

# Empirische Wirtschaftsforschung

## Die OLS-Schätzung in Matrixschreibweise

Ulrich Fritsche

Universität Hamburg

Email: [Ulrich.Fritsche@wiso.uni-hamburg.de](mailto:Ulrich.Fritsche@wiso.uni-hamburg.de)

## Wiederholung: Der OLS-Schätzer in Skalarschreibweise

Im einfachsten Schätzmodell schreibt man für das tatsächliche Modell:

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_m x_{tm} + u_t; t = 1, 2, \dots, T \quad (1)$$

mit  $t$  als Anzahl der Beobachtungen, und  $m$  als Anzahl der Regressoren.

Der OLS-Schätzer minimiert die Residuenquadratsumme (RSS) für das geschätzte Modell:

$$\min_{(\hat{\beta}_1, \dots, \hat{\beta}_m)} \sum_{t=1}^T (y_t - b_1 - b_2 x_{t2} - \dots - b_m x_{tm})^2 = \sum_{t=1}^T \hat{u}_t^2 \quad (2)$$

Durch Berechnung der partiellen Ableitungen ergibt sich das **Normalgleichungssystem**:

$$Tb_1 + b_2 \sum_{t=1}^T x_{t2} + \dots + b_m \sum_{t=1}^T x_{tm} - \sum_{t=1}^T y_t = 0 \quad (3)$$

$$b_1 \sum_{t=1}^T x_{t2} + b_2 \sum_{t=1}^T x_{t2}^2 + \dots + b_m \sum_{t=1}^T x_{tm}^2 - \sum_{t=1}^T x_{t2} y_t = 0$$

⋮

$$b_1 \sum_{t=1}^T x_{tm} + b_2 \sum_{t=1}^T x_{tm} x_{t2} + \dots + b_m \sum_{t=1}^T x_{tm}^2 - \sum_{t=1}^T x_{tm} y_t = 0$$

Dieses System lässt sich für  $m = 2$  analytisch lösen:

$$\sum_{t=1}^T y_t = b_1 T + b_2 \sum_{t=1}^T x_{t2} \quad (4)$$

$$\sum_{t=1}^T y_t x_{t2} = b_1 \sum_{t=1}^T x_{t2} + b_2 \sum_{t=1}^T x_{t2}^2$$

und man erhält als OLS-Schätzer:

$$b_1 = \bar{y} - b_2 \bar{x}_2; \quad (5)$$

mit  $\bar{y}, \bar{x} = \frac{1}{T} \sum_{t=1}^T y_t, x_{t2}$

und

$$b_2 = \frac{\text{Cov}(x_2, y)}{\text{Var}(x_2)} \quad (6)$$

# Der OLS-Schätzer in Matrixform

Das multivariate Schätzmodell

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_m x_{tm} + u_t; t = 1, 2, \dots, T \quad (1)$$

lässt sich kompakter in Matrixform schreiben als

$$\mathbf{y}_{T \times 1} = \mathbf{X}_{T \times m} \boldsymbol{\beta}_{m \times 1} + \mathbf{u}_{T \times 1} \quad (7)$$

Ausgeschrieben:

$$\mathbf{y}_{T \times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \quad \mathbf{X}_{T \times m} = \begin{pmatrix} 1 & X_{12} & \cdots & X_{1m} \\ 1 & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{T2} & \cdots & X_{Tm} \end{pmatrix} \quad \beta_{m \times 1} =$$
$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} \quad \mathbf{u}_{T \times 1} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}$$

Das tatsächliche Modell:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (8)$$

Das geschätzte Modell:

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}} \quad (9)$$

Minimierung der RSS:

$$\min_{\hat{\boldsymbol{\beta}}} \text{RSS} = \sum_{t=1}^T \hat{u}_t^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} \quad (10)$$

$\hat{\mathbf{u}}'\hat{\mathbf{u}}$  ist das **Skalarprodukt** oder **inneres Produkt** und  $'$  die **Transponierte** einer Matrix oder eines Vektors, d.h., Zeilen und Spalten werden vertauscht. Es gilt  $(\mathbf{A}')' = \mathbf{A}$   
Ausgeschrieben:

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = (\hat{u}_1 \hat{u}_2 \cdots \hat{u}_T) \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_T \end{pmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \dots + \hat{u}_T^2 = \sum_{t=1}^T (\hat{u}_t)^2$$

Durch Einsetzen von Gleichung (9) in (10) folgt

$$\hat{\mathbf{u}}' \hat{\mathbf{u}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (11)$$

**Matrixrechenregeln:**

Es gilt:

1.  $(\mathbf{A} \pm \mathbf{B})' = \mathbf{A}' \pm \mathbf{B}'$
2.  $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

