Skript für

Kapitel 3 (**Ökonometrie**, MSc.)

Dr. Lars Metzger

Stand 12. Mai 2025

Inhaltsverzeichnis

1 Das muliple lineare Regressionsmodell

1.1 Vor- und Nachteile der multiplen linearen Regression

1.1.1 Vorteile

Allgemeinere Funktionale Formen für das ökonometrische Modell sind möglich.

Erweiterung des Erklärungsgehalts des Regressionsmodells. Kausale Beziehungen zum Regressanden, die zuvor nicht enthalten waren, können nun zusätzlich genutzt werden um die Variation des Regressanden zu erklären.

Plausiblere Begründung der Annahme der strikten Exogenität.

1.1.2 Nachteil

Bei multiplen Regressoren verändert sich der Rechenaufwand. Wird bei der Summennotation verharrt, so werden die Rechnungen aufwändiger und unübersichtlicher.

1.2 KQ bei zwei Regressoren und Summennotation

1.2.1 Notation

Es sei $\overline{a} = \frac{1}{n}\sum_{i=1}^{n} a_i$ und $\overline{ab} = \sum_{i=1}^{n} a_i b_i$.

Es sei $s_{aa} = \overline{aa} - \overline{a} \cdot \overline{a}$ und $s_{ab} = \overline{ab} - \overline{a} \cdot \overline{b}$.

1.2.2 Regressionsmodell

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + u_i,\ i = 1, \ldots, n\ ,$$

wobei x_i und z_i zwei unterschiedliche Regressoren mit $|\rho_{xz}| \neq 1$ und $s_{xx}, s_{zz} > 0$ seien. Für beliebige Parameter b_0, b_1 und b_2 sei die Prognose $\hat{y}_i$ definiert durch:

$$\hat{y}_i = b_0 + b_1 x_i + b_2 z_i$$

Der Prognosefehler (Residuum) $\hat{u}_i$ beträgt dann:

$$\hat{u}_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_i + b_2 z_i)$$

Die Summe der quadrierten Residuen ist:

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i - b_2 z_i\right)^2$$

Die kleinsten Quadrate (KQ)-Schätzer $\hat{\beta}_0, \hat{\beta}_1$ und $\hat{\beta}_2$ für β_0, β_1 und β_2 minimieren nun die Summe der quadrierten Residuen und erfüllen die notwendigen Bedingungen:

$$\sum_{i=1}^{n}(-2) \cdot \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 z_i\right) = 0$$

$$\sum_{i=1}^{n}(-2x_{i1}) \cdot \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 z_i\right) = 0$$

$$\sum_{i=1}^{n}(-2x_{i2}) \cdot \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 z_i\right) = 0$$

Beidseitiges Teilen durch n:

$$\begin{array}{rcccccc} \overline{y} & = & \hat{\beta}_0 & + & \hat{\beta}_1 \overline{x} & + & \hat{\beta}_2 \overline{z} \\ \overline{xy} & = & \hat{\beta}_0 \overline{x} & + & \hat{\beta}_1 \overline{xx} & + & \hat{\beta}_2 \overline{xz} \\ \overline{zy} & = & \hat{\beta}_0 \overline{z} & + & \hat{\beta}_1 \overline{xz} & + & \hat{\beta}_2 \overline{zz} \end{array}$$

Die erste Gleichung kann leicht nach b_0 aufgelöst werden:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} - \hat{\beta}_2 \overline{z}$$

und $\hat{\beta}_0$ kann dann in die beiden verbleibenden Gleichungen eingesetzt werden. Es ergibt sich dann:

$$\begin{array}{rcl} s_{xy} & = & \hat{\beta}_1 \cdot s_{xx} + \hat{\beta}_2 \cdot s_{xz} \\ s_{zy} & = & \hat{\beta}_1 \cdot s_{xz} + \hat{\beta}_2 \cdot s_{zz} \end{array}$$

Unter der Voraussetzung $s_{xx}s_{zz} - s_{xz}s_{zx} \neq 0 \Leftrightarrow |\rho_{xz}| \neq 1$ hat dieses System die eindeutige Lösung $(\hat{\beta}_1, \hat{\beta}_2)$:

$$\hat{\beta}_1 = \left(s_{zz}s_{xy} - s_{xz}s_{zy}\right) / \left(s_{xx}s_{zz} - s_{xz}s_{xz}\right)$$

$$\hat{\beta}_2 = \left(s_{xx}s_{zy} - s_{xz}s_{xy}\right) / \left(s_{xx}s_{zz} - s_{xz}s_{xz}\right)$$

und

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} - \hat{\beta}_2 \overline{z}$$

1.3 Matrizen und Vektoren

Eine **Matrix** ist eine Tabelle mit Zeilen und Spalten. Die Anzahl der Zeilen und Spalten definiert die **Ordnung** der Matrix. Die Einträge der Matrix sind reelle Zahlen. Wir bezeichnen mit a_{ij} den Eintrag der i-ten Zeile und j-ten Spalte der Matrix A.

$$\underset{m\times n}{\mathbf{A}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Besteht die Matrix nur aus einer Zeile oder nur aus einer Spalte, sprechen wir von einem **Zeilen**- bzw. von einem **Spaltenvektor**. Wenn nicht spezifiziert ist, ob $\mathbf{v}$ ein Zeilen- oder ein Spaltenvektor ist, so handelt es sich um einen Spaltenvektor. Da Vektoren auch Matrizen sind, gelten die Rechenregeln für Matrizen auch für Vektoren. Besteht die Matrix nur aus einer Zeile und einer Spalte, handelt es sich um einen **Skalar**.

Eine Matrix ist **quadratisch**, falls die Anzahl ihrer Zeilen der Anzahl ihrer Spalten entspricht.

Die **Transponierte** $\mathbf{A}'$ einer Matrix $\mathbf{A}$ wird erreicht, indem Zeilen und Spalten vertauscht werden:

$$\underset{n\times m}{\mathbf{A}'} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}' = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{pmatrix}$$

Eine Matrix $\mathbf{A}$ ist **symmetrisch**, falls $\mathbf{A} = \mathbf{A}'$. Symmetrische Matrizen sind notwendig quadratisch.

1.3.1 Besondere Matrizen und Vektoren

Die **Einheitsmatrix** der Ordnung n, I_n, ist eine quadratische $n \times n$-Matrix, bei welcher auf der Diagonalen nur einsen stehen und abseits der Diagonalen nur nullen.

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

Die **Nullmatrix** besteht nur aus Nullen, sie muss nicht notwendig quadratisch sein und wird mit $\mathbf{N}$ bezeichnet:

$$\mathbf{N} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

Der **Einsvektor** ist ein Vektor mit m Komponenten, welche alle gleich eins sind:

$$\boldsymbol{\iota} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

1.3.2 Addition von Matrizen

Die Summe zweier Matrizen der gleichen Ordnung ist eine Matrix dieser Ordnung. Das Element der i-ten Zeile und j-ten Spalte der Summe entspricht der Summe der jeweiligen Elemente der i-ten Zeile und j-ten Spalte der beiden Matrizen:

$$\mathbf{A} + \mathbf{B} = \mathbf{C}$$

mit $c_{ij} = a_{ij} + b_{ij}$. Es gilt $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.

1.3.3 Multiplikation einer Matrix mit einem Skalar

Das Ergebnis der Multiplikation einer Matrix $\mathbf{A}$ mit einem Skalar b ist die Multiplikation jedes Elements der Matrix mit diesem Skalar:

$$b \cdot \mathbf{A} = \mathbf{C}$$

mit $c_{ij} = b \cdot a_{ij}$. Es gilt $b \cdot \mathbf{A} = \mathbf{A} \cdot b$.

Eine **Linearkombination** zweier Vektoren gleicher Ordnung ist die Summe des Produktes der zwei Vektoren mit Skalaren:

$$\mathbf{v} \cdot a + \mathbf{w} \cdot b = \begin{pmatrix} v_1 \cdot a + w_1 \cdot b \\ v_2 \cdot a + w_2 \cdot b \\ \vdots \\ v_n \cdot a + w_n \cdot b \end{pmatrix},$$

für $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ und $a, b \in \mathbb{R}$.

1.3.4 Multiplikation zweier Vektoren (Skalarprodukt)

Für zwei Vektoren $\mathbf{v}$ und $\mathbf{w}$ der gleichen Ordnung n, $\mathbf{v}$ ein Zeilenvektor und $\mathbf{w}$ ein Spaltenvektor, ist das Skalarprodukt die Summe der Produkte der Komponenten gleichen Indexes:

$$\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} v_i \cdot w_i$$

Es gilt $\mathbf{v} \cdot \mathbf{w} = 0$ genau dann, wenn $\mathbf{v}$ und $\mathbf{w}$ senkrecht zueinander sind.

Zudem gilt $\mathbf{w}' \cdot \mathbf{w} \geq 0$ und $\mathbf{w}' \cdot \mathbf{w} = 0 \Leftrightarrow w_i = 0$ für $i = 1, \ldots, n$.

$||\mathbf{w}|| = \sqrt{\mathbf{w}' \cdot \mathbf{w}}$ bezeichnet die Euklidische Norm eines Vektors $\mathbf{w} \in \mathbb{R}^n$.

