universität innsbruck

OLS in Matrixschreibweise

Grundlagen der Ökonometrie

www.hsto.info/econometrics

OLS in Matrixschreibweise

Wir wieder von einer einfachen linearen SRF aus (mit $i = 1, \dots, n$)

$$y_i = x_{i1}\widehat{\beta}_1 + x_{i2}\widehat{\beta}_2 + \cdots + x_{ih}\widehat{\beta}_h + \cdots + x_{ik}\widehat{\beta}_k + \hat{\varepsilon}_i$$

Dies ist eine Kurzschreibweise für

$$\begin{aligned}
y_1 &= x_{11}\widehat{\beta}_1 + x_{12}\widehat{\beta}_2 + \cdots + x_{1k}\widehat{\beta}_k + \hat{\varepsilon}_1 \\
y_2 &= x_{21}\widehat{\beta}_1 + x_{22}\widehat{\beta}_2 + \cdots + x_{2k}\widehat{\beta}_k + \hat{\varepsilon}_2 \\
\vdots &= \quad \vdots \quad + \quad \vdots \quad + \ddots + \quad \vdots \quad + \vdots \\
y_i &= x_{i1}\widehat{\beta}_1 + x_{i2}\widehat{\beta}_2 + \cdots + x_{ik}\widehat{\beta}_k + \hat{\varepsilon}_i \\
\vdots &= \quad \vdots \quad + \quad \vdots \quad + \ddots + \quad \vdots \quad + \vdots \\
y_n &= x_{n1}\widehat{\beta}_1 + x_{n2}\widehat{\beta}_2 + \cdots + x_{nk}\widehat{\beta}_k + \hat{\varepsilon}_n
\end{aligned}$$

mit $x_{\bullet 1} = \mathbf{1}$ (Regressionskonstante)

OLS in Matrixschreibweise

$$\begin{aligned}
y_1 &= x_{11}\widehat{\beta}_1 + x_{12}\widehat{\beta}_2 + \cdots + x_{1k}\widehat{\beta}_k + \hat{\varepsilon}_1 \\
y_2 &= x_{21}\widehat{\beta}_1 + x_{22}\widehat{\beta}_2 + \cdots + x_{2k}\widehat{\beta}_k + \hat{\varepsilon}_2 \\
\vdots &= \quad \vdots \quad + \quad \vdots \quad + \ddots + \quad \vdots \quad + \vdots \\
y_i &= x_{i1}\widehat{\beta}_1 + x_{i2}\widehat{\beta}_2 + \cdots + x_{ik}\widehat{\beta}_k + \hat{\varepsilon}_i \\
\vdots &= \quad \vdots \quad + \quad \vdots \quad + \ddots + \quad \vdots \quad + \vdots \\
y_n &= x_{n1}\widehat{\beta}_1 + x_{n2}\widehat{\beta}_2 + \cdots + x_{nk}\widehat{\beta}_k + \hat{\varepsilon}_n
\end{aligned}$$

in Matrixschreibweise:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix} + \begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_n \end{pmatrix}$$

oder kompakter

$$\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\varepsilon}}$$

Matrizen

Notation:

- Matrizen: fette Großbuchstaben
- Vektoren: fette Kleinbuchstaben

Spezielle Matrizen:

- Spaltenvektor: $(n \times 1)$; Zeilenvektor: $(1 \times n)$
- Quadratische Matrizen: $(n \times n)$
- Einheitsmatrix: $a_{ii} = 1$ für $i = 1, 2, \dots, n$ und $a_{ij} = 0$ wenn $i \neq j$, bzw.

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Spezialfall einer Diagonalmatrix

Matrizen**addition**

Seien $\mathbf{A}$ und $\mathbf{B}$ Matrizen der gleichen Dimension, dann ist $\mathbf{C} = \mathbf{A} + \mathbf{B}$

$$c_{ij} = a_{ij} + b_{ij}$$

z.B.

