Kapitel 3:

Multiple lineare Regression mit exogenen Regressoren

Moodle

Lehrbuch

Das klären wir in diesem Kapitel:

Matrixnotation im einfachen Regressionsmodell

Multiple Regressoren

Strikte Exogenität

Parameterschätzung

Kleinste Quadrate-Schätzung

Der Erwartungswert der OLS-Schätzer

Die bedingte Varianz des OLS-Schätzers

Matrixnotation im einfachen Regressionsmodell

Einfaches Regressionsmodell

$$y_i = \beta_0 + \beta_1 x_i + u_i \text{ für } i = 1, \ldots, n$$

Betrachte die **Vektoren**:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad \boldsymbol{\iota} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ und } \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

Wir können schreiben:

$$\mathbf{y} = \overbrace{\beta_0 \cdot \boldsymbol{\iota} + \beta_1 \cdot \mathbf{x}}^{\text{Linearkombination}} + \mathbf{u}$$

Einfaches Regressionsmodell

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

Fasse die beiden Vektoren $\boldsymbol{\iota}$ und $\mathbf{x}$ zur **Regressormatrix** X

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = (\boldsymbol{\iota}, \mathbf{x})$$

und die beiden Parameter β_0, β_1 zum **Parametervektor** $\boldsymbol{\beta}$ zusammen:

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

Dann gilt:

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

Multiple Regressoren

Multiple Regressoren

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + u_i \quad \text{für } i = 1, ..., n$$

$\beta_0, ..., \beta_k$: unbekannte Parameter

y_i: beobachtete zu erklärende Variable, Regressand

x_{ij}: beobachtete erklärende Variablen, Regressoren ($j = 1, \ldots, k$)

u_i: unbeobachtete Störterme

Fall $k = 1$: einfaches lineares Regressionsmodell

Fall $k > 1$: multiples lineares Regressionsmodell.

Das Regressionsmodell in Vektornotation

$$y_1 = \beta_0 + \beta_1 x_{11} + ... + \beta_k x_{1k} + u_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + ... + \beta_k x_{2k} + u_2$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{n1} + ... + \beta_k x_{nk} + u_n$$

$$\Leftrightarrow$$

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{y}} = \overbrace{\beta_0 \underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{\boldsymbol{\iota}} + \beta_1 \underbrace{\begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix}}_{\mathbf{x}_{\bullet 1}} + \ldots + \beta_k \underbrace{\begin{pmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{nk} \end{pmatrix}}_{\mathbf{x}_{\bullet k}}}^{\text{Linearkombination}} + \underbrace{\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}}_{\mathbf{u}}$$

Das Regressionsmodell in Vektornotation

$$\mathbf{y} = \beta_0 \boldsymbol{\iota} + \beta_1 \mathbf{x}_{\bullet 1} + \ldots + \beta_k \mathbf{x}_{\bullet k} + \mathbf{u}$$

Regressand $\mathbf{y}$:
Spaltenvektor der n zu erklärenden Beobachtungen

$\boldsymbol{\iota}$ („iota“):
Spaltenvektor mit n Einsen

Regressor $\mathbf{x}_{\bullet j}$ (für $j = 1, \ldots, k$):
Spaltenvektor der n erklärenden Beobachtungen

Störterm $\mathbf{u}$:
Spaltenvektor der n unbeobachteten sonstigen Einflüsse

$\boldsymbol{\beta}$:
Spaltenvektor der $k + 1$ unbekannten Parameter

Das Regressionsmodell in Matrixnotation

Zusammenfassen des Eins-Vektors $\boldsymbol{\iota}$ und aller Regressoren $\mathbf{x}_{\bullet 1}, \ldots, \mathbf{x}_{\bullet k}$ zu einer **Regressormatrix**:

$$X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1j} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2j} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{i1} & \ldots & x_{ij} & \ldots & x_{ik} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{n1} & \ldots & x_{nj} & \ldots & x_{nk} \end{pmatrix} = (\boldsymbol{\iota}, \mathbf{x}_{\bullet 1}, \mathbf{x}_{\bullet 2}, \ldots, \mathbf{x}_{\bullet j}, \ldots, \mathbf{x}_{\bullet k})$$