1.3.5 Multiplikation zweier Matrizen

Für zwei Matrizen $\mathbf{A} \in \mathbb{R}^{m \times n}$ und $\mathbf{B} \in \mathbb{R}^{n \times o}$ gilt für das **Matrixprodukt**:

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{C} \in \mathbb{R}^{m \times o} \ ,$$

wobei c_{ij} das Skalarprodukt der i-ten Zeile von $\mathbf{A}$ mit der j-ten Spalte von $\mathbf{B}$ ist:

$$c_{ij} = \sum_{t=1}^{n} a_{it} \cdot b_{tj}$$

Multiplikation einer Matrix mit einem Spaltenvektor von rechts

Das Produkt einer Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ mit einem Vektor $\mathbf{v} \in \mathbb{R}^n$ ist die Linearkombination aller Spalten von A, wobei der Kofaktor der j-ten Spalte von $\mathbf{A}$ die j-te Komponente von $\mathbf{v}$ ist:

$$\begin{aligned}
\mathbf{A} \cdot \mathbf{v} &= \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} \cdot v_1 + \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} \cdot v_2 + \ldots + \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \cdot v_n \\
&= \begin{pmatrix} a_{11} \cdot v_1 + a_{12} \cdot v_2 + \ldots + a_{1n} \cdot v_n \\ a_{21} \cdot v_1 + a_{22} \cdot v_2 + \ldots + a_{2n} \cdot v_n \\ \vdots \\ a_{m1} \cdot v_1 + a_{m2} \cdot v_2 + \ldots + a_{mn} \cdot v_n \end{pmatrix}
\end{aligned}$$

Wichtig: Die Anzahl der Komponenten des Vektors $\mathbf{v}$ muss der Anzahl der Spalten von $\mathbf{A}$ entsprechen!

Multiplikation eine Matrix mit einem Zeilenvektor von links

Dies ist analog zur Multiplikation mit einem Vektor von rechts. Abweichend muss der von links zu multiplizierende Vektor $\mathbf{v}$ ein Zeilenvektor sein und dessen Anzahl von Komponenten muss der Anzahl der Zeilen der Matrix entsprechen.

Das Ergebnis ist dann eine Linearkombination der Zeilen der Matrix.

1.3.6 Transponieren von Matrixprodukten

Sei $\underset{m\times o}{\mathbf{C}} = \mathbf{AB}$ das Produkt der Matrizen $\underset{m\times n}{\mathbf{A}}$ und $\underset{n\times o}{\mathbf{B}}$. Dann ist das transponierte Matrixprodukt gegeben durch

$$\underset{o\times m}{\mathbf{C}'} = (\mathbf{AB})' = \underset{o\times n}{\mathbf{B}'}\underset{n\times m}{\mathbf{A}'}$$

1.4 Stochastische Matrizen und Vektoren

1.4.1 Erwartungswerte und bedingte Erwartungswerte

Seien a_{ij} Zufallsvariablen mit Erwartungswerten $\mathbb{E}[a_{ij}]$ für $i = 1, \ldots, m$, $j = 1, \ldots, n$. Dann ist die Matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

eine stochastische Matrix mit Erwartungswert

$$\mathbb{E}[\mathbf{A}] = \begin{pmatrix} \mathbb{E}[a_{11}] & \mathbb{E}[a_{12}] & \ldots & \mathbb{E}[a_{1n}] \\ \mathbb{E}[a_{21}] & \mathbb{E}[a_{22}] & \ldots & \mathbb{E}[a_{2n}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[a_{m1}] & \mathbb{E}[a_{m2}] & \ldots & \mathbb{E}[a_{mn}] \end{pmatrix}$$

1.4.2 Rechenregeln für den Erwartungswert

Sei $\mathbf{A}$ eine stochastische Matrix der Ordnung $m \times n$ mit Erwartungswert $\mathbb{E}[\mathbf{A}]$. Sei $\mathbf{B}$ eine feste Matrix der Ordnung $o \times m$ und sei $\mathbf{C}$ eine feste Matrix der Ordnung $o \times n$. Dann gilt:

$$\mathbb{E}\left[\mathbf{B} \cdot \mathbf{A} + \mathbf{C}\right] = \mathbf{B} \cdot \mathbb{E}\left[\mathbf{A}\right] + \mathbf{C}$$

Sei $\mathbf{v}$ ein stochastischer Vektor und $\mathbf{A}$ eine stochastische Matrix mit passender Ordnung. Wir benutzen häufig den auf $\mathbf{A}$ bedingten Erwartungswert einer Funktion von $\mathbf{v}$. Ist $\mathbf{A}$ Teil dieser Funktion, so kann $\mathbf{A}$ im auf $\mathbf{A}$ bedingten Erwartungswert als konstant betrachtet werden:

$$\mathbb{E}[\mathbf{A} \cdot \mathbf{v} | \mathbf{A}] = \mathbf{A} \cdot \mathbb{E}[\mathbf{v} | \mathbf{A}]$$

Wird der Erwartungswert bezüglich $\mathbf{A}$ des auf $\mathbf{A}$ bedingten Erwartungswertes bezüglich $\mathbf{v}$ gebildet, so ist dies der unbedingte Erwartungswert von $\mathbf{v}$:

$$\mathbb{E}[\mathbb{E}[\mathbf{v} | \mathbf{A}]] = \mathbb{E}[\mathbf{v}]$$

1.4.3 Kovarianz-Matrizen von Vektoren

Die Kovarianz-Matrix zweier stochastischer Vektoren $\mathbf{v}$ und $\mathbf{w}$ gleicher Ordnung n ist definiert durch:

$$\boldsymbol{\Sigma}(\mathbf{v}, \mathbf{w}) = \mathbb{E}\left[\left(\mathbf{v} - \mathbb{E}[v]\right)\left(\mathbf{w} - \mathbb{E}[w]\right)'\right]$$

Ausmultiplizieren ergibt

$$\boldsymbol{\Sigma}(\mathbf{v}, \mathbf{w}) = \mathbb{E}\left[\begin{pmatrix} v_1 - \mathbb{E}[v_1] \\ v_2 - \mathbb{E}[v_2] \\ \vdots \\ v_n - \mathbb{E}[v_n] \end{pmatrix} (w_1 - \mathbb{E}[w_1], w_2 - \mathbb{E}[w_2], \ldots, w_n - \mathbb{E}[w_n])\right]$$

$$= \mathbb{E}\left[\begin{array}{cccc} (v_1 - \mathbb{E}[v_1])(w_1 - \mathbb{E}[w_1]) & (v_1 - \mathbb{E}[v_1])(w_2 - \mathbb{E}[w_2]) & \ldots & (v_1 - \mathbb{E}[v_1])(w_n - \mathbb{E}[w_n]) \\ (v_2 - \mathbb{E}[v_2])(w_1 - \mathbb{E}[w_1]) & (v_2 - \mathbb{E}[v_2])(w_2 - \mathbb{E}[w_2]) & \ldots & (v_2 - \mathbb{E}[v_2])(w_n - \mathbb{E}[w_n]) \\ \vdots & \vdots & \ddots & \vdots \\ (v_n - \mathbb{E}[v_n])(w_1 - \mathbb{E}[w_1]) & (v_n - \mathbb{E}[v_n])(w_2 - \mathbb{E}[w_2]) & \ldots & (v_n - \mathbb{E}[v_n])(w_n - \mathbb{E}[w_n]) \end{array}\right]$$

Mit $\mathbb{E}[(v_i - \mathbb{E}[v_i])(w_j - \mathbb{E}[w_j])] = Cov(v_i, w_j)$:

$$\Sigma(\mathbf{v}, \mathbf{w}) = \begin{pmatrix} Cov(v_1, w_1) & Cov(v_1, w_2) & \dots & Cov(v_1, w_n) \\ Cov(v_2, w_1) & Cov(v_2, w_2) & \dots & Cov(v_2, w_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(v_n, w_1) & Cov(v_n, w_2) & \dots & Cov(v_n, w_n) \end{pmatrix}$$

Gilt $\mathbf{v} = \mathbf{w}$ spricht man von der Varianz-Kovarianz-Matrix

$$\Sigma(\mathbf{v}) = \begin{pmatrix} Var(v_1) & Cov(v_1, v_2) & \dots & Cov(v_1, v_n) \\ Cov(v_2, v_1) & Var(v_2) & \dots & Cov(v_2, v_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(v_n, v_1) & Cov(v_n, v_2) & \dots & Var(v_n) \end{pmatrix}$$

Die Kovarianz-Matrix zweier Vektoren der Ordnung n hat die Ordnung $n \times n$.

1.4.4 Rechenregeln für Kovarianz-Matrizen

Es seien $\mathbf{v}$ und $\mathbf{w}$ zwei stochastische Vektoren der Ordnung n, $\mathbf{A}$ und $\mathbf{B}$ feste Matrizen der Ordnung $m \times n$ und $\mathbf{a}$ und $\mathbf{b}$ ein feste Vektoren der Ordnung m. Dann gilt:

$$\Sigma(\mathbf{A} \cdot \mathbf{v} + \mathbf{a}, \mathbf{B} \cdot \mathbf{w} + \mathbf{b}) = \mathbf{A} \cdot \Sigma(\mathbf{v}, \mathbf{w}) \cdot \mathbf{B}'$$

Gilt $\mathbf{v} = \mathbf{w}$, dann folgt

$$\Sigma(\mathbf{A} \cdot \mathbf{v} + \mathbf{a}) = \mathbf{A} \cdot \Sigma(\mathbf{v}) \cdot \mathbf{A}'$$

1.5 Das lineare Regressionsmodell in Matrixnotation

Es sollen die Werte des Regressands $y_1, \dots, y_n$ durch eine Konstante und durch die Werte der Regressoren $(x_{11}, \dots, x_{n1})$, $(x_{12}, \dots, x_{n2})$, ..., $(x_{1K}, \dots, x_{nK})$ wie folgt erklärt werden:

$$y_i = \beta_0 + x_{i1} \cdot \beta_1 + \ldots + x_{iK} \cdot \beta_k + u_i,\ i = 1, \dots, n \tag{1}$$

Hierbei sind $\beta_0, \dots, \beta_K$ Parameter und $u_1, \dots, u_n$ Störterme.