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$$

oder

$$\begin{pmatrix} 2 & 3 \\ 4 & 1 \end{pmatrix} + \begin{pmatrix} 3 & 1 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 5 & 4 \\ 6 & 5 \end{pmatrix}$$

Rechenregeln für Matrizenaddition

Für geeignet dimensionierte Matrizen gilt

- $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
- $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
- $\mathbf{A} + \mathbf{0} = \mathbf{A}$, wobei $\mathbf{0}$ die Nullmatrix ist, d.h. aus lauter Nullen besteht.
- $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$
- $(c_1 + c_2)\mathbf{A} = c_1\,\mathbf{A} + c_2\,\mathbf{A}$ (c_1 und c_2 sind Skalare)
- $c(\mathbf{A} + \mathbf{B}) = c\,\mathbf{A} + c\,\mathbf{B}$

Matrizen**multiplikation**

Dimension:

$$\underset{(n\times k)}{\mathbf{A}}\ \underset{(k\times m)}{\mathbf{B}} = \underset{(n\times m)}{\mathbf{C}}$$

$$\begin{pmatrix} a_{11} & \cdots & a_{1h} & \cdots & a_{1k} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ih} & \cdots & a_{ik} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nh} & \cdots & a_{nk} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1m} \\ \vdots & & \vdots & & \vdots \\ b_{g1} & \cdots & b_{gj} & \cdots & b_{gm} \\ \vdots & & \vdots & & \vdots \\ b_{k1} & \cdots & b_{kj} & \cdots & b_{km} \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1j} & \cdots & c_{1m} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \cdots & c_{ij} & \cdots & c_{im} \\ \vdots & & \vdots & & \vdots \\ c_{n1} & \cdots & c_{nj} & \cdots & c_{nm} \end{pmatrix}$$

$$c_{ij} = \sum_{r=1}^{k} a_{ir} b_{rj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj}$$

$$(\underbrace{n\times k}\overbrace{)(k\times}\underbrace{m}) $$

Matrixmultiplikation

Beispiel 1:

$$\underset{2\times 3}{\mathbf{A}} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \qquad \underset{3\times 2}{\mathbf{B}} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}$$

$$\underset{(2\times 2)}{\mathbf{C}} = \underset{(2\times 3)(3\times 2)}{\mathbf{A}\,\mathbf{B}} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix}$$

Beispiel 2:

$$\begin{pmatrix} 1 & 2 \\ 4 & 3 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 4 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 1 & 2 \\ 22 & 4 & 3 \\ -2 & 0 & -1 \end{pmatrix}$$

Matrixmultiplikation

Rechenregen für Matrizenmultiplikation:

Für geeignet dimensionierte Matrizen gilt

1. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ (Assoziativgesetz)
2. $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$ (linksseitiges Distributivgesetz)
3. $(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{BA}+\mathbf{CA}$ (rechtsseitiges Distributivgesetz)

Achtung: das Kommutativgesetz $\mathbf{AB} = \mathbf{BA}$ gilt im allgemeinen **nicht**!!!
Deshalb muss bei Matrizen zwischen *Vor- und Nachmultiplikation* unterschieden werden.
Zum Beispiel:

$$\mathbf{AB} = \begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 3 & 2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 6 & 2 \\ 2 & 0 \end{pmatrix}; \qquad \mathbf{BA} = \begin{pmatrix} 3 & 2 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 13 \\ 1 & 3 \end{pmatrix}$$

Transponierte einer Matrix

Die Transponierte einer Matrix $\mathbf{A}$, geschrieben $\mathbf{A}'$, erhält man durch Austauschen von Zeilen und Spalten, z.B.

$$\underset{2\times 3}{\mathbf{A}} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \quad \Rightarrow \quad \underset{3\times 2}{\mathbf{A}'} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}$$

Eine quadratische Matrix $\mathbf{A}$ ist symmetrisch wenn und nur wenn gilt $\mathbf{A}' = \mathbf{A}$.

Rechenregeln für Transponierte:

für geeignet dimensionierte Matrizen gilt:

1. $(\mathbf{A}')' = \mathbf{A}$
2. $(\mathbf{A}+\mathbf{B})' = \mathbf{A}'+\mathbf{B}'$
3. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, bzw. $(\mathbf{ABCD})' = \mathbf{D}'\mathbf{C}'\mathbf{B}'\mathbf{A}'$

Inverse einer Matrix

Multiplikation einer quadratischen $k \times k$ Matrix $\mathbf{A}$ mit ihrer Inversen $\mathbf{A}^{-1}$ (sofern diese existiert) gibt die Einheitsmatrix $\mathbf{I}$

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_k$$

Die Inverse einer Matrix existiert nur, wenn deren Determinante *nicht* den Wert Null hat, das heißt, wenn die Matrix **nicht singulär** ist!