Dann gilt

$$X\boldsymbol{\beta} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{i1} & \ldots & x_{ik} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \beta_0 \boldsymbol{\iota} + \beta_1 \mathbf{x}_{\bullet 1} + \ldots + \beta_k \mathbf{x}_{\bullet k}$$

Das Regressionsmodell in Matrixnotation

Mit der Regressormatrix

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1j} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2j} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{i1} & \dots & x_{ij} & \dots & x_{ik} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nj} & \dots & x_{nk} \end{pmatrix} = (\boldsymbol{\iota}, \mathbf{x}_{\bullet 1}, \mathbf{x}_{\bullet 2}, \dots, \mathbf{x}_{\bullet j}, \dots, \mathbf{x}_{\bullet k})$$

gilt also

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

Motivation für die Multiple Regression

Wenn wir weitere Regressoren im Schätzverfahren zulassen,

- können wir den Erklärungsgehalt weiterer Faktoren nutzen,
- die Annahme strikter Exogenität besser stützen und
- flexiblere funktionale Formen zulassen.

Hierdurch können wir einen größeren Anteil der Variation von $\mathbf{y}$ erklären.

Beispiel: die Lohn-Gleichung

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u$$

Wir inkludieren nun die Expertise, also die Arbeitserfahrung (in Jahren) eines Individuums, um dessen Stundenlohn zu erklären.

Dieser Faktor war im einfachen Regressionsmodell (mit nur einem Regressor, *educ*) implizit Teil der unbeobachtbaren Größe u.

Wie zuvor müssen wir annehmen, dass der Erwartungswert der unbeobachtbaren Größen $\mathbf{u}$ nicht durch *exper* beeinflusst wird.

Dadurch, dass wir den Faktor *exper* nun direkt kontrollieren, ist die Annahme der strikten Exogenität des Faktors *educ* in Bezug auf $\mathbf{u}$ etwas weniger problematisch.

Beispiel: Durchschn. Testergebnisse und Ausgaben pro SuS

$$avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u$$

avgscore: durchschn. standardisiertes Testergebnis der Schule

expend: Ausgaben pro SuS der Schule

avginc: durchschnittliches Familieneinkommen von SuS der Schule

Es ist plausibel, dass *expend* mit *avginc* korreliert sind.

Würde *avginc* kein eigener Regressor, sondern impliziert in $\mathbf{u}$ enthalten, so wäre der verbleibende Regressor *expend* mit $\mathbf{u}$ korreliert.

Je mehr erklärende Variablen direkt kontrolliert werden, desto plausibler ist die Annahme strikter Exogenität.

Beispiel: Familieneinkommen und Familienkonsum

$$cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u$$

cons: Ausgaben für Konsum einer Familie

inc: Einkommen einer Familie

inc^2: Quadrat des Familieneinkommens

Wir können mit dem multiplen Regressionsmodell weitere funktionale Formen zulassen.

Vorsicht!

Der Parameter β_1 misst nun nicht mehr den Effekt einer Veränderung von *inc* auf *cons* bei konstanten anderen Faktoren, da inc^2 offensichtlich nicht konstant in *inc* ist.

Beispiel: Familieneinkommen und Familienkonsum

$$cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u$$

Es gilt

$$\frac{\Delta cons}{\Delta inc} = \frac{cons(inc + \Delta inc) - cons(inc)}{\Delta inc} = \beta_1 + \beta_2(2inc + \Delta inc)$$

$$\approx \beta_1 + 2\beta_2 inc$$

Der marginale Effekt des Einkommens auf den Konsum hängt nun ab von β_1, β_2 und der Höhe des Einkommens.