Regressand $\mathbf{y} \in \mathbb{R}^n$ (beobachtete Zufallsstichprobe)

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Die Werte des Regressanden $\mathbf{y}$ sollen durch ein Modell erklärt werden. Die Werte von $\mathbf{y}$ werden beobachtet und sind also bekannt.

Absolutglied $\boldsymbol{\iota} \in \mathbb{R}^n$

$$\boldsymbol{\iota} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Das Absolutglied bestimmt den Wert des Regressanden, der unabhängig von allen anderen erklärenden Variablen entsteht.

Regressoren $\mathbf{x}_1, \ldots, \mathbf{x}_K \in \mathbb{R}^n$ (beobachtete Zufallsstichprobe)

$$\mathbf{x}_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix} \text{ für } j = 1, \ldots, K$$

Die Werte der Regressoren $\mathbf{x}_1$ bis $\mathbf{x}_K$ fließen in das Modell ein, welches den Regressanden erklärt. Die Werte von $\mathbf{x}_1$ bis $\mathbf{x}_K$ werden beobachtet und sind also bekannt.

Parameter $\boldsymbol{\beta} \in \mathbb{R}^{K+1}$ (unbeobachtet, fest)

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}$$

Die Parameter $\beta_0, \ldots, \beta_K$ fließen <u>linear</u> in das Modell ein, welches den Regressanden erklärt. Diese Parameter können aber nicht beobachtet werden. Sie sind aber für alle Beobachtungen unverändert.

Störterm u $\in \mathbb{R}^n$ (unbeobachtet, stochastisch)

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

Auch die Störterme bestimmen den Regressanden, aber auf unsystematische –unerklärte– Art und Weise. Sie können nicht beobachtet werden und sie können für jede Beobachtung i einen anderen Wert annehmen.

Wir können nun die n durch (1) definierten Gleichungen (etwas) kompakter darstellen:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \beta_0 + \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} \cdot \beta_1 + \ldots + \begin{pmatrix} x_{1K} \\ x_{2K} \\ \vdots \\ x_{nK} \end{pmatrix} \cdot \beta_K + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

beziehungsweise:

$$\mathbf{y} = \boldsymbol{\iota}\beta_0 + \mathbf{x}_1\beta_1 + \ldots + \mathbf{x}_K\beta_K + \mathbf{u}$$

Im nächsten Schritt fassen wir $\boldsymbol{\iota}\beta_0 + \mathbf{x}_1\beta_1 + \ldots + \mathbf{x}_K\beta_K$ zu einem Matrixprodukt zusammen. Hierfür definieren wir zunächst die Regressormatrix.

Regressormatrix X $\in \mathbb{R}^{n\times(K+1)}$

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1K} \\ 1 & x_{21} & \ldots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{nK} \end{pmatrix} = \left(\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} \ldots \begin{pmatrix} x_{1K} \\ x_{2K} \\ \vdots \\ x_{nK} \end{pmatrix} \right) = (\boldsymbol{\iota}, \mathbf{x}_1, \ldots, \mathbf{x}_K)$$

Die Regressormatrix enthält das Absolutglied $\boldsymbol{\iota}$ und alle Regressoren $\mathbf{x}_j$, $j = 1, \ldots, K$. Sie ist von der Ordnung $n \times (K+1)$. Dies bedeutet, dass sie n Zeilen und $K+1$ Spalten hat.

Produkt $\mathbf{X}\boldsymbol{\beta}$

$$
\begin{aligned}
\mathbf{X}\boldsymbol{\beta} &= \begin{pmatrix} 1 & x_{11} & \ldots & x_{1K} \\ 1 & x_{21} & \ldots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{nK} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} \\
&= \begin{pmatrix} 1 \cdot \beta_0 + x_{11} \cdot \beta_1 + \ldots + x_{1K} \cdot \beta_K \\ 1 \cdot \beta_0 + x_{21} \cdot \beta_1 + \ldots + x_{2K} \cdot \beta_K \\ \vdots \\ 1 \cdot \beta_0 + x_{n1} \cdot \beta_1 + \ldots + x_{nK} \cdot \beta_K \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \beta_0 + \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} \beta_1 + \ldots + \begin{pmatrix} x_{1K} \\ x_{2K} \\ \vdots \\ x_{nK} \end{pmatrix} \beta_K \\
&= \boldsymbol{\iota}\beta_0 + \mathbf{x}_1\beta_1 + \ldots + \mathbf{x}_K\beta_K
\end{aligned}
$$

Das Matrixprodukt $\mathbf{X}\boldsymbol{\beta}$ ist also eine Linearkombination der Spalten der Regressormatrix $\mathbf{X}$, wobei die Gewichte der $K+1$ Spalten den $K+1$ Parametern entsprechen.

Wir können nun das lineare Regressionsmodell sehr kompakt wir folgt darstellen:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{2}$$

Diese Darstellung werden wir als Annahme **MLR 1** einfordern. Sie bedeutet, dass im Regressionsmodell der Regressand $\mathbf{y}$ linear in den Parametern $\boldsymbol{\beta}$ ist.

1.6 Lineare Unabhängigkeit und Spaltenrang einer Matrix

Eine Menge von Vektoren gleicher Ordnung ist **linear unabhängig**, wenn keiner der Vektoren durch eine Linearkombination der verbleibenden Vektoren dargestellt werden kann.

Der **Spaltenrang** einer Matrix ist die größte Anzahl linear unabhängiger Spalten dieser Matrix.

Hat eine Matrix $\mathbf{A}$ vollen Spaltenrang (alle Spalten sind linear unabhängig), so gilt:

$$\mathbf{A} \cdot \mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v} = \mathbf{0}$$

1.7 Modellannahmen

MLR 1 Linearität in den Parametern

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

mit $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n\times(K+1)}$, $\boldsymbol{\beta} \in \mathbb{R}^{K+1}$ und $\mathbf{u} \in \mathbb{R}^n$.

Der Parametervektor $\boldsymbol{\beta}$ ist unbekannt und ex ante unspezifiziert.

MLR 2 Zufallsstichprobe

Die Stichprobe $\mathbf{y}$ und $\mathbf{X}$ wird zufällig erhoben.

MLR 3 Voller Spaltenrang der Regressormatrix X

$$rk(\mathbf{X}) = K + 1$$

Alle Spalten von $\mathbf{X}$ sind linear unabhängig.

MLR 4 Strikte Exogenität der Regressoren

$$\mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbb{E}[\mathbf{u}] = \mathbf{0}$$

Die erste Gleichung bedeutet, dass die Realisation der Regressormatrix keinen Einfluss auf die Realisation der Störterm hat. Die zweite Gleichung gilt ohne Beschränkung der Allgemeinheit.

MLR 5 Homoskedastizität und serielle Unkorreliertheit der Störterme

$$\mathbb{V}(\mathbf{u}|\mathbf{X}) = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 \cdot I_n$$

Homoskedastizität: Die auf $\mathbf{X}$ bedingten Varianzen der Störterme $u_i,\ i = 1, \dots, n$ sind alle identisch $(Var(u_i|\mathbf{X}) = \sigma^2)$
Serielle Unkorreliertheit: Die auf $\mathbf{X}$ bedingten Kovarianzen unterschiedlicher Störterme sind gleich null: $(Cov(u_i, u_j|\mathbf{X}) = 0,\ i \neq j)$

MLR 6 Normalverteile homoskedastische und seriell unkorrelierte Störterme

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$$

Diese Annahme impliziert **MLR 4** und **MLR 5**.

1.8 Anmerkungen zu den Modellannahmen

MLR 1

Diese Annahme fordert Linearität in den Parametern $\beta_0, \beta_1, \ldots, \beta_K$ und nicht in den Regressoren. Es dürfen also problemlos nichtlineare Beziehungen wie x_i^2 oder $\ln(x_i)$ untersucht werden.

MLR 2

In der Analyse unterscheiden wir oft zwischen einer ex-ante Betrachtung, in welcher die Regressoren $\mathbf{X}$ noch nicht gezogen wurden und einer ex post Betrachtung, in welcher die Regressoren $\mathbf{X}$ beobachtet wurden und fest gegeben sind. Der unbedingte Erwartungswert, wie zum Beispiel $\mathbb{E}[\hat{\boldsymbol{\beta}}]$, bezieht sich hierbei auf die ex-ante Betrachtung und der bedingte Erwartungswert, wie zum Beispiel $\mathbb{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]$, bezieht sich dann auf die ex post Betrachtung.

MLR 3

Sind die Regressoren $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K$ linear abhängig, so gibt es einen Vektor $\mathbf{v} \in \mathbb{R}^{K+1}$, $\mathbf{v} \neq \mathbf{0}$ mit

$$\sum_{j=0}^{K} v_j \mathbf{x}_j = 0$$

Wir können dann einen Regressor j^* mit $v_{j^*} \neq 0$ durch die anderen Regressoren darstellen:

$$\sum_{j=0}^{K} v_j \mathbf{x}_j = v_{j^*} \mathbf{x}_{j^*} + \sum_{\substack{j=0 \\ j \neq j^*}}^{K} v_j \mathbf{x}_j = \mathbf{0} \Leftrightarrow \mathbf{x}_{j^*} = \frac{1}{v_{j^*}} \sum_{\substack{j=0 \\ j \neq j^*}}^{K} v_j \mathbf{x}_j$$

Das lineare Regressionsmodell benötigt also die in $\mathbf{x}_{j^*}$ gespeicherten Informationen nicht und wir können auf $\mathbf{x}_{j^*}$ verzichten. Die Annahme $rk(\mathbf{X}) = K+1$ hat viele wichtige Implikationen auf die Ergebnisse der Regressionsanalyse, auf die wir später im Einzelnen eingehen werden.

