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)

$$
\begin{aligned}
MLR\ 1: \quad & \mathbf{y} &=& \quad X\boldsymbol{\beta} + \mathbf{u} \\
MLR\ 3: \quad & rk(X) &=& \quad k+1 \\
MLR\ 4: \quad & E[\mathbf{u}|X] &=& \quad \mathbf{0} \\
MLR\ 5: \quad & \Sigma(\mathbf{u}|X) &=& \quad \sigma^2 I
\end{aligned}
$$

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

Verallgemeinertes Regressionsmodell (GLS)

MLR 1, 3 & 4 wie OLS und

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

wobei Ω eine positiv definit symmetrische (pds) $n \times n$ – Matrix ist.

Eigenschaften positiv definiter Matrizen

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

Dann hat A vollen Rang und es existiert die Inverse A^{-1}, welche ebenfalls positiv definit ist.

Für eine positiv definite Matrix A existiert eine reguläre Matrix Z, sodass $A = ZZ'$. Deswegen gilt:

$$A' = \left(ZZ'\right)' = \left(Z'\right)' Z' = ZZ' = A,$$

d.h. jede positiv definite Matrix ist symmetrisch.

Da A^{-1} ebenfalls positiv definit ist, existiert auch eine reguläre Matrix P, sodass:

$$A^{-1} = P'P$$

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

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
- homoskedastische und seriell korrelierte Störterme

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

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

Wir schreiben dann

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

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

Dieser Fall ist häufig in Querschnittsmodellen.

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}:

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

Spezialfall homoskedastisch, seriell korreliert

Die Störterme haben identische Varianzen, sind aber untereinander korreliert:

$$Var(u_i|X) = \sigma^2 > 0 \;\forall\; i = 1, \dots, n \;\&\; Cov(u_s, u_t|X) \neq 0 \;\; s \neq t$$

$$\begin{aligned}
\Sigma(\mathbf{u}|X) &= \begin{pmatrix} \sigma^2 & Cov(u_1,u_2|X) & \dots & Cov(u_1,u_n|X) \\ Cov(u_2,u_1|X) & \sigma^2 & \dots & Cov(u_2,u_n|X) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(u_n,u_1|X) & Cov(u_n,u_2) & \dots & \sigma^2 \end{pmatrix} \\
&= \sigma^2 \underbrace{\begin{pmatrix} 1 & corr(u_1,u_2|X) & \dots & corr(u_1,u_n|X) \\ corr(u_2,u_1|X) & 1 & \dots & corr(u_2,u_n|X) \\ \vdots & \vdots & \ddots & \vdots \\ corr(u_n,u_1|X) & corr(u_n,u_2|X) & \dots & 1 \end{pmatrix}}_{\Omega}
\end{aligned}$$

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

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

auf die OLS-Schätzer?

Wir schätzen den Parametervektor $\boldsymbol{\beta}$ zunächst mit OLS:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (X'X)^{-1}X'\mathbf{y} && \text{(mit MLR 3)}\\
&= (X'X)^{-1}X'(X\boldsymbol{\beta}+\mathbf{u}) && \text{(mit MLR 1)}\\
&= \boldsymbol{\beta}+(X'X)^{-1}X'\mathbf{u}
\end{aligned}$$

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

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

Der OLS-Schätzer ist also nach wie vor erwartungstreu!

Konsequenzen für OLS

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

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

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

Konsequenzen für OLS

Es gelte *MLR* $5'$: $\Sigma(\mathbf{u}|X) = \sigma^2\Omega$ mit Ω 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 $\Sigma(\mathbf{u}|X) = \sigma^2\Omega$ wäre bekannt und positiv definit symmetrisch.

Dann existiert eine reguläre Matrix P mit:

$$P'P = \Omega^{-1} \text{ bzw. } P^{-1}\left(P^{-1}\right)' = \Omega \text{ bzw. } P\Omega P' = I$$

Da P regulär ist, hat P vollen Rang. Ferner gilt $n = rk(\Omega^{-1}) = rk(P'P) = rk(P)$. Daher muss P der Ordnung $n \times n$ sein.