Rechenregeln für Inverse:

Seien $\mathbf{A}$ und $\mathbf{B}$ invertierbare quadratische Matrizen, dann gilt

- $\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$
- $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ wenn beide Inversen existieren.
- $(c\mathbf{A})^{-1} = c^{-1}\mathbf{A}^{-1}$
- $\left|\mathbf{A}^{-1}\right| = \frac{1}{|\mathbf{A}|}$

OLS Schätzfunktion

Summe der quadrierten Residuen: $\longrightarrow$ inneres Produkt

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \begin{pmatrix} \hat{\varepsilon}_1 & \hat{\varepsilon}_2 & \cdots & \hat{\varepsilon}_n \end{pmatrix} \begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_n \end{pmatrix} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

mit $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$

Gesucht ist $\hat{\boldsymbol{\beta}}$ welcher $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$ minimiert:

$$\hat{\boldsymbol{\beta}} = \arg\min \left[(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})\right]$$

OLS Schätzfunktion

Quadratsumme der Residuen:

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} &= (\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) \\ &= (\boldsymbol{y}'-\hat{\boldsymbol{\beta}}'\boldsymbol{X}')(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{y}'\boldsymbol{y}-\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}-\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}}+\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} \\ &= \boldsymbol{y}'\boldsymbol{y}-2\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}+\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}\end{aligned}$$

mit $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = (\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y})' = \boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$ da alles Skalare! (beachte $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$)

OLS Schätzfunktion

Ableitung linearer und quadratischer Formen:

$$\frac{\partial(\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y})}{\partial\hat{\boldsymbol{\beta}}} = \boldsymbol{X}'\boldsymbol{y} \qquad \frac{\partial(\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}})}{\partial\hat{\boldsymbol{\beta}}} = 2\,\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$$

Also

$$\min_{\hat{\boldsymbol{\beta}}}(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}) = \min_{\hat{\boldsymbol{\beta}}}\left[\boldsymbol{y}'\boldsymbol{y}-2\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}+\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}\right]$$

$$\frac{\partial(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}})}{\partial\hat{\boldsymbol{\beta}}} = -2\boldsymbol{X}'\boldsymbol{y}+2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} \stackrel{!}{=} 0$$

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y} \quad \Rightarrow \quad (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

OLS-Punktschätzer

$$\boxed{\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}}$$

OLS Schätzfunktion

Beispiel: Die Matrix **X** sei ein $n \times 1$ Vektor mit lauter Einsen, also einer Regression auf die Regressionskonstante

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \left[\begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}\right]^{-1}\begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \\ &= \left(1^2+1^2\cdots+1^2\right)^{-1}(y_1+y_2+\cdots+y_n) \\ &= n^{-1}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}\end{aligned}$$

Gauss-Markov Annahmen

A1 **Linearität der PRF:** (korrekte Spezifikation)

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$$

A2 **Voller Spaltenrang:**

$$\mathrm{rk}(\boldsymbol{X}) = k$$

mit rk Rang: Maximalzahl linear unabhängiger Spaltenvektoren
impliziert $n \geq k$ und keine lineare Abhängigkeit zwischen Regressoren.
(Wenn man aus den Zeilen der $n \times k$ Matrix $\boldsymbol{X}$ eine nichtsinguläre $k \times k$ Matrix $\boldsymbol{X}'\boldsymbol{X}$ bilden kann (d.h. eine Matrix, deren Determinante nicht Null ist), hat $\boldsymbol{X}$ vollen Spaltenrang.)

A3 Exogene Regressoren

$$\mathrm{E}(\varepsilon_i|\boldsymbol{X}) = 0$$

(keine *'omitted variables'*, keine simultane Kausalität, etc.)