MLR 4

Die Annahme $\mathbb{E}[\mathbf{u}] = \mathbf{0}$ ist ohne Beschränkung der Allgemeinheit. Diese Annahme besteht aus zwei Teilen: Erstens sollen alle Störterme u_i den gleichen Erwartungswert haben und zweitens soll dieser gleich null sein. Nehmen wir zunächst an, dass die Störterme u_i unterschiedliche Erwartungswerte μ_i hätten, also $\mu_i \neq \mu_j$ für mindestens ein Paar i, j mit $i \neq j$. Dann könnten wir den Vektor $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ als zusätzlichen Regressor in das Modell aufnehmen. Das neue Modell lautete dann:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} \cdot \beta_{K+1} + \tilde{\mathbf{u}}$$

wobei wir $\tilde{\mathbf{u}} = \mathbf{u} - \boldsymbol{\mu}$ definieren. Es gilt dann offensichtlich $\mathbb{E}[\tilde{\mathbf{u}}] = \mathbf{0}$. Falls $\mu_i = \mu$ für alle i, dann würde kein zusätzlicher Regressor benötigt, da mit $\tilde{\beta}_0 = \beta_0 + \mu$ der Regressor $\boldsymbol{\iota}$ den Erwartungswert von $\mathbf{u}$ „auffängt“.

1.9 Verletzungen der Modellannahmen

MLR 1:

$$\mathbf{y} = \beta_0\boldsymbol{\iota} + \beta_1\mathbf{x} + \mathbf{z} + \mathbf{u}$$

Hier wird implizit $\beta_2 = 1$ spezifiziert. Laut **MLR 1** müssen aber alle Parameter unspezifiziert sein.

$$\mathbf{y} = \beta_0 \cdot (\boldsymbol{\iota} + \beta_1\mathbf{x}) + \mathbf{u}$$

$$\mathbf{y} = \beta_0 \cdot \mathbf{x}^{\beta_1} \cdot \mathbf{z}^{\beta_2} \cdot \mathbf{u}$$

Diese Modelle sind nicht linear in allen Parametern.

MLR 3:

Es sei $x_i = 1$, falls Beobachtung i eine Katze ist und $x_i = 0$, falls Beobachtung i keine Katze ist. Es sei $z_i = 0$, falls i eine Katze ist und $z_i = 1$ andernfalls.

$$\mathbf{y} = \beta_0 \cdot \boldsymbol{\iota} + \beta_1 \cdot \mathbf{x} + \beta_2 \cdot \mathbf{z} + \mathbf{u}$$

Wegen $\mathbf{x} + \mathbf{z} = \boldsymbol{\iota}$ sind $\boldsymbol{\iota}$, $\mathbf{x}$ und $\mathbf{z}$ linear abhängig.

MLR 4:

Der Stundenlohn y_i sei abhängig von der Anzahl der Bildungsjahre x_i und von der Anzahl der Jahre Berufserfahrung z_i. Die Anzahl der Jahre Berufserfahrung konnte aber nicht erfasst werden und es wird das Modell

$$y_i = \beta_0 + \beta_1 x_i + u_i \ , \ i = 1, \ldots, n$$

geschätzt. Hierbei ist z_i implizit ein Teil von u_i. Da aber die Anzahl der Jahre Berufserfahrung z_i mit der Anzahl der Bildungsjahre x_i korreliert ist, ist auch x_i mit u_i korreliert. Dann ist aber MLR 4 verletzt.

MLR 5:

Es sei $\mathbf{y}$ die Konsumausgaben und $\mathbf{x}$ das Haushaltseinkommen. Offensichtlich streuen die Konsumausgaben stärker bei hohen Haushaltseinkommen. Mit $\Sigma(\mathbf{y}|\mathbf{x}) = \Sigma(\beta_0 \cdot \boldsymbol{\iota} + \beta_1 \cdot \mathbf{x} + \mathbf{u}|\mathbf{x}) = \Sigma(\mathbf{u}|\mathbf{x})$ muss daher auch die auf $\mathbf{x}$ bedingte Varianz der Störterme $\mathbf{u}$ von $\mathbf{x}$ abhängen.

Der Störterm $\mathbf{u}$ erfülle MLR 5. Es sei $\mathbf{y}_t$ das Bruttoinlandsprodukt von Jahr t. Die Regressionsgleichuung laute mit $\beta_2 \neq 0$.

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + u_t \ ,$$

das aktuelle BIP hängt demnach auch vom BIP letzten Jahres ab. Wird y_{t-1} in die Regressionsgleichung eingesetzt, ergibt sich

$$y_t = \beta_0 + \beta_1 x_t + \beta_2(\beta_0 + \beta_1 x_{t-1} + \beta_2 y_{t-2} + u_{t-1}) + u_t \ ,$$

Mit $\tilde{\beta}_0 = \beta_0(1 + \beta_2)$, $\tilde{\beta}_1 = \beta_1$, $\tilde{\beta}_2 = \beta_1\beta_2$, $\tilde{\beta}_3 = \beta_2\beta_2$ und $\tilde{u}_t = \beta_2 u_{t-1} + u_t$ können wir das Modell wie folgt aufschreiben:

$$y_t = \tilde{\beta}_0 + \tilde{\beta}_1 x_t + \tilde{\beta}_2 x_{t-1} + \tilde{\beta}_3 y_{t-2} + \tilde{u}_t$$

Wegen $Cov(\tilde{u}_t, \tilde{u}_{t-1}) = Cov(\beta_2 u_{t-1} + u_t, \beta_2 u_{t-2} + u_{t-1}) = \beta_2 Var(u_{t-1}) \neq 0$ ist nun MLR 5 für $\tilde{u}_t$ verletzt.

MLR 6:

Im Störterm $\mathbf{u}$ seien ganzzahlige und nichtnegative unbeobachtete Faktoren enthalten, die den Regressanden $\mathbf{u}$ erklären, wie Monate Arbeitslosigkeit, Anzahl Kinder im Haushalt, etc. Da die Ausprägungen von normalverteilten Zufallsvariablen aber beliebige reelle Zahlen sein können, kann $\mathbf{u}$ nicht normalverteilt sein.

2 Schätzer für $\boldsymbol{\beta}$

2.1 Lineare Funktionen

$f : \mathbb{R}^n \to \mathbb{R}^m$ ist eine **lineare Funktion**, falls

$$f(\mathbf{x}_1 + \mathbf{x}_2) = f(\mathbf{x}_1) + f(\mathbf{x}_2) \text{ für alle } \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$$

und

$$f(t \cdot \mathbf{x}) = t \cdot f(\mathbf{x}) \text{ für alle } \mathbf{x} \in \mathbb{R} \text{ und } t \in \mathbb{R}$$

Sei $\mathbf{A}$ eine Matrix der Ordnung $m \times n$ und $\mathbf{x}$ ein Vektor der Ordnung n. Dann ist die Funktion $f(\mathbf{x}) = \mathbf{A} \cdot \mathbf{x}$ eine lineare Funktion.

2.1.1 Jacobi-Matrix linearer Funktionen

Die Jacobi Matrix einer Funktion besteht aus den ersten partiellen Ableitungen einer Funktion.

Ist eine Funktion $f : \mathbb{R}^m \to \mathbb{R}^n$ durch

$$f(\mathbf{x}) = \mathbf{A} \cdot \mathbf{x}$$

definiert, so ist $\mathbf{A}$ die Jacobi-Matrix von f:

$$\mathbf{J}_f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_m} \end{pmatrix} = \mathbf{A}'$$

2.2 Quadratische Formen und positiv Definitheit

Es sei $\mathbf{A}$ eine quadratische Matrix der Ordnung $n \times n$ und $\mathbf{x}$ ein Vektor der Ordnung n. Dann ist die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ mit

$$f(\mathbf{x}) = \mathbf{x}' \cdot \mathbf{A} \cdot \mathbf{x}$$

eine **quadratische Form**.

Gilt

$$\mathbf{x}' \cdot \mathbf{A} \cdot \mathbf{x} \geq 0$$

für alle $\mathbf{x} \in \mathbb{R}^n$, so heißt die Matrix **A positiv semi definit**. Ist die Ungleichung strikt für alle $\mathbf{x} \neq \mathbf{0}$, so heißt die Matrix **A positiv definit**.

Der Gradient von f an der Stelle $\mathbf{x}$ ist gegeben durch

$$\nabla_f(\mathbf{x}) = (\mathbf{A} + \mathbf{A}') \cdot \mathbf{x}$$

und die Hessematrix von f ist für alle Stellen $\mathbf{x}$ gegeben durch:

$$H_f = \mathbf{A}' + \mathbf{A}$$

Die Hessematrix H_f ist symmetrisch.

Ist $\mathbf{A}$ positiv definit, so H_f positiv definit symmetrisch (pds) und f ist streng konvex.

2.3 Löwner kleiner

Gegeben seien zwei $n \times n$ Matrizen $\mathbf{A}$ und $\mathbf{B}$. Die Matrix $\mathbf{A}$ ist **Löwner kleiner** als die Matrix $\mathbf{B}$, $\mathbf{A} \leq \mathbf{B}$, falls

$$\mathbf{x}' \cdot \mathbf{A} \cdot \mathbf{x} \leq \mathbf{x}' \cdot \mathbf{B} \cdot \mathbf{x} \text{ für alle } \mathbf{x} \in \mathbb{R}^n$$

Löwner kleiner $\mathbf{A} \leq \mathbf{B}$ ist äquivalent zu $\mathbf{B} - \mathbf{A}$ positiv semidefinit.