Damit wird Gleichung (11) zu

$$\hat{\mathbf{u}}' \hat{\mathbf{u}} = (\mathbf{y}' - (\hat{\boldsymbol{\beta}}'\mathbf{X}')) (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (12)$$

Mit

$$3. (\mathbf{B} \pm \mathbf{C})\mathbf{A} = \mathbf{BA} \pm \mathbf{CA}$$

ergibt sich durch Ausmultiplizieren:

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \quad (13)$$

Unter Anwendung von Regel (2) lässt sich schreiben:  
(Hinweis: Das Produkt  $\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}$  ist ein Skalar:

$$\mathbf{y}'_{1 \times T} \mathbf{X}_{T \times m} \hat{\boldsymbol{\beta}}_{m \times 1} = a_{1 \times 1})$$

$$\begin{aligned} \hat{\mathbf{u}}'\hat{\mathbf{u}} &= \mathbf{y}'\mathbf{y} - (\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}})' - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \end{aligned} \quad (14)$$

Ableitungsregeln für Matrizen:

4. Skalarprodukte:  $\frac{\partial \mathbf{a}'\beta}{\partial \beta} = \frac{\partial \beta' \mathbf{a}}{\partial \beta} = \mathbf{a}$ , mit  $\mathbf{a}$  als Vektor.

5. Quadratische Form:  $\frac{\partial \beta' \mathbf{A} \beta}{\partial \beta} = (\mathbf{A} + \mathbf{A}')\beta$

Mit  $\mathbf{A}$  symmetrisch, i.e.  $\mathbf{A} = \mathbf{A}'$ :  $\frac{\partial \beta' \mathbf{A} \beta}{\partial \beta} = 2\mathbf{A}\beta$ .

Damit haben wir:

$$\frac{\partial \hat{\mathbf{u}}' \hat{\mathbf{u}}}{\partial \beta} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = 0 \quad (15)$$

da mit  $(\mathbf{X}'_{m \times T} \mathbf{X}_{T \times m})' = (\mathbf{X}'\mathbf{X})_{m \times m}$  Symmetrie erfüllt ist.

## Matrixinversion:

Eine quadratische Matrix  $A$  heißt **invertierbar** wenn gilt:

$$6. \mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$


Dies entspricht **ungefähr** der Division im Skalarfall.<sup>1</sup>

Die Inverse der quadratischen Matrix  $\mathbf{A}$  kann man berechnen mit

$$7. \mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \cdot \mathit{adj}(\mathbf{A})$$

, wobei  $|\mathbf{A}|$  die **Determinante** von  $\mathbf{A}$  ist.

---

<sup>1</sup>Für rechteckige Matrizen lassen sich s.g. Pseudoinverse berechnen. 

Damit  $\mathbf{A}$ , bzw.  $(\mathbf{X}'\mathbf{X})_{m \times m}$  invertierbar ist, muss gelten:

$$\text{Rang}(\mathbf{X}'\mathbf{X}) = m,$$

d.h., die Spalten der Matrix  $\mathbf{X}'\mathbf{X}$  müssen unabhängig sein.<sup>2</sup>

Auflösen von Gleichung (15) ergibt den OLS-Schätzer:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (16)$$

---

<sup>2</sup>Ist dies nicht der Fall, spricht man von Multikollinearität.

# BLUE-Eigenschaften des multiplen OLS-Schätzers

Erwartungstreue (Unbiasedness) des OLS-Schätzers  $\hat{\beta}$ :

$$E(\hat{\beta}) = E(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (17)$$

Aus der Gleichung des tatsächlichen Modells (8):

$$\begin{aligned} E(\hat{\beta}) &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u})) & (18) \\ &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + \mathbf{X}'\mathbf{X}^{-1}\mathbf{X}'\mathbf{u}) \\ &= E(\mathbf{I}\beta) + E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}) \\ &= \beta \end{aligned}$$

mit der Regel  $E(b + aX) = b + aE(X)$ , der Annahme, dass die Variablen in  $X$  nicht stochastisch sind, und dass für die Störterme gilt  $\mathbf{u} \sim N(0, \sigma^2)$ .