Beispiel: CEO Gehalt, Umsätze und Beschäftigungsdauer

$$log(salary) = \beta_0 + \beta_1 log(sales) + \beta_2 ceoten + \beta_3 ceoten^2 + u$$

salary: Gehalt (von CEO)

sales: Umsatz der Firma

ceoten: Beschäftigungsdauer der/des CEO bei der Firma

Interpretation der Parameter:

β_1: c.p. Elastizität von *salary* in Bezug auf *sales*

Falls $\beta_3 = 0$:

β_2: c.p. %-Anstieg von *salary* bei $+1$ Jahr *ceoten*

Falls $\beta_3 \neq 0$: komplizierter

Lineares Regressionsmodell bedeutet: **Linearität in Parametern**

Ausflug: lineare Algebra

Seien $\beta_0, \beta_1, \ldots, \beta_k \in \mathbb{R}$ und seien $\boldsymbol{\iota}, \mathbf{x}_{\bullet 1}, \ldots, \mathbf{x}_{\bullet k} \in \mathbb{R}^n$.

Dann ist

$$\beta_0 \boldsymbol{\iota} + \beta_1 \mathbf{x}_{\bullet 1} + \ldots + \beta_k \mathbf{x}_{\bullet k} = X\boldsymbol{\beta} \in \mathbb{R}^n$$

eine **Linearkombination** der Vektoren $\boldsymbol{\iota}, \mathbf{x}_{\bullet 1}, \ldots, \mathbf{x}_{\bullet k}$.

Beispiel für $k = 1$:

Für $\beta_0 = 3$ und $\beta_1 = -4$ gilt

$$\tilde{\mathbf{y}} = 3 \underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{\iota} + (-4) \underbrace{\begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix}}_{\mathbf{x}_{\bullet 1}} = \begin{pmatrix} 3 - 4x_{11} \\ 3 - 4x_{21} \\ \vdots \\ 3 - 4x_{n1} \end{pmatrix}$$

Spann

Die Menge aller möglichen Linearkombinationen der Vektoren $\boldsymbol{\iota}, \mathbf{x}_{\bullet 1}, \ldots, \mathbf{x}_{\bullet k}$ heißt Spann von $\{\boldsymbol{\iota}, \mathbf{x}_{\bullet 1}, \ldots, \mathbf{x}_{\bullet k}\}$:

$$span\{\boldsymbol{\iota}, \mathbf{x}_{\bullet 1}, \ldots, \mathbf{x}_{\bullet k}\} = \left\{ \tilde{\mathbf{y}} \in \mathbb{R}^n \ : \ \exists \ \mathbf{b} \in \mathbb{R}^{k+1} \text{ sodass } \tilde{\mathbf{y}} = X\mathbf{b} \right\}$$

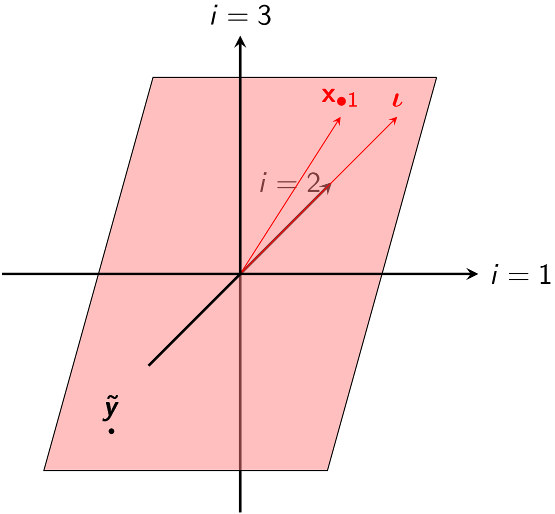
Da die Vektoren $\boldsymbol{\iota}, \mathbf{x}_{\bullet 1}, \ldots, \mathbf{x}_{\bullet k}$ die Spalten der Matrix X sind, nennen wir diesen Spann auch **Spaltenraum von** X.

Die Dimension des Spaltenraums von X entspricht dem Rang der Matrix X.

Beispiel für $k = 1$:

Falls $\boldsymbol{\iota}$ und $\mathbf{x}_{\bullet 1}$ linear unabhängig sind, gilt $rk(X) = 2$ und der Spann ist eine (Hyper-)Ebene im $\mathbb{R}^n$.

Grafische Darstellung Spaltenraum für $n = 3$ & $k = 1$



Jeder Punkt $\tilde{\boldsymbol{y}}$ in der Ebene ist durch geeignete Wahl von $b_0, b_1 \in \mathbb{R}$ darstellbar (hier: $b_0 = b_1 = -\frac{1}{2}$).