A4 i.i.d. **Störterme:**

$$\varepsilon_i \sim \text{i.i.d.}(0,\sigma^2)$$

Erwartungstreue

Einsetzen der PRF in Schätzfunktion

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\underbrace{(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})}_{\boldsymbol{y}} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \\ \mathrm{E}(\hat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} + \mathrm{E}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon})\end{aligned}$$

das heißt

$$\boxed{\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \qquad \text{wenn} \qquad \mathrm{E}(\boldsymbol{X}'\boldsymbol{\varepsilon}) = 0}$$

mit

$$\boldsymbol{X}'\boldsymbol{\varepsilon} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \sum_i x_{i1}\varepsilon_i \\ \sum_i x_{i2}\varepsilon_i \\ \vdots \\ \sum_i x_{ik}\varepsilon_i \end{pmatrix}$$

Varianz-Kovarianzmatrix von $\hat{\boldsymbol{\beta}}$

Definition der Varianz-Kovarianzmatrix von $\hat{\boldsymbol{\beta}}$:

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \mathrm{E}\left[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})\right]\left[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})\right]'$$

Beispiel: 2×2 Varianz-Kovarianzmatrix

$$\begin{aligned}\mathrm{var}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}\left[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})\right]\left[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})\right]' \\ &= \mathrm{E}\left[\begin{pmatrix} \widehat{\beta}_1 - \mathrm{E}(\widehat{\beta}_1) \\ \widehat{\beta}_2 - \mathrm{E}(\widehat{\beta}_2) \end{pmatrix}\begin{pmatrix} \widehat{\beta}_1 - \mathrm{E}(\widehat{\beta}_1) & \widehat{\beta}_2 - \mathrm{E}(\widehat{\beta}_2) \end{pmatrix}\right] \\ &= \begin{pmatrix} \mathrm{E}[\widehat{\beta}_1 - \mathrm{E}(\widehat{\beta}_1)]^2 & \mathrm{E}[\widehat{\beta}_1 - \mathrm{E}(\widehat{\beta}_1)][\widehat{\beta}_2 - \mathrm{E}(\widehat{\beta}_2)] \\ \mathrm{E}[\widehat{\beta}_1 - \mathrm{E}(\widehat{\beta}_1)][\widehat{\beta}_2 - \mathrm{E}(\widehat{\beta}_2)] & \mathrm{E}[\widehat{\beta}_2 - \mathrm{E}(\widehat{\beta}_2)]^2 \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{var}(\widehat{\beta}_1) & \mathrm{cov}(\widehat{\beta}_1, \widehat{\beta}_2) \\ \mathrm{cov}(\widehat{\beta}_1, \widehat{\beta}_2) & \mathrm{var}(\widehat{\beta}_2) \end{pmatrix}\end{aligned}$$

Varianz-Kovarianzmatrix von $\hat{\boldsymbol{\beta}}$

Herleitung:
Unter Gauss-Markov Annahmen gilt $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$ und $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$
Also:

$$\begin{aligned}\underset{(k\times k)}{\mathrm{var}(\hat{\boldsymbol{\beta}})} &= \mathrm{E}\left[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})\right]\left[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})\right]' \\ &= \mathrm{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}][(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}]' \\ &= \mathrm{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]\end{aligned}$$

da $(\boldsymbol{A}\boldsymbol{B})' = \boldsymbol{B}'\boldsymbol{A}'$ sowie $(\boldsymbol{A}')' = \boldsymbol{A}$; $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ ist symmetrisch
Für deterministische $\boldsymbol{X}$ (*sandwich formula*)

$$\boxed{\mathrm{var}(\hat{\boldsymbol{\beta}}) = \underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}}_{\text{bread}}\underbrace{\boldsymbol{X}'\,\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\boldsymbol{X}}_{\text{meat}}\underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}}_{\text{bread}}}$$

mit $\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$: Varianz-Kovarianzmatrix der *Störterme*!

