf{X} (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \\&= \sigma^2 (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \\&= \sigma^2 (\mathbf{X}'\mathbf{P}\mathbf{P}\mathbf{X})^{-1} \\&= \sigma^2 ((\mathbf{P}\mathbf{X})' \mathbf{P}\mathbf{X})^{-1} = \sigma^2 (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1}\end{aligned}$$

Das Gauss-Markov-Aitken Theorem

Unter den Annahmen *MLR* 1, 3, 4, 5' ist ein Schätzer innerhalb der Klasse der linearen und unverzerrten Schätzer varianzminimal (effizient), falls es sich um $\tilde{\beta}$ handelt.

Die Struktur des Beweises ist die folgende:

- ▶ Definiere die Klasse der linearen Schätzer.
- ▶ Charakterisiere darin die Unterklasse der unverzerrten Schätzer.
- ▶ Zeige Effizienz von $\tilde{\beta}$.

t - und F -Tests bei bekannten Parametern σ^2 und Ω

Die Varianz des GLS-Schätzers beträgt

$$\text{Var}(\tilde{\beta}|\mathbf{X}) = \sigma^2 (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}$$

Alle darauf basierenden Statistiken (Standardfehler, t -Test, F -Test) behalten ihre Gültigkeit.

GLS: Schätzer für σ^2

Falls die Matrix $\mathbf{\Omega}$ bekannt ist, kann der Parameter σ^2 analog zu OLS geschätzt werden.

GLS-Residuen:

$$\tilde{\mathbf{u}} = \mathbf{Py} - \mathbf{PX}\tilde{\boldsymbol{\beta}} = \mathbf{P}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}]\mathbf{u}$$

Schätzer $\tilde{\sigma}^2$ für σ^2 :

$$\tilde{\sigma}^2 = \frac{1}{n - K - 1} \tilde{\mathbf{u}}' \tilde{\mathbf{u}}$$

FGLS: Feasible Generalized Least Squares

FGLS: Feasible Generalized Least Squares

- ▶ GLS ist nicht durchführbar (infeasible), weil Ω bzw. P , mit der die Daten transformiert werden müssen, nicht bekannt ist.
- ▶ Um eine durchführbare (feasible) Variante zu erhalten, müssen zuerst die Elemente von Ω geschätzt werden. Dies wird mit **feasible generalized least squares** (FGLS) bezeichnet.
- ▶ Nach Ausnutzung der Symmetrie von Ω bleiben aber $\frac{n(n+1)}{2}$ Elemente, die geschätzt werden müssten – das sind zu viele!
- ▶ Nehme an, dass Elemente von Ω von einem Parameter-Vektor α funktional abhängen und schätze α durch $\hat{\alpha}$.
- ▶ Berechne nun $\hat{\Omega} = \Omega(\hat{\alpha})$ und berechne $\tilde{\beta}^{FGLS}$ mit $\hat{\Omega}$.

Beispiel für Funktionale Form

$$\Omega = \begin{pmatrix} 1 & s & s^2 & s^3 & & & s^{n-1} \\ s & 1 & s & s^2 & & & s^{n-2} \\ s^2 & s & 1 & s & & & \vdots \\ s^3 & s^2 & s & 1 & & & \vdots \\ & \vdots & & & \ddots & & s \\ s^{n-1} & s^{n-2} & & & & s & 1 \end{pmatrix}$$

Eigenschaften der OLS-, GLS- und FGLS-Schätzer

- ▶ Der OLS-Schätzer ist auch unter MLR 5' $\mathbb{E}(\mathbf{u}) = \sigma^2 \mathbf{\Omega}$ (pds) erwartungstreu.
- ▶ Der OLS-Schätzer ist aber nicht mehr effizient.
- ▶ Der GLS-Schätzer ist BLUE (und damit effizient).
- ▶ Der GLS-Schätzer ist aber nicht feasible.
- ▶ 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 $\mathbf{\Omega}(\alpha)$ unterstellt wird.