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

GLS: Generalized Least Squares

Lineare Transformation der Daten:

$$\tilde{\mathbf{y}} = P\mathbf{y}, \; \tilde{X} = PX$$

Schätze das Modell

$$\tilde{\mathbf{y}} = \tilde{X}\boldsymbol{\beta} + \tilde{\mathbf{u}} \qquad \text{(mit } \tilde{\mathbf{u}} = P\mathbf{u}\text{)}$$

per OLS.

Beachte: $\boldsymbol{\beta}$ wird nicht transformiert!

GLS: Generalized Least Squares

Prüfe die klassischen Annahmen für das transformierte Modell:

MLR 1:

Prämultiplikation von P ist eine lineare Transformation ✓

MLR 3:

$$rk(\tilde{X}) = rk(PX) = \min\{rk(P), rk(X)\} = rk(X) = k + 1 \checkmark$$

MLR 4:

$$E[\tilde{\mathbf{u}}|\tilde{X}] = E[P\mathbf{u}|X] = PE[\mathbf{u}|X] = P\mathbf{0} \checkmark$$

MLR 5:

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

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

GLS: Generalized Least Squares

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

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

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

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

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

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

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

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

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 $\tilde{\boldsymbol{\beta}}$ handelt.

Die Struktur des Beweises ist die folgende:

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

Beweis Gauss-Markov-Aitken („dann“)

Sei $\check{\boldsymbol{\beta}}$ ein Schätzer für $\boldsymbol{\beta}$, welcher linear in $\mathbf{y}$ ist, also

$$\check{\boldsymbol{\beta}} = C\mathbf{y}$$

für eine geeignete Matrix C.

Der Erwartungswert von $\check{\boldsymbol{\beta}}$ beträgt

$$\begin{aligned} E[\check{\boldsymbol{\beta}}|X] &= E[C\mathbf{y}|X] &= CE[\mathbf{y}|X] \\ &= CE[X\boldsymbol{\beta} + \mathbf{u}|X] &= CX\boldsymbol{\beta} + CE[\mathbf{u}|X] \\ &= CX\boldsymbol{\beta} \end{aligned}$$

Erwartungstreue verlangt nun:

$$CX\boldsymbol{\beta} = \boldsymbol{\beta} \; \forall \; \boldsymbol{\beta} \in \mathbb{R}^{k+1}, \text{ also } CX = I$$

Beweis Gauss-Markov-Aitken („dann")

Ein linearer unverzerrter Schätzer $\check{\boldsymbol{\beta}} = C\mathbf{y}$ erfüllt also $CX = I$.

Es gilt also:
$$\begin{aligned} \check{\boldsymbol{\beta}} &= C\mathbf{y} &&= C\,(X\boldsymbol{\beta} + \mathbf{u}) \\ &= CX\boldsymbol{\beta} + C\mathbf{u} &&= \boldsymbol{\beta} + C\mathbf{u} \end{aligned}$$

Die Varianz von $\check{\boldsymbol{\beta}}$ berechnet sich durch:

$$\Sigma(\check{\boldsymbol{\beta}}|X) = \Sigma(\boldsymbol{\beta} + C\mathbf{u}|X) = C\,\Sigma(\mathbf{u}|X)C' = \sigma^2 C\Omega C'$$

Die Varianz von $\tilde{\boldsymbol{\beta}}$ ist gegeben durch:

$$\Sigma(\tilde{\boldsymbol{\beta}}|X) = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1}$$

Beweis Gauss-Markov-Aitken („dann“)

Es bleibt nun zu Zeigen, dass

$$\mathbb{V}(\check{\beta}|X) = \sigma^2 C\Omega C' \geq \sigma^2 \left(X'\Omega^{-1}X\right)^{-1} = \mathbb{V}(\tilde{\beta}|X)$$

für alle C mit $CX = I$ im Sinne der Loewner-Ordnung (s. Kap. 3)

Für die Argumentation benutzen wir folgende Eigenschaften:

- $CX = I$
- $\Omega = P^{-1}\left(P^{-1}\right)'$ bzw. $\Omega^{-1} = P'P$
- Falls für eine Matrix A das Produkt AA' definiert ist, so ist AA' psds. (Es gilt $\mathbf{x}'AA'\mathbf{x} = \left(A'\mathbf{x}\right)' A'\mathbf{x} = ||A'\mathbf{x}|| \geq 0 \; \forall \; \mathbf{x}$.)