2.4 Inverse einer Matrix

Sei $\mathbf{A}$ eine Matrix der Ordnung $m \times n$. Eine Matrix $\mathbf{B}$ der Ordnung $n \times m$ mit der Eigenschaft $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$ heißt **Pseudoinverse** von $\mathbf{A}$.

Sei $\mathbf{A}$ eine Matrix der Ordnung $n \times n$. Eine Matrix $\mathbf{B}$ der Ordnung $n \times n$ mit der Eigenschaft $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$ und $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$ heißt **Inverse** von $\mathbf{A}$. Wir bezeichnen die Inverse von $\mathbf{A}$ mit $\mathbf{A}^{-1}$.

Nur quadratische Matrizen können eine Inverse haben und nicht jede quadratische Matrix hat eine Inverse. Existiert eine Inverse, so ist sie eindeutig.

Ist eine Matrix positiv definit, so ist dies hinreichend für die Existenz einer Inversen.

Für eine inverse Matrix gilt:

$$\left(\mathbf{A}^{-1}\right)' = (\mathbf{A}')^{-1}$$

2.5 Schätzer

Ein **Schätzer** ist eine Funktion f, welche jede mögliche Realisation einer Zufallsstichprobe $(\mathbf{y}, \mathbf{X})$ in die Menge der möglichen Parameter abbildet. Da diese nicht ausschließlich den Vektor $\boldsymbol{\beta} \in \mathbb{R}^{K+1}$ betrifft, sondern zum Beispiel auch Varianzen und Kovarianzen der Störterme, bezeichnen wir die Menge der möglichen Parameter hier allgemein mit Θ und einen bestimmten Vektor von Parametern mit $\boldsymbol{\theta}$.

Folgende Eigenschaften spielen für die lineare Regressionsanalyse eine wichtige Rolle:

Ein Schätzer f ist **linear** in $\mathbf{y}$, falls eine Matrix $\underset{m\times n}{\mathbf{A}}$ existiert, sodass:

$$f(\mathbf{y}, \mathbf{X}) = \mathbf{A} \cdot \mathbf{y}$$

Ein Schätzer f ist **erwartungstreu** für einen unbekannten Parameter $\boldsymbol{\theta}$, falls

$$\mathbb{E}[f(\mathbf{y}, \mathbf{X})] = \boldsymbol{\theta} \text{ für alle } \boldsymbol{\theta} \in \Theta$$

Ein Schätzer f ist **effizient**, falls f erwartungstreu ist und die Varianz-Kovarianz-Matrix von f Löwner kleiner ist als die Varianz-Kovarianz-Matrix jedes anderen erwartungstreuen Schätzers.

2.6 Lineare und erwartungstreue Schätzer für $\boldsymbol{\beta}$

Sei f eine in $\mathbf{y}$ lineare Schätzfunktion für $\boldsymbol{\beta}$:

$$f(\mathbf{y}) = \mathbf{A}\mathbf{y} \ ,$$

wobei die Matrix $\mathbf{A}$ für eine gegebene Regressormatrix $\mathbf{X}$ nicht stochastisch ist.

Um die statistischen Eigenschaften eines Schätzers zu bestimmen ist es hilfreich, diesen mit MLR 1 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ in folgende Form zu bringen:

$$f(\mathbf{y}) = \mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{X}\boldsymbol{\beta} + \mathbf{A}\mathbf{u}$$

Der auf $\mathbf{X}$ bedingte Erwartungswert dieser Schätzfunktion lautet

$$\mathbb{E}[\mathbf{A}\mathbf{y}|\mathbf{X}] = \mathbb{E}[\mathbf{A}\mathbf{X}\boldsymbol{\beta} + \mathbf{A}\mathbf{u}|\mathbf{X}] = \mathbf{A}\mathbf{X}\boldsymbol{\beta} + \mathbf{A}\mathbb{E}[\mathbf{u}|\mathbf{X}]$$

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

$$\mathbb{E}[\mathbf{Ay}|\mathbf{X}] = \mathbf{AX}\boldsymbol{\beta}$$

Es gilt dann:

$$\mathbb{E}[\mathbf{Ay}] = \mathbb{E}[\mathbb{E}[\mathbf{Ay}|\mathbf{X}]] = \mathbb{E}[\mathbf{AX}]\boldsymbol{\beta}$$

Der Ausdruck $\mathbb{E}[\mathbf{AX}]\boldsymbol{\beta}$ ist gleich $\boldsymbol{\beta}$ für alle $\boldsymbol{\beta} \in \mathbb{R}^{K+1}$, falls $\mathbf{AX} = \mathbf{I}_{K+1}$. Es ist also hinreichend für die Erwartungstreue von f, wenn $\mathbf{A}$ eine Pseudoinverse von $\mathbf{X}$ ist.

Die auf $\mathbf{X}$ bedingte Kovarianzmatrix eines linearen Schätzers lautet:

$$\Sigma(\mathbf{Ay}|\mathbf{X}) = \Sigma(\mathbf{AX}\boldsymbol{\beta} + \mathbf{Au}|\mathbf{X}) = \mathbf{A}\Sigma(\mathbf{u}|\mathbf{X})\mathbf{A}'$$

Mit MLR 5 $\Sigma(\mathbf{u}|\mathbf{X}) = \sigma^2 \cdot \mathbf{I}_n$ folgt:

$$\Sigma(\mathbf{Ay}|\mathbf{X}) = \mathbf{A} \cdot \sigma^2 \cdot \mathbf{I}_n\mathbf{A}' = \sigma^2 \cdot \mathbf{AA}'$$

Ein effizienter linearer Schätzer $\hat{\boldsymbol{\beta}} = \mathbf{Ay}$ für $\boldsymbol{\beta}$ erfüllt also:

$$\mathbf{AX} = \mathbf{I}_{K+1} \text{ und } \mathbf{AA}' \leq \mathbf{BB}' \text{ für alle } \mathbf{B} \in \mathbb{R}^{K+1\times n} \text{ mit } \mathbf{BX} = \mathbf{I}_{K+1}$$

3 Kleinste Quadrate Schätzung

3.1 Prognose und Residuum

Sei $\mathbf{b} \in \mathbb{R}^{K+1}$ eine Vermutung über für den unbekanten Parametervektor $\boldsymbol{\beta} \in \mathbb{R}^{K+1}$. Dann ist

$$\hat{\mathbf{y}} = \mathbf{Xb}$$

Die **Prognose** für $\mathbf{y}$ bei gegebener Regressormatrix $\mathbf{X}$ und

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{Xb}$$

heißt **Prognosefehler** oder **Residuum**.

3.2 Summe der quadrierten Residuen

Für eine gegebene Stichprobe $(\mathbf{y}, \mathbf{X})$ sei die Summe der quadrierten Residuen durch $S : \mathbb{R}^{K+1} \to \mathbb{R}$ definiert:

$$\begin{aligned} S(\mathbf{b}|\mathbf{y}, \mathbf{X}) &= \sum_{i=1}^{n} \hat{u}_i^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= \mathbf{y}'\mathbf{y} - (\mathbf{X}\mathbf{b})'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} + (\mathbf{X}\mathbf{b})'\mathbf{X}\mathbf{b} \\ &= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \end{aligned}$$

Da $\mathbf{b}'\mathbf{X}'\mathbf{y}$ ein Skalar ist, gilt

$$\mathbf{b}'\mathbf{X}'\mathbf{y} = (\mathbf{b}'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\mathbf{b}$$

und damit

$$S(\mathbf{b}|\mathbf{y}, \mathbf{X}) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$$

3.3 Multivariate Optimierung

Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine stetige und differenzierbare Funktion.

Die Stelle $\mathbf{x}_0 \in \mathbb{R}^n$ heißt **stationär**, falls $\nabla_f(\mathbf{x}_0) = \mathbf{0}$.

Ist f streng konvex und ist $\mathbf{x}_0$ stationär, so ist $\mathbf{x}_0$ die einzige Minimumstelle von f, d.h. $f(\mathbf{x}_0) \leq f(\mathbf{x})$ für alle $\mathbf{x} \in \mathbb{R}^n$.

3.4 Kleinste Quadrate

Zu minimieren ist die Summe der quadrierten Residuen:

$$S(\mathbf{b}|\mathbf{y}, \mathbf{X}) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$$

Die Funktion $S(\mathbf{b}|\mathbf{y}, \mathbf{X})$ ist stetig und differenzierbar.

Der Gradient von S lautet:

$$\nabla_S(\mathbf{b}) = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}$$

Für die Lösung des Problems $\min\limits_{\mathbf{b} \in \mathbb{R}^{K+1}} S(\mathbf{b}|\mathbf{y}, \mathbf{X})$ muss die notwenige Bedingung erster Ordnung gelten, d.h. die Minumstelle $\mathbf{b}_0$ muss stationär sein: $\nabla_S(\mathbf{b}) = \mathbf{0}$. Dies ist äquivalent zu:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y} \tag{3}$$

Das obige System von $K+1$ Gleichungen in $K+1$ Unbekannten $b_0, \ldots, b_K$ definiert die notwendigen Bedingungen für eine Minimumstelle.

Wir werden Gleichung (3) später lösen. Auf der linken Seite der Gleichung sehen wir das Matrixprodukt $\mathbf{X}'\mathbf{X}$. Wir nennen es auch die „Momentematrix“. Diese wird im Laufe der Vorlesung noch oft eine wichtige Rolle spielen, daher untersuchen wir zunächst ihre Eigenschaften.