Ökonometrische Fragestellung

Können wir mit den Beobachtungen X die Beobachtungen $\mathbf{y}$ erklären?

Können wir bei gegebenen Beobachtungen $\mathbf{y}$ und X für das ökonometrische Modell

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

einen Punkt $\tilde{\boldsymbol{y}}$ im Spaltenraum von X finden, welcher die Beobachtung $\mathbf{y}$ möglichst gut erklärt?

Bemerkungen zur Notation

Für eine gegebene Regressormatrix

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1j} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2j} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{i1} & \dots & x_{ij} & \dots & x_{ik} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nj} & \dots & x_{nk} \end{pmatrix}$$

bezeichnet

- $\mathbf{x}_{\bullet j}$ den j-ten $(n \times 1)$-**Spaltenvektor** für $j = 0, 1, \dots, k$

  und

- $\mathbf{x}_{i \bullet}$ den i-ten $(1 \times k + 1)$-**Zeilenvektor** für $i = 1, \dots, n$

Strikte Exogenität

Strikte Exogenität

- Ceteris-paribus Interpretation der Parameter $\beta_1, \ldots, \beta_k$:

  Wenn x_{ij} sich um eine Einheit ändert (*und sonst ändert sich nichts*), um wieviele Einheiten ändert sich dann y_i?

- Voraussetzung für ceteris-paribus Betrachtungsweise:

  Die $\mathbf{x}_{\bullet j}$ - Variablen dürfen nicht mit dem Störterm $\mathbf{u}$ zusammenhängen.

- Die durch den Regressor $\mathbf{x}_{\bullet j}$ verursachte Variation in $\mathbf{y}$ muss unterscheidbar sein von der durch den Störterm $\mathbf{u}$ verursachten Variation in $\mathbf{y}$.

  Das geht nur, wenn $E[\mathbf{u}|X] = E[\mathbf{u}]$.

Strikte Exogenität

$$E\left[\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}\right] = \begin{pmatrix} E[u_1] \\ \vdots \\ E[u_n] \end{pmatrix}$$

Die Bedingung $E[\mathbf{u}|X] = E[\mathbf{u}]$ bedeutet

$$E[u_i|\mathbf{x}_{\bullet 1}, \ldots, \mathbf{x}_{\bullet k}] = E[u_i] \text{ für } i = 1, \ldots, n$$

In Worten:

Die Werte der Regressormatrix X haben keinen Einfluss auf die Erwartungswerte der Störterme.

Diese Annahme bzw. Anforderung wird auch als **strikte Exogenität** der Regressoren bezeichnet. Sie ist von zentraler Bedeutung für Regressionsmodelle.

Analog zur Argumentation in Kapitel 2 nehmen wir ohne Beschränkung der Allgemeinheit an, dass $E[\mathbf{u}] = \mathbf{0}$.

$$\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \qquad \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Strikte Exogenität und Unkorreliertheit

$$E[x_{ij} u_i] - E[x_{ij}] \cdot \underbrace{E[u_i]}_{=0}$$

Unter strikter Exogenität gilt für die Kovarianz einer Regressorvariable x_{ij} und eines Störterms u_i:

$$Cov(x_{ij}, u_i) = E[x_{ij} u_i] = E[x_{ij} \underbrace{E[u_i|x]}_{= E[u_i] = 0}] = 0$$

Die Begründung ist analog zur entsprechenden Argumentation in Kapitel 2 (Folie 20).

Parameterschätzung

Residuen

Schätze die unbekannten Parameter $\beta_0, \beta_1, \dots, \beta_k$

unbekannt

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i, \; i = 1, \dots, n$$

zunächst durch beliebige Zahlen $b_0, b_1, \dots, b_k$.

$\rightarrow$ prognostizierten Werte $\hat{y}_i$ (**fitted values**) mit

$$\hat{y}_i = b_0 + b_1 x_{i1} + \dots + b_k x_{ik}, \; i = 1, \dots, n$$

in Matrix-Notation:

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

mit $\mathbf{b}' = (b_0, b_1, \dots, b_k)$.