Varianz-Kovarianzmatrix der **Störterme**

$$\underset{(n\times n)}{\mathrm{var}(\boldsymbol{\varepsilon})} = \mathrm{E}\begin{pmatrix} \varepsilon_1 - \mathrm{E}(\varepsilon_1) \\ \varepsilon_2 - \mathrm{E}(\varepsilon_2) \\ \vdots \\ \varepsilon_n - \mathrm{E}(\varepsilon_n) \end{pmatrix}\begin{pmatrix} \varepsilon_1 - \mathrm{E}(\varepsilon_1) & \varepsilon_2 - \mathrm{E}(\varepsilon_2) & \cdots & \varepsilon_n - \mathrm{E}(\varepsilon_n) \end{pmatrix}$$

$$= \mathrm{E}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}\begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \end{pmatrix} = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \qquad (\text{wenn } \mathrm{E}(\varepsilon_i) = 0)$$

$$\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \begin{pmatrix} \mathrm{E}(\varepsilon_1)^2 & \mathrm{E}(\varepsilon_1\varepsilon_2) & \dots & \mathrm{E}(\varepsilon_1\varepsilon_n) \\ \mathrm{E}(\varepsilon_2\varepsilon_1) & \mathrm{E}(\varepsilon_2)^2 & \dots & \mathrm{E}(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(\varepsilon_n\varepsilon_1) & \mathrm{E}(\varepsilon_n\varepsilon_2) & \dots & \mathrm{E}(\varepsilon_n)^2 \end{pmatrix} = \begin{pmatrix} \mathrm{var}(\varepsilon_1) & \mathrm{cov}(\varepsilon_1\varepsilon_2) & \dots & \mathrm{cov}(\varepsilon_1\varepsilon_n) \\ \mathrm{cov}(\varepsilon_2\varepsilon_1) & \mathrm{var}(\varepsilon_2) & \dots & \mathrm{cov}(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\varepsilon_n\varepsilon_1) & \mathrm{cov}(\varepsilon_n\varepsilon_2) & \dots & \mathrm{var}(\varepsilon_n) \end{pmatrix}$$

Varianz-Kovarianzmatrix der **Störterme**

$$\underset{(n\times n)}{\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')} = \begin{pmatrix} \mathrm{var}(\varepsilon_1) & \mathrm{cov}(\varepsilon_1\varepsilon_2) & \dots & \mathrm{cov}(\varepsilon_1\varepsilon_n) \\ \mathrm{cov}(\varepsilon_2\varepsilon_1) & \mathrm{var}(\varepsilon_2) & \dots & \mathrm{cov}(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\varepsilon_n\varepsilon_1) & \mathrm{cov}(\varepsilon_n\varepsilon_2) & \dots & \mathrm{var}(\varepsilon_n) \end{pmatrix}$$

Wenn A4: $\varepsilon_i \sim \text{i.i.d.}(0, \sigma^2)$

$$\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = \sigma^2 \boldsymbol{I}_n$$

Varianz-Kovarianzmatrix von $\hat{\boldsymbol{\beta}}$

$$\begin{aligned} \mathrm{var}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}\left[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})\right]\left[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})\right]' = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\underbrace{\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')}_{=\sigma^2\boldsymbol{I}}\boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \sigma^2\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right] \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \end{aligned}$$

Also: **wenn A4** $\varepsilon_i \sim \text{i.i.d.}(0, \sigma^2)$

$$\boxed{\mathrm{var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}}$$

aber: σ^2 unbeobachtbarer Parameter!
Erwartungstreue Schätzfunktion für σ^2 (siehe Skript):

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-k} = \frac{\sum_i \hat{\varepsilon}_i^2}{n-k}$$

OLS Schätzfunktion

OLS Schätzfunktion:

Mit Gauss-Markov Annahmen A1 – A4:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

$$\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}}) := \hat{\boldsymbol{\sigma}}^2_{\hat{\boldsymbol{\beta}}} = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-k}$$