Ausgeschrieben lautet die Momentematrix:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \ldots & \sum_{i=1}^n x_{iK} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}x_{i1} & \ldots & \sum_{i=1}^n x_{i1}x_{iK} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{iK} & \sum_{i=1}^n x_{iK}x_{i1} & \ldots & \sum_{i=1}^n x_{iK}x_{iK} \end{pmatrix}$$

Teilen wir alle Einträge der Momentematrix durch n (und prämultiplizieren die Matrix mit n), können wir die Momentematrix kürzer aufschreiben:

$$n \cdot \frac{1}{n} \cdot \mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & \overline{x_1} & \ldots & \overline{x_K} \\ \overline{x_1} & \overline{x_1 x_1} & \ldots & \overline{x_1 x_K} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{x_K} & \overline{x_K x_1} & \ldots & \overline{x_K x_K} \end{pmatrix}$$

In dieser Matrix sind nun die Informationen gespeichert, die zur Berechnung der ersten und zweiten empirischen Momente der Regressoren benötigt werden.

Das Matrixproduk $\mathbf{X}'\mathbf{X}$ ist symmetrisch:

$$(\mathbf{X}'\mathbf{X})' = \mathbf{X}' (\mathbf{X}')' = \mathbf{X}'\mathbf{X}$$

Die Momentematrix $\mathbf{X}'\mathbf{X}$ ist positiv definit.
Hierzu bilden wir mit einem $K+1$-Vektor $\mathbf{v}$ und der Momentematrix eine quadratische Form:

$$\mathbf{v}'\mathbf{X}'\mathbf{X}\mathbf{v} = (\mathbf{X}\mathbf{v})' \mathbf{X}\mathbf{v}$$

Das Produkt $\underset{n\times K+1}{\mathbf{X}} \underset{K+1\times 1}{\mathbf{v}}$ ist ein n-Vektor, den wir $\mathbf{z}$ nennen. Damit gilt:

$$\mathbf{v}'\mathbf{X}'\mathbf{X}\mathbf{v} = (\mathbf{X}\mathbf{v})' \mathbf{X}\mathbf{v} = \mathbf{z}'\mathbf{z} = \sum_{i=1}^n z_i^2 \geq 0 \ ,$$

womit wir gezeigt haben, dass die Momentematrix positiv semidefinit ist. Für positiv Definitheit darf die Summe der Quadrate der z_i aber nur dann gleich null sein (in diesem Fall müssen alle z_i gleich null sein), falls $\mathbf{v}$ der Nullvektor ist. Betrachten wir also einen beliegigen $K+1$-Vektor $\mathbf{v}$, welcher nicht nur aus nullen besteht. Da mit Annahme MLR 3 die Regressormatrix $\mathbf{X}$ vollen Spaltenrang hat, also alle Spalten von $\mathbf{X}$ linear unabhängig sind, muss $\mathbf{Xv} \neq \mathbf{0}$ gelten. Damit ist $\mathbf{z} = \mathbf{Xv} \neq \mathbf{0}$ für alle $\mathbf{v} \neq \mathbf{0}$ und es gilt:

$$\mathbf{v}'\mathbf{X}'\mathbf{X}\mathbf{v} = \mathbf{z}'\mathbf{z} > 0 \text{ für alle } \mathbf{v} \neq \mathbf{0}$$

Damit ist die Momentematrix $\mathbf{X}'\mathbf{X}$ positiv definit.

Nun betrachten wir die hinreichende Bedingung für ein Minimum: Falls S streng konvex ist, dann ist eine stationäre Stelle die eindeutige Minimumstelle. S ist streng konvex, falls die Hessematrix von S positiv definit ist. Die Hessematrix von S lautet:

$$H_S = (\mathbf{X}'\mathbf{X})' + \mathbf{X}'\mathbf{X} = 2\mathbf{X}'\mathbf{X}$$

Da $\mathbf{X}'\mathbf{X}$ positiv definit ist, ist auch die Hessematrix von S positiv definit. Die notwendigen Bedingungen (3) definieren also das eindeutige Minimum von S. Deren Lösung nennen wir die kleinsten Quadrate Schätzer.

3.5 Momente Methode

Ber der Momente Methode („method of moments“) wird eine Annahme an die theoretischen Momente von Zufallsvariablen getroffen und dann Parameter des Modells so angepasst, dass die empirischen Momente den theoretischen Momenten entsprechen.

Die Zufallsvariablen die hier in der Momente Methode Anwendung finden, sind die Zufallsvariablen hinter $\mathbf{y}$ und $\mathbf{X}$, welche gemäß MLR 2 einer Zufallsstichprobe entstehen. Die theoretischen Momente sind durch $\mathbb{E}[\mathbf{X}'\mathbf{y}]$ und $\mathbb{E}[\mathbf{X}'\mathbf{X}]$ gegeben.

Mit der Gleichung (2) des linearen Regressionsmodells gilt:

$$\mathbb{E}[\mathbf{X}'\mathbf{y}] = \mathbb{E}[\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u})] = \mathbb{E}[\mathbf{X}'\mathbf{X}]\boldsymbol{\beta} + \mathbb{E}[\mathbf{X}'\mathbf{u}]$$

Aus der Annahme MLR 4 (strikt exogene Regressoren) $\mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbb{E}[\mathbf{u}] = \mathbf{0}$ folgt:

$$\mathbb{E}[\mathbf{X}'\mathbf{u}] = \mathbb{E}[\mathbf{X}'\mathbb{E}[\mathbf{u}|\mathbf{X}]] = \mathbb{E}[\mathbf{X}'\mathbb{E}[\mathbf{u}]] = \mathbb{E}[\mathbf{X}'\mathbf{0}] = \mathbf{N}$$

Damit gilt auch

$$\mathbb{E}\left[\mathbf{X}'\mathbf{y}\right] = \mathbb{E}\left[\mathbf{X}'\mathbf{X}\right]\boldsymbol{\beta}$$

Die empirischen Pendants zu den theoretischen Momenten $\mathbb{E}\left[\mathbf{X}'\mathbf{y}\right]$ und $\mathbb{E}\left[\mathbf{X}'\mathbf{X}\right]$ lauten $\mathbf{X}'\mathbf{y}$ und $\mathbf{X}'\mathbf{X}$. Daher müssen nach der Momente Methode der vermutete Vektor $\mathbf{b}$ für den unbekanten Parametervektor $\boldsymbol{\beta}$ so angepasst werden, dass

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\mathbf{b}$$

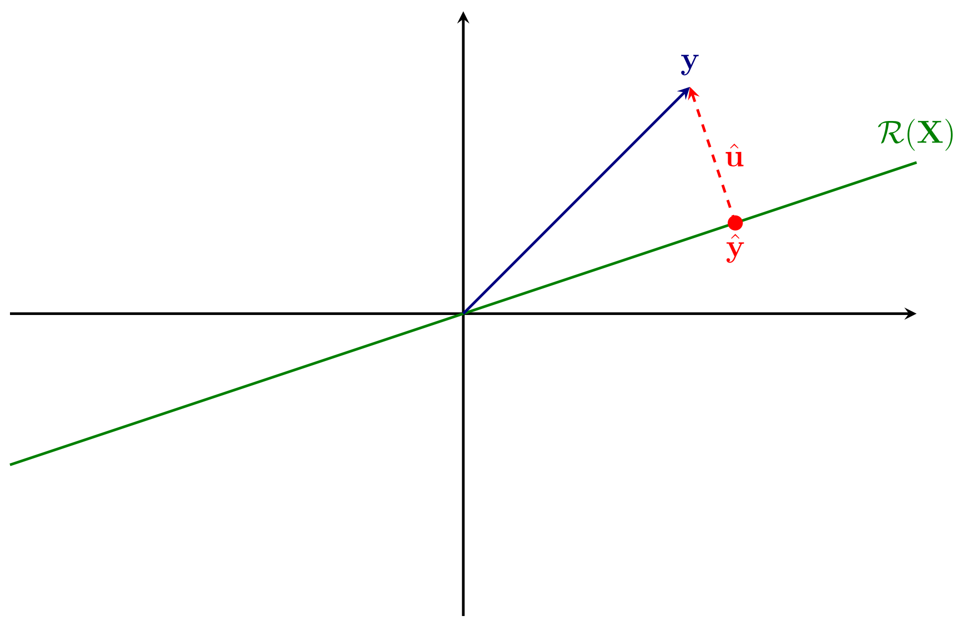
gilt. Diese Gleichung ist identisch zu (3).