Begründung für: $\tilde{\beta}$ effizient

Definiere $D := CP^{-1}$ und $\tilde{X} := PX$.

Es gilt dann:

- $D\tilde{X} = CP^{-1}PX = CX = I$
- $\sigma^2 DD' = \sigma^2 CP^{-1}(P^{-1})'C' = \sigma^2 C\Omega C' = \Sigma(\check{\beta}|X)$
- $\sigma^2(\tilde{X}'\tilde{X})^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1} = \Sigma(\tilde{\beta}|X)$

Begründung für: $\tilde{\beta}$ effizient

Schritt 1: Das Matrixprodukt
$\left(D - \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)\left(D - \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)'$ ist psds.

Schritt 2: Dieses Matrixprodukt lässt sich ausmultiplizieren zu:

$$
\begin{aligned}
& DD' - D\left(\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)' - \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'D' + \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1} \\
= \; & DD' - D\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1} - \left(\tilde{X}'\tilde{X}\right)^{-1}\left(D\tilde{X}\right)' + \left(\tilde{X}'\tilde{X}\right)^{-1} \\
= \; & DD' - \left(\tilde{X}'\tilde{X}\right)^{-1} \\
= \; & C\Omega C' - \left(X'\Omega^{-1}X\right)^{-1} \; \text{psds}
\end{aligned}
$$

Begründung für: $\tilde{\boldsymbol{\beta}}$ effizient

Wir haben gezeigt:

Für eine gegebene Matrix C mit $CX = I$ gilt

$$C\Omega C' - \left(X'\Omega^{-1}X\right) \text{ psds} \overset{\text{Loewner}}{:\Leftrightarrow} C\Omega C' \geq \left(X'\Omega^{-1}X\right)$$

Also gilt mit

$$\Sigma(\check{\boldsymbol{\beta}}|X) = \sigma^2 C\Omega C'$$

und

$$\Sigma(\tilde{\boldsymbol{\beta}}|X) = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1} \quad ,$$

dass die Varianz von $\tilde{\boldsymbol{\beta}}$ Loewner-kleiner ist als die Varianz von einem beliebigen linearen erwartungstreuen Schätzer $\check{\boldsymbol{\beta}}$. $\square$

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

Die Varianz des GLS-Schätzers beträgt

$$\Sigma(\tilde{\beta}|X) = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1}$$

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

GLS: Schätzer für σ^2

Falls die Matrix Ω bekannt ist, kann der Parameter σ^2 analog zu OLS geschätzt werden.

GLS-Residuen:

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

Schätzer $\tilde{\sigma}^2$ für σ^2:

$$\tilde{\sigma}^2 = \frac{1}{n-k-1}\tilde{\boldsymbol{u}}'\tilde{\boldsymbol{u}}$$

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 $\boldsymbol{\alpha}$ funktional abhängen und schätze $\boldsymbol{\alpha}$ durch $\hat{\boldsymbol{\alpha}}$.
- Berechne nun $\hat{\Omega} = \Omega(\hat{\boldsymbol{\alpha}})$ und berechne $\tilde{\boldsymbol{\beta}}^{FGLS}$ mit $\hat{\Omega}$.

Eigenschaften der OLS-, GLS- und FGLS-Schätzer

- Der OLS-Schätzer ist auch unter MLR 5' $\Sigma(\mathbf{u}) = \sigma^2\Omega$ (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 $\Omega(\boldsymbol{\alpha})$ unterstellt wird.