Zusammenfassung

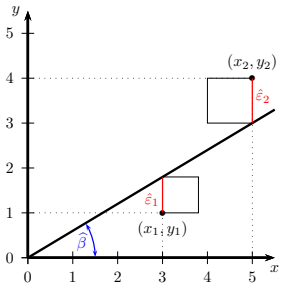
- OLS Schätzfunktion: $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$
- Varianz-Kovarianzmatrix von $\hat{\boldsymbol{\beta}}$ ohne i.i.d. Störterme
$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$
- Varianz-Kovarianzmatrix der Störterme $\boldsymbol{\varepsilon}$ *ohne* i.i.d. Störterme
$$\mathrm{var}(\boldsymbol{\varepsilon}) = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$$
- Varianz-Kovarianzmatrix der Störterme $\boldsymbol{\varepsilon}$ *mit* i.i.d. Störtermen
$$\mathrm{var}(\boldsymbol{\varepsilon}) = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{I}$$
- Varianz-Kovarianzmatrix von $\hat{\boldsymbol{\beta}}$ *mit* Gauss-Markov Annahmen
$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$
- Eine Schätzfunktion für σ^2 mit Gauss-Markov Annahmen
$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-k} = \frac{\sum_i \hat{\varepsilon}_i^2}{n-k}$$

Geometrische Interpretation der OLS Schätzfunktion

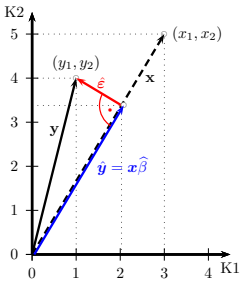
Beispiel: Sei $\boldsymbol{y} = \boldsymbol{x}\widehat{\beta} + \hat{\boldsymbol{\varepsilon}}$ mit

$$\boldsymbol{y} = \begin{pmatrix} 1 \\ 4 \end{pmatrix} \quad , \quad \boldsymbol{x} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$$

x, y - Raum:



Spaltenraum:

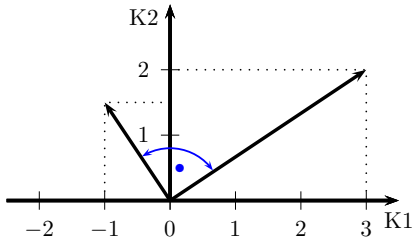


Geometrische Interpretation der OLS Schätzfunktion

Orthogonalität:
Zwei Vektoren $\boldsymbol{x}_1$ und $\boldsymbol{x}_2$ stehen rechtwinkelig aufeinander, d.h. sie sind *orthogonal* $\boldsymbol{x}_1 \perp \boldsymbol{x}_2$, wenn und nur wenn $\boldsymbol{x}_1'\boldsymbol{x}_2 = 0$

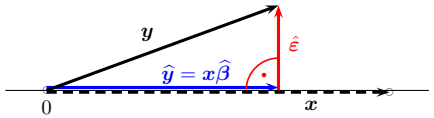
$$\boldsymbol{x}_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \quad , \quad \boldsymbol{x}_2 = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$$

$$\begin{aligned} \boldsymbol{x}_1'\boldsymbol{x}_2 &= \begin{pmatrix} 3 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 1.5 \end{pmatrix} \\ &= 3 \times (-1) + 2 \times 1.5 \\ &= 0 \end{aligned}$$



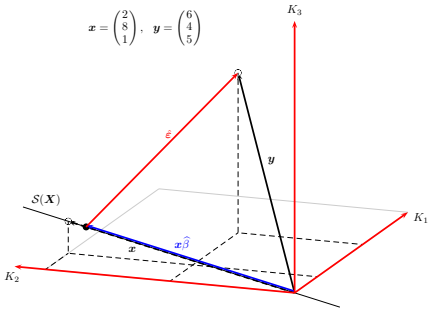
Geometrische Interpretation der OLS Schätzfunktion

OLS-Schätzfunktion $\widehat{\beta}$ als Ergebnis einer orthogonalen Projektion in den Spaltenraum (für $n = 2$ und $k = 1$)



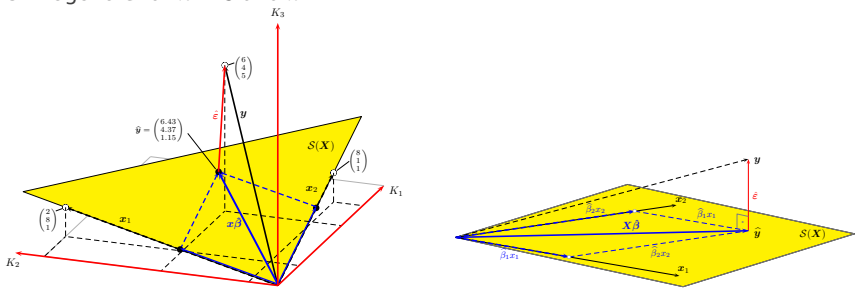
Geometrische Interpretation der OLS Schätzfunktion