**Effizienz (Best)** des OLS-Schätzers:

Matrixrechenregel:

$$8. ((\mathbf{X}'\mathbf{X})^{-1})' = ((\mathbf{X}'\mathbf{X})')^{-1} = (\mathbf{X}'\mathbf{X})^{-1}.$$

$$\begin{aligned} \text{Var}(\hat{\beta}) &= E \left[ (\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))' \right] && (19) \\ &= E \left[ ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) - \beta)((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) - \beta)' \right] \\ &= E \left[ ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta - \beta) \right. \\ &\quad \left. ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta - \beta)' \right] \\ &= E \left[ ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})' \right] \\ &= E \left[ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \right] \end{aligned}$$

Unter der Annahme nicht-stochastischer erklärender Variablen kann man schreiben:

$$\text{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' E[\mathbf{u}\mathbf{u}'] \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad (20)$$

Der Ausdruck  $E[\mathbf{u}\mathbf{u}']$  ist das **äußere Produkt**. Dieses wird berechnet als:

$$\begin{aligned} E[\mathbf{u}\mathbf{u}'] &= E \begin{bmatrix} u_1 u_1 & u_1 u_2 & \cdots & u_1 u_T \\ u_2 u_1 & u_2 u_2 & \cdots & u_2 u_T \\ \vdots & \vdots & \ddots & \vdots \\ u_T u_1 & u_T u_2 & \cdots & u_T u_T \end{bmatrix} \\ &= \begin{bmatrix} E[(u_1)^2] & E[u_1 u_2] & \cdots & E[u_1 u_T] \\ E[u_2 u_1] & E[(u_2)^2] & \cdots & E[u_2 u_T] \\ \vdots & \vdots & \ddots & \vdots \\ E[u_T u_1] & E[u_T u_2] & \cdots & E[(u_T)^2] \end{bmatrix} \end{aligned}$$

Da per Definition:

$$\text{Var}(u_i) = E[(u_i - E[u_i])(u_i - E[u_i])]$$

führt die Annahme  $E[u_i] = 0$

zu  $\text{Var}(u_i) = E[u_i^2]$  für  $i = j$

Analog wird

$$\text{Cov}(u_i, u_j) = E[(u_i - E[u_i])(u_j - E[u_j])]$$

unter  $E[u_i], E[u_j] = 0$

zu  $\text{Cov}(u_i, u_j) = E[u_i u_j]$  für  $i \neq j$ .

## Die Annahmen

- ▶  $Var(u_i) = \sigma_u^2$  konstant  $\forall i$ : Homoskedastie
- ▶  $Cov(u_i, u_j) = 0 \forall i \neq j$ : Keine Autokorrelation

führen zur **Varianz-Kovarianz-Matrix von  $\mathbf{u}$** :

$$\mathbf{V}(\mathbf{u}) = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \sigma^2 \mathbf{I}_T$$

Damit kann man Gleichung (20)

$$\text{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' E[\mathbf{u}\mathbf{u}'] \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad (20)$$

schreiben als

$$\text{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \sigma^2 \mathbf{I} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad (21)$$

Da  $\mathbf{A}\mathbf{a}\mathbf{B} = \mathbf{a}(\mathbf{A}\mathbf{B})$ , und  $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{I}$  erhält man die **Varianz-Kovarianz von  $\beta$**  als :

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad (22)$$


Da  $\sigma^2$  unbekannt ist, muss man auch hierfür einen Schätzer verwenden. Diesen gewinnt man über die Residuen

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\mathbf{b} \quad (23)$$

Problem:  $\text{Var}[\mathbf{u}] \neq \text{Var}[\hat{\mathbf{u}}]$

$$\begin{aligned} \hat{\sigma}^2(T - m) &= \hat{\mathbf{u}}'\hat{\mathbf{u}} & (24) \\ E[\hat{\sigma}^2](T - m) &= E[\hat{\mathbf{u}}'\hat{\mathbf{u}}] \\ \hat{\sigma}^2 \frac{T - m}{T} &= \frac{1}{T} \hat{\mathbf{u}}'\hat{\mathbf{u}} \\ \hat{\sigma}^2 &= \frac{T}{T - m} \frac{1}{T} \hat{\mathbf{u}}'\hat{\mathbf{u}} \\ \hat{\sigma}^2 &= \frac{1}{T - m} \hat{\mathbf{u}}'\hat{\mathbf{u}} \end{aligned}$$

Dass heißt, der Term  $\frac{1}{T-m}$  korrigiert den Bias, und man erhält einen erwartungstreuen Schätzer für die unbekannte Varianz der Regression.<sup>3</sup>

<sup>3</sup>Siehe von Auer (2007, S.194/195) für einen ausführlichen Beweis. 

# Zusammenfassung

1. Der OLS-Schätzer in Matrixschreibweise:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (25)$$

2. Die Varianz-Kovarianz-Matrix von  $\beta$ :

$$\begin{aligned} \mathbf{V}(\beta) &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \\ &= \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad (26)$$

3. OLS als Best-Linear-Unbiased-Estimator im multivariaten Fall
4. Literatur: Verbeek (2008): Kapitel 2.1 - 2.5