3.6 Orthogonale Projektion

Jede Vermutung $\mathbf{b}$ über die unbekannten Parameter $\boldsymbol{\beta}$ generieret eine Prognose $\hat{\mathbf{y}}$ des Regressanden bei gegebener Regressormatrix $\mathbf{X}$:

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$$

Da $\hat{\mathbf{y}}$ eine Linearkombination der Spalten von $\mathbf{X}$ ist, befindet sich $\hat{\mathbf{y}}$ in der durch die Spalten von $\mathbf{X}$ aufgespannten Hyperebene, dem sogenannten Spaltenraum von $\mathbf{X}$, $\mathcal{R}(\mathbf{X})$. Dieser Spaltenraum von $\mathbf{X}$ hat aufgrund der linearen Unabhängigkeit aller $K+1$ Spalten von $\mathbf{X}$ die Dimension $K+1$. Der tatsächlich beobachtete Regressand $\mathbf{y}$ hat die Dimension n und liegt allgemein nicht in $\mathcal{R}(\mathbf{X})$. Je näher der Prognosevektor $\hat{\mathbf{y}}$ am Beobachtungsvektor $\mathbf{y}$ liegt, desto höher ist der Erklärungsgehalt der Vermutung $\mathbf{b}$. Der Erklärungsgehalt von $\mathbf{b}$ ist daher am größten, wenn der Differenzenvektor von $\mathbf{y}$ zu $\hat{\mathbf{y}}$, also das Residuum $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}$, ausgehend von $\mathbf{y}$ senkrecht auf den Spaltenraum von $\mathbf{X}$, $\mathcal{R}(\mathbf{X})$ zeigt. Somit wird der Vektor $\mathbf{y}$ orthogonal auf den Spaltenraum $\mathcal{R}(\mathbf{X})$ projiziert:



Damit muss gelten, dass das Residuum $\hat{\mathbf{u}}$ senkrecht zu allen Spalten der Regressormatrix ist:

$$\mathbf{x}_j'\hat{\mathbf{u}} = 0 \text{ für alle } j = 0, \ldots, K \Leftrightarrow \mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$$

Mit $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{b}$ folgt wieder Gleichung (3):

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\mathbf{b}$$

3.7 Lösung von Gleichung (3)

Die notwendige Bedingung für eine Minimumstelle der Summe der quadrierten Residuen (3) lautet:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$$

Da die Momentematrix $\mathbf{X}'\mathbf{X}$ positiv definit ist, existiert die Inverse $(\mathbf{X}'\mathbf{X})^{-1}$. Wir multiplizieren diese Inverse von links auf beiden Seiten der Gleichung und erhalten die eindeutige Lösung, welche wir mit $\hat{\boldsymbol{\beta}}$ bezeichnen:

$$\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}}_{=\mathbf{I}_{K+1}}\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$\Rightarrow$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{4}$$

Wir bezeichnen $\hat{\boldsymbol{\beta}}$ als **kleinsten Quadrate Schätzer für $\boldsymbol{\beta}$**. Er ist gleichzeitig der Momente Methode Schätzer für $\boldsymbol{\beta}$ und der Schätzer, welcher durch die orthogonale Projektion von $\mathbf{y}$ auf $\mathcal{R}(\mathbf{X})$ berechnet wird.

3.8 Gauß-Markov-Theorem

Unter den Annahmen **MLR 1**, **2**, **3**, **4** und **5** ist der Schätzer $\hat{\boldsymbol{\beta}}$ innerhalb der linearen Schätzer effizient für $\boldsymbol{\beta}$. Auf Englisch wird effizient mit „best" bezeichnet und das Gauß-Markov-Theorem wird oft durch das Akronym „BLUE" wiedergegeben: **B**est **L**inear **U**nbiased **E**stimator.

Wir gehen in der Begründung dieser Aussage nun schrittweise vor:

E: Es handelt sich hier um einen Schätzer (estimator) für $\boldsymbol{\beta}$ ✓

L: Mit $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ ist die Matrix $\mathbf{A}$, welche die lineare Funktion $\mathbf{A}\mathbf{y}$ definiert durch

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

gegeben. $\mathbf{A}$ hat die Ordnung $K+1 \times n$. Damit ist $\hat{\boldsymbol{\beta}}$ linear in $\mathbf{y}$.

Anmerkung:
Die Inverse $(\mathbf{X}'\mathbf{X})^{-1}$ existiert wegen **MLR 3**.

U: Wir hatten zuvor die hinreichende Bedingung $\mathbf{A}\mathbf{X} = \mathbf{I}_{K+1}$ für Unverzerrtheit hergeleitet. Mit $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ gilt:

$$\mathbf{A}\mathbf{X} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}_{K+1} \checkmark$$

Anmerkung:
Hierfür werden die Annahmen **MLR 1**, **3** und **4** benötigt.

B: Zu zeigen ist für $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ und für alle $\underset{n\times K+1}{\mathbf{B}}$ mit $\mathbf{B}\mathbf{X} = \mathbf{I}$:

$$\mathbb{V}(\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{A}\mathbf{A}' \leq \sigma^2\mathbf{B}\mathbf{B}' \Leftrightarrow \mathbf{x}'(\mathbf{B}\mathbf{B}' - \mathbf{A}\mathbf{A}')\,\mathbf{x} \geq 0 \text{ für alle } \mathbf{x} \in \mathbb{R}^{K+1}$$

Hierzu stellen wir zunächst fest, dass wegen $\mathbf{B}\mathbf{X} = \mathbf{I}$ gilt:

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

Damit gilt:

$$(\mathbf{B} - \mathbf{A})(\mathbf{B} - \mathbf{A})' = \mathbf{B}\mathbf{B}' - \mathbf{A}\mathbf{B}' - \mathbf{B}\mathbf{A}' + \mathbf{A}\mathbf{A}' = \mathbf{B}\mathbf{B}' - \mathbf{A}\mathbf{A}'$$

und

$$\mathbf{x}'\left(\mathbf{B}\mathbf{B}'-\mathbf{A}\mathbf{A}'\right)\mathbf{x}=\underbrace{\mathbf{x}'\left(\mathbf{B}-\mathbf{A}\right)}_{=:\mathbf{z}'}\cdot\underbrace{\left(\mathbf{B}-\mathbf{A}\right)'\mathbf{x}}_{=:\mathbf{z}}=\mathbf{z}'\mathbf{z}\geq 0 \text{ für alle } \mathbf{x}\in\mathbb{R}^{K+1} \checkmark$$

4 Schätzer für σ^2

4.1 Idempotent symmetrische Matrizen

Wie in Abschnitt 3.6 bereits dargestellt wird mit der orthogonalen Projektion auf den linearen Unterraum $\mathcal{R}(\mathbf{X})$ des $\mathbb{R}^n$ einem beliebigen Punkt $\mathbf{y}\in\mathbb{R}^n$ eine Projektion $\hat{\mathbf{y}}\in\mathcal{R}(\mathbf{X})$ zugeordet, welche die Eigenschaft hat, dass die Differenz $\mathbf{y}-\hat{\mathbf{y}}=\hat{\mathbf{u}}$ orthogonal zum Raum $\mathcal{R}(\mathbf{X})$ und damit auch zu allen Spalten von $\mathbf{X}$ ist.

Wir haben nun hergeleitet, dass die Projektion von $\mathbf{y}$ auf $\hat{\mathbf{y}}$ durch die lineare Abbildung $\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$ erreicht wird. Wir nennen die Matrix $\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ eine **Projektionsmatrix**.

Da $\hat{\mathbf{y}}$ bereits im Spaltenraum von $\mathbf{X}$ liegt, also $\hat{\mathbf{y}}\in\mathcal{R}(\mathbf{X})$, würde eine Projektion von $\hat{\mathbf{y}}$ auf $\mathcal{R}(\mathbf{X})$ nicht verändern:

$$\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{y}}=\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}=\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}=\hat{\mathbf{y}}$$

Diese Eigenschaft definieren wir wir folgt:

Eine $n\times n$-Matrix $\mathbf{A}$ heißt **idempotent**, falls gilt:

$$\mathbf{A}\mathbf{A}=\mathbf{A}$$

Orthogonale Projektionsmatrizen müssen idempotent und symmetrisch sein.

Ist $\mathbf{A}$ eine orthogonale Projektionsmatrix, dann ist auch $\mathbf{I}-\mathbf{A}$ eine orthogonale Projektionsmatrix. Konkret projiziert die Matrix

$$\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$$

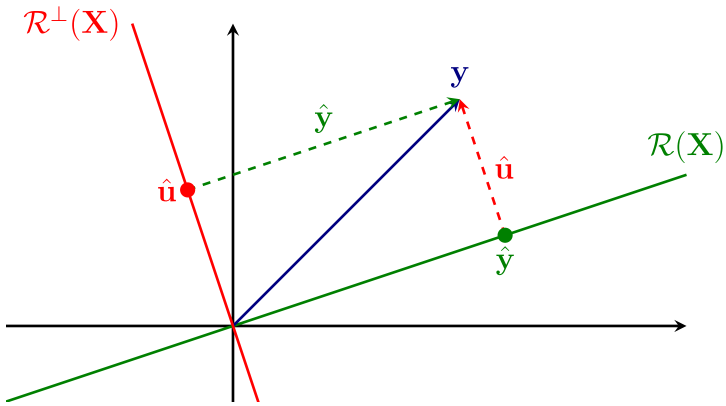
auf das orthogonale Komplement von $\mathcal{R}(\mathbf{X})$, welches wir mit $\mathcal{R}^{\perp}(\mathbf{X})$ bezeichnen. Auch $\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ ist idempotent symmetrisch.

Wir benötigen $\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$, um die Residuen $\hat{\mathbf{u}}$ zu berechnen:

$$\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\mathbf{y}=\mathbf{y}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{y}-\hat{\mathbf{y}}=\hat{\mathbf{u}}$$

Die Residuen sind also Elemente des orthogonalen Komplements von $\mathcal{R}(\mathbf{X})$, also $\hat{\mathbf{u}} \in \mathcal{R}^{\perp}(\mathbf{X})$ und sind damit orthogonal zu allen Spalten von $\mathbf{X}$:

$$\mathbf{X}\hat{\mathbf{u}} = \mathbf{0}$$



$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}}$$

mit

$$\hat{\mathbf{y}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \text{ und } \hat{\mathbf{u}} = \left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\mathbf{y} \ .$$

Ab nun benennen wir die beiden Projektionsmatrizen wie folgt:

$$\mathbf{P_X} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' \text{ und } \mathbf{P_X^{\perp}} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$$

Mit $\hat{\mathbf{u}} = \mathbf{P_X^{\perp}}\mathbf{y}$ und $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ folgt:

$$\hat{\mathbf{u}} = \mathbf{P_X^{\perp}}\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}\right) = \mathbf{P_X^{\perp}}\mathbf{X}\boldsymbol{\beta} + \mathbf{P_X^{\perp}}\mathbf{u}$$

Da $\mathbf{P_X^{\perp}}$ auf das orthogonale Komplement des Spaltenraums von $\mathbf{X}$ projiziert, und $\mathbf{X}$ offensichtlich im Spaltenraum von $\mathbf{X}$ liegt, gilt $\mathbf{P_X^{\perp}}\mathbf{X} = \mathbf{N}$. Also folgt:

$$\hat{\mathbf{u}} = \mathbf{P_X^{\perp}}\mathbf{u}$$

Da $\mathbf{P_X}$ auf den Spaltenraum von $\mathbf{X}$ projiziert und nach Annahme **MLR 3** $rk(\mathbf{X}) = K + 1$ die Matrix $\mathbf{X}$ vollen Spaltenrang hat, gilt auch $rk(\mathbf{P_X}) = K + 1$.