Residuen:

$$\hat{\boldsymbol{u}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \mathbf{y} - X\mathbf{b}$$

Schätzung mit der Momentenmethode (für $k > 1$)

Unter der Annahme strikt exogener Regressoren gilt für $j = 0, 1, \ldots k$ (mit $\mathbf{x}_{\bullet 0} = \boldsymbol{\iota}$):

$$Cov(x_{ij}, u_i) = 0 \; \forall \; i = 1, \ldots, n$$

$\overline{\hat{u}} = 0$

$\overline{\hat{u} x_1} = 0$

$\vdots$

$\overline{\hat{u} x_k} = 0$

Wähle $\mathbf{b} = \hat{\boldsymbol{\beta}}$, sodass für jedes $j = 0, 1, \ldots, k$

$$\frac{1}{n} \sum_{i=1}^{n} x_{ij} \hat{u}_i = \underline{\frac{1}{n} \sum_{i=1}^{n} x_{ij} \left(y_i - \mathbf{x}_{i\bullet} \hat{\boldsymbol{\beta}} \right) = 0}$$

?

In Matrixnotation ist dies äquivalent zu:

X': Transponierte Matrix von X

$$X' \mathbf{y} = X' X \hat{\boldsymbol{\beta}}$$

$(X'X)^{-1} \cdot X'y = \underbrace{(X'X)^{-1} X'X}_{I} \hat{\beta} = \hat{\beta}$

Falls $(X'X)^{-1}$ existiert, ergibt sich

$X'X$: Momentematrix

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

Ausflug lineare Algebra

Momentematrix $\underbrace{\underset{k+1\times n}{X'}\ \underset{n\times k+1}{X}}_{k+1\times k+1}$

Die $(k+1) \times (k+1)$-Matrix $\boxed{X'X}$ ist **invertierbar**, falls sie den **Rang** $k+1$ hat.

Der Rang des Produktes $X'X$ der zwei Matrizen X' und X entspricht dem Minimum der Ränge beider Matrizen:

$$rk(X'X) = \min\{rk(X'), rk(X)\} = rk(X)$$

Der Rang einer Matrix ist die Anzahl linear unabhängiger **Spalten** bzw. **Zeilen**.

Die Matrix $X'X$ ist also invertierbar, falls die Regressormatrix X vollen Spaltenrang hat (wegen $k+1 < n$).

X hat vollen Spaltenrang, falls keine Spalte von X durch Linearkombinationen der anderen Spalten erzeugt werden kann.

Diese Voraussetzung bezeichnet man auch als **Freiheit von Multikollinearität**.

Kleinste Quadrate-Schätzung

Kleinste Quadrate-Schätzung

Eine andere Schätzmöglichkeit besteht in der Vorgabe einer Zielfunktion.

Die gebräuchlichste ist:

Minimiere die Summe der quadrierten Abweichungen zwischen der erklärten Variablen und ihren durch die Schätzung vorhergesagten Werten – den Residuen.

Dies ist das gewöhnliche Kleinste-Quadrate-Prinzip (**OLS, ordinary least squares**).

Kleinste Quadrate-Schätzung

$$(a_1, a_2, \dots, a_n)\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a_1^2 + a_2^2 + \dots + a_n^2 = \sum_{i=1}^{n} a_i^2$$

Der durch die Wahl von **b** verursachte Prognosefehler erzeugt **Residuen $\hat{\boldsymbol{u}}$**: die Abweichungen von Beobachtungen y_i und prognostizierten Werten $\hat{y}_i$.

$$\hat{\boldsymbol{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - X\mathbf{b}$$

Je besser der Schätzer **b**, desto kürzer der Vektor der Residuen $\hat{\boldsymbol{u}}$.

$\Rightarrow$ Zielfunktion: Minimiere $||\hat{\boldsymbol{u}}||$ bzw. $\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$!

$$\min_{\mathbf{b}\in\mathbb{R}^{K+1}} \underbrace{(\mathbf{y} - X\mathbf{b})'}_{\hat{u}'} \underbrace{(\mathbf{y} - X\mathbf{b})}_{\hat{u}}$$

Wiederholung lineare Algebra (siehe Vorkurs)

Sei A eine Matrix und $\mathbf{x}$ ein Vektor (mit passenden Dimensionen).

Dann gilt für $f(\mathbf{x}) = \mathbf{x}'A$: $= (A' \cdot x)'$

"nabla"

Gradient $$\nabla f(\mathbf{x}) = A$$ Vektor der 1. Ableitungen

und für $g(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$:

$$\nabla g(\mathbf{x}) = 2A\mathbf{x}$$

$$S(b) = \underbrace{(y - Xb)'}_{\substack{\hat{u}' \\ 1\times n}} \cdot \underbrace{(y - Xb)}_{\substack{\hat{u} \\ n\times 1}} = \underbrace{y' \cdot y}_{1\times 1} - \underbrace{\overbrace{y'Xb}^{=(y'Xb)'}}_{1\times 1} - \underbrace{(Xb)'y}_{1\times 1} + \underbrace{(Xb)'Xb}_{1\times 1}$$

$$1\times 1$$

$$= y' \cdot y - b'X'y - b'X'y + b'X'Xb$$

$$= y' \cdot y - 2b'\underbrace{X'y}_{A} + b'\underbrace{X'X}_{\tilde{A}}b$$

$$\nabla S(b) = -2X'y + 2X'Xb \overset{BEO}{=} 0$$

$$\Leftrightarrow \cancel{2}X'Xb = \cancel{2}X'y$$

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

$$HS(b) = 2X'X$$

Kleinste Quadrate-Schätzung

Damit lautet die notwendige Bedingung erster Ordnung des Minimierungsproblems:

$$\frac{\partial\left(\mathbf{y}-X\mathbf{b}\right)'\left(\mathbf{y}-X\mathbf{b}\right)}{\partial\mathbf{b}} = -2X'\mathbf{y}+2X'X\mathbf{b} \overset{!}{=} 0$$

Dies ist äquivalent zu

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

und analog zu MM folgt bei Invertierbarkeit von $X'X$:

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

Algebraische Eigenschaften

$$\underbrace{\underset{n\times k+1}{X}\;\underset{k+1\times 1}{b}}_{n\times 1}$$

Bemerkungen zu $X'X$

Es gilt:

$$\left(X'X\right)' = (X)'\left(X'\right)' = X'X$$

Die Matrix $X'X$ ist **symmetrisch**.

Voller Spaltenrang: (alle Spalten von X sind lin. unabh.)

Wegen $rk(X) = k+1$ gilt $X\mathbf{b} = \mathbf{0}$ für $\mathbf{b} \in \mathbb{R}^{k+1} \Rightarrow \mathbf{b} = \mathbf{0}$

Demnach gilt für $\mathbf{b} \neq \mathbf{0}$: $X\mathbf{b} \neq \mathbf{0}$.

Für $\mathbf{b} \in \mathbb{R}^{k+1} \setminus \{\mathbf{0}\}$ gilt

$$\mathbf{b}'X'X\mathbf{b} = (X\mathbf{b})' X\mathbf{b} = ||X\mathbf{b}||^2 > 0$$

positiv definit
symmetrisch
"pds"

Die Matrix $X'X$ is **positiv definit**, falls rk(X)=k+1.

Def. führende Hauptminoren für 3X3 Matrizen

$$A = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$$

Führender Hauptminor 1. Ordnung

$$\begin{matrix} a & b & c \\ b & d & e \\ c & e & f \end{matrix}$$

$$= \det(a) = a$$

2. Ordnung

$$\begin{matrix} a & b & c \\ b & d & e \\ c & e & f \end{matrix}$$

$$= \det\begin{pmatrix} a & b \\ b & d \end{pmatrix} = a \cdot d - b \cdot b$$

3. Ordnung $= \det(A)$

Eine Matrix $\underset{n \times n}{A}$ ist positiv definit, falls alle führenden Hauptminoren positiv sind.

Falls $X'X$ pd ist, muss also gelten $\det(X'X) > 0$

$\Rightarrow (X'X)^{-1}$ existiert

Algebraische Eigenschaften

Behauptung:

Der Schätzer $\hat{\boldsymbol{\beta}}$ ist die eindeutige Minimumstelle.

Die Hesse-Matrix der Zielfunktion

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

lautet $2X'X$.

Da $X'X$ positiv definit ist, ist die Zielfunktion streng konvex und

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

ist die eindeutige Minimumstelle.

Algebraische Eigenschaften

$$\underbrace{\underbrace{X'}_{k+1\times n}\underbrace{X}_{n\times k+1}}_{k+1\times k+1}$$

Es gilt

$\hat{\beta}(y+\tilde{y}) = \hat{\beta}(y) + \hat{\beta}(\tilde{y})$

$\hat{\beta}(\lambda y) = \lambda \hat{\beta}(y)$

Der Schätzer

$$\hat{\boldsymbol{\beta}} = \underbrace{\underbrace{\left(X'X\right)^{-1}}_{k+1\times k+1}\underbrace{X'}_{k+1\times n}}_{k+1\times n}\mathbf{y}$$

ist linear in $\mathbf{y}$.

Prognose:

$$\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}} = \underbrace{X\left(X'X\right)^{-1}X'}_{n\times n}\mathbf{y}$$

Auch die Prognose $\hat{\mathbf{y}}$ ist linear in $\mathbf{y}$.

Residuen:

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

Auch die Residuen sind linear in $\mathbf{y}$.

Algebraische Eigenschaften

Behauptung:

Die Matrix $\overbrace{X(X'X)^{-1}X'}^{n \times n}$ projiziert den Vektor $\mathbf{y}$ orthogonal auf den Spaltenraum von X.

Wir nennen $X(X'X)^{-1}X'$ deshalb auch **Projektionsmatrix**.

Begründung:

1. Die Spalten von X sind orthogonal zum Vektor $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}$:

$$X'\underbrace{(\mathbf{y} - \hat{\mathbf{y}})}_{\hat{u}} = X'\left(\mathbf{y} - \underbrace{X\underbrace{\left(X'X\right)^{-1}X'\mathbf{y}}_{\hat{\beta}}}_{\hat{y}}\right) = X'\mathbf{y} - \underbrace{X'X\left(X'X\right)^{-1}}_{I}X'\mathbf{y} = \mathbf{0}$$

2. $\hat{y} = X\hat{\boldsymbol{\beta}}$ liegt im Spaltenraum von X:

$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$ ist ein $(k+1) \times 1$-Spaltenvektor. Daher ist
$\hat{y} = \underset{n \times k+1}{X}\underset{k+1 \times 1}{\hat{\boldsymbol{\beta}}} = X(X'X)^{-1}X'\mathbf{y}$ eine Linearkombination der Spalten von X.

Algebraische Eigenschaften

Für die OLS-Residuen gilt also

$$X'(\mathbf{y} - \hat{\mathbf{y}}) = X'\hat{\mathbf{u}} = \mathbf{0}$$

Zeile für Zeile geschrieben bedeutet diese Gleichung:

$$\sum_{i=1}^{n} x_{ij}\hat{u}_i = 0 \text{ für } j = 0, 1, \ldots, k$$

Diese Eigenschaft hatten wir schon für die OLS-Residuen in Kapitel 2 für $k = 1$ gezeigt:

$$\sum_{i=1}^{n} 1 \cdot \hat{u}_i = 0 \text{ und } \sum_{i=1}^{n} x_i\hat{u}_i = 0$$

Algebraische Eigenschaften

Die Stichprobendurchschnitte liegen auf der "Regressionsgeraden":

$$\bar{y} = \sum_{j=0}^{k} \hat{\beta}_j \bar{x}_j$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij} = \frac{1}{n} (1, 1, \dots, 1) \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix} = \frac{1}{n} \iota' \cdot x_{\cdot j}$$

$$\sum_{i=1}^{n} \overset{\overset{1}{\downarrow}}{x_{i0}} \cdot \hat{u}_i = 0 \Leftrightarrow \sum_{i=1}^{n} 1 \cdot \hat{u}_i = 0 \Leftrightarrow \sum_{i=1}^{n} \hat{u}_i = 0$$