Orthogonale für $n = 3$ und $k = 1$



Geometrische Interpretation der OLS Schätzfunktion

Orthogonale für $n = 3$ und $k = 2$



Zwei wichtige Matrizen: M & P

Residuenerzeugende Matrix M:

Einsetzen der OLS Schätzfunktion $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ in $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ gibt

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \underbrace{\left[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right]}_{\boldsymbol{M}}\boldsymbol{y}$$

Also

$$\underset{(n\times n)}{\boldsymbol{M}} := \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

und

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{M}\boldsymbol{y}$$

Zwei wichtige Matrizen: M & P

Eigenschaften der residuenerzeugenden Matrix M:

- $\boldsymbol{M}$ ist symmetrisch, d.h. $\boldsymbol{M} = \boldsymbol{M}'$
- $\boldsymbol{M}$ ist idempotent, d.h. $\boldsymbol{M}\boldsymbol{M} = \boldsymbol{M}$
- Die Matrizen $\boldsymbol{M}$ und $\boldsymbol{X}$ sind orthogonal, d.h. $\boldsymbol{M}\boldsymbol{X} = \mathbf{O}$

$$\text{weil} \quad \boldsymbol{M}\boldsymbol{X} = \left[I - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right]\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = \mathbf{O}$$

Daraus folgt auch

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{M}\boldsymbol{\varepsilon}$$

$$\text{weil} \quad \hat{\boldsymbol{\varepsilon}} = \boldsymbol{M}\boldsymbol{y} = \boldsymbol{M}(\underbrace{\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}}_{\boldsymbol{y}}) = \mathbf{O}\boldsymbol{\beta} + \boldsymbol{M}\boldsymbol{\varepsilon} = \boldsymbol{M}\boldsymbol{\varepsilon}$$

- $\boldsymbol{M}$ ist singulär mit Rang $n - k$

Zwei wichtige Matrizen: M & P

Projektionsmatrix P: (Hutmatrix)

$$\underset{(n\times n)}{\boldsymbol{P}} := \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{I} - \boldsymbol{M}$$

$$\widehat{\boldsymbol{y}} = \boldsymbol{y} - \hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{M}\boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{M})\boldsymbol{y} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{P}\boldsymbol{y}$$

$\rightarrow$ Hutmatrix (*hat matrix*)

Eigenschaften der Projektionsmatrix P:

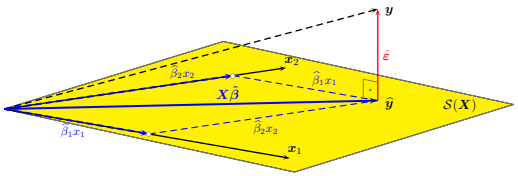
- symmetrisch $\boldsymbol{P} = \boldsymbol{P}'$
- idempotent $\boldsymbol{P} = \boldsymbol{P}\boldsymbol{P}$
- $\boldsymbol{P}$ und $\boldsymbol{M}$ orthogonal, d.h. $\boldsymbol{P}\boldsymbol{M} = \boldsymbol{M}\boldsymbol{P} = \mathbf{O}$
- $\boldsymbol{P}$ ist singulär mit Rang k

Zwei wichtige Matrizen: M & P

$\boldsymbol{P}$ und $\boldsymbol{M}$ werden für zahlreiche Beweise benötigt.

Darüber hinaus haben sie eine grafische Interpretation (orthogenale Zerlegung):

$$\begin{aligned} \boldsymbol{y} &= \widehat{\boldsymbol{y}} + \hat{\boldsymbol{\varepsilon}} \\ &= \boldsymbol{P}\boldsymbol{y} + \boldsymbol{M}\boldsymbol{y} \\ &= \text{Projektion} + \text{Residuum} \end{aligned}$$



Matrix transform

$$\begin{bmatrix} \cos 90^\circ & \sin 90^\circ \\ -\sin 90^\circ & \cos 90^\circ \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

Quelle: `https://xkcd.com/184/`

Thanks!