Da $\mathbf{P_X^{\perp}}$ auf das orthogonale Komplement des Spaltenraums von $\mathbf{X}$ projiziert gilt entsprechend $rk(\mathbf{P_X^{\perp}}) = n - (K + 1)$.

4.2 Streuungszerlegung

Wir können die Variation des Regressanden $\mathbf{y}$ zerlegen in einen durch die Regressoren $\mathbf{X}$ erklärten Teil und in einen unerklärten Teil:

SST **t**otal **s**um of **s**quares

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

SSE **e**xplained **s**um of **s**quares

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

SSR **s**um of **s**quared **r**esiduals

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}$$

Da $\boldsymbol{\iota}$ die erste Spalte von $\mathbf{X}$ ist und $\hat{\mathbf{u}} \in \mathcal{R}(\mathbf{X})^\perp$, gilt auch $\boldsymbol{\iota} \perp \hat{\mathbf{u}}$, also $\boldsymbol{\iota}'\hat{\mathbf{u}} = 0$. Mit $\bar{\hat{u}} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = \frac{1}{n}\boldsymbol{\iota}'\hat{\mathbf{u}}$ gilt also $\bar{\hat{\mathbf{u}}} = 0$.

Mit $\hat{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{u}}$ gilt $\bar{\hat{\mathbf{y}}} = \bar{\mathbf{y}} - \bar{\hat{\mathbf{u}}} = \bar{\mathbf{y}}$.

Somit können wir schreiben:

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i + \hat{u}_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n} \hat{y}_i\hat{u}_i - 2\bar{y}\sum_{i=1}^{n} \hat{u}_i + \sum_{i=1}^{n} \hat{u}_i^2$$

Da $\hat{\mathbf{y}} \in \mathcal{R}(\mathbf{X})$ und $\hat{\mathbf{u}} \in \mathcal{R}(\mathbf{X})^\perp$ gilt $\hat{\mathbf{y}} \perp \hat{\mathbf{u}}$, also $\hat{\mathbf{y}}'\hat{\mathbf{u}} = \sum_{i=1}^{n} \hat{y}_i\hat{u}_i = 0$. Zudem gilt $\sum_{i=1}^{n} \hat{u}_i = \boldsymbol{\iota}'\hat{\mathbf{u}} = 0$.

Damit folgt die Streuungszerlegung

$$SST = SSE + SSR$$

4.3 Bestimmtheitsmaß R^2

Um die Güte einer Regression zu beurteilen, berechnen wir den Anteil der erklärten quadratischen Abweichung des Regressanden an der gesamten quadratischen Abweichung des Regressanden:

$$R^2 = \frac{SSE}{SST}$$

Da $SSE \leq SST$ gilt (ansonsten müsste $SSR < 0$ sein) und wegen $SSE \geq 0$ folgt $0 \leq R^2 \leq 1$. Deswegen können wir R^2 als prozentualen Anteil des Erklärungsgehalts des ökonometrischen Modells interpretieren. „Die Variation des Regressors $\mathbf{y}$ wird zu $R^2 \cdot 100\%$ durch die Variation der Regressoren $\mathbf{X}$ erklärt.“

4.3.1 R^2 bei zusätzlichen Regressoren

Die Regressormatrix $\mathbf{X}$ werde um die linear unabhängige und exogene Spalte $\mathbf{z}$ erweitert. $\hat{\boldsymbol{\beta}}$, SSR und R^2 werden ohne $\mathbf{z}$ berechnet, $\widetilde{SSR}$ und $\tilde{R}^2$ mit $\mathbf{z}$.

$\hat{\boldsymbol{\beta}}$ minimiert die Summe der quadrierten Residuen SSR. Dieses Minimum kann auch bei Hinzunahme der Spalte $\mathbf{z}$ erreicht werden, wenn die ersten $K+1$ Komponenten des neuen $K+2$-Schätzers den ersten $K+1$ Komponenten von $\hat{\boldsymbol{\beta}}$ entsprechen und die mit $\mathbf{z}$ assozierte Komponente des neuen Schätzers gleich null gesetzt wird. Das Minimum $\widetilde{SSR}$ muss also mindestens so klein sein wie SSR, es gilt $\widetilde{SSR} \leq SSR$. Da SST nur durch $\mathbf{y}$ berechnet wird, hat $\mathbf{z}$ keinen Einfluss auf SST und es gilt $\frac{\widetilde{SSR}}{\widetilde{SST}} \leq \frac{SSR}{SST} \Leftrightarrow \tilde{R}^2 \geq R^2$.

Der Erklärungsgehalt des um $\mathbf{z}$ erweiterten Modells ist also mindestens so groß wie ohne $\mathbf{z}$.

4.4 Spur einer Matrix und zyklische Vertauschungen

Die **Spur** einer $n \times n$-Matrix $\mathbf{A}$ ist die Summe der Elemente von $\mathbf{A}$ der Diagonalen:

$$tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$$

Ist die Matrix $\mathbf{A}$ idempotent, so gleicht die Spur von $\mathbf{A}$ dem Rang von $\mathbf{A}$.

Für die Projektionsmatrizen $\mathbf{P_X}$ und $\mathbf{P_X^\perp}$ gilt daher:

$$tr(\mathbf{P_X}) = rk(\mathbf{P_X}) = K + 1 \text{ und } tr(\mathbf{P_X^\perp}) = rk(\mathbf{P_X^\perp}) = n - K - 1$$

Der **Satz über zyklische Vertauschungen** besagt, dass die Spur eines Matrixproduktes wie folgt berechnet werden kann:

$$tr(\mathbf{ABC}) = tr(\mathbf{CAB}) \ ,$$

wobei die Matrizen $\mathbf{A}$, $\mathbf{B}$ und $\mathbf{C}$ passende Ordnungen haben.

4.5 Schätzer $\hat{\sigma}^2$ für die Varianz σ^2 der Störterme

Es sei ein Schätzer für σ^2 wie folgt definiert:

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

Um den Erwartungswert von $\hat{\sigma}^2$ zu untersuchen, ist es hilfreich den Schätzer als Funktion der unbeobachtbaren Störterme $\mathbf{u}$ darzustellen. Mit

$$\mathbf{\hat{u}} = \mathbf{P_X^\perp u}$$

gilt:

$$\mathbf{\hat{u}'\hat{u}} = \left(\mathbf{P_X^\perp u}\right)' \mathbf{P_X^\perp u} = \mathbf{u}' \left(\mathbf{P_X^\perp}\right)' \mathbf{P_X^\perp u} = \mathbf{u'P_X^\perp P_X^\perp u} = \mathbf{u'P_X^\perp u}$$

Da $\mathbf{\hat{u}'\hat{u}}$ ein Skalar ist, gilt $\mathbf{\hat{u}'\hat{u}} = tr(\mathbf{\hat{u}'\hat{u}})$. Mit dem Satz über zyklische Vertauschungen und mit $\mathbf{\hat{u}'\hat{u}} = \mathbf{u'P_X^\perp u}$ gilt

$$\mathbf{\hat{u}'\hat{u}} = tr(\mathbf{\hat{u}'\hat{u}}) = tr(\mathbf{u'P_X^\perp u}) = tr(\mathbf{uu'P_X^\perp})$$

Da $\mathbb{E}$ und tr lineare Operatoren sind, können diese vertauscht werden:

$$\mathbb{E}[\mathbf{\hat{u}'\hat{u}}|\mathbf{X}] = \mathbb{E}[tr(\mathbf{\hat{u}'\hat{u}})|\mathbf{X}] = \mathbb{E}[tr(\mathbf{uu'P_X^\perp})|X] = tr\left(\mathbb{E}\left[\mathbf{uu'}\ \mathbf{P_X^\perp}\middle|\ \mathbf{X}\right]\right) = tr\left(\mathbb{E}\left[\mathbf{uu'}|\mathbf{X}\right]\mathbf{P_X^\perp}\right)$$

Mit $\mathbb{E}[\mathbf{uu'}|\mathbf{X}] = \mathbb{E}[(\mathbf{u} - \mathbb{E}[\mathbf{u}])(\mathbf{u} - \mathbb{E}[\mathbf{u}])'|\mathbf{X}] = \mathbb{V}(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I}_n$ (**MLR 5**) folgt:

$$\mathbb{E}[\mathbf{\hat{u}'\hat{u}}|X] = tr\left(\sigma^2\mathbf{I}_n\mathbf{P_X^\perp}\right) = \sigma^2 tr\left(\mathbf{P_X^\perp}\right) = \sigma^2(n - (K + 1)) \text{ für alle } \mathbf{X}$$

und damit auch $\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}\left[\frac{1}{n-K-1}\mathbf{\hat{u}'\hat{u}}\right] = \sigma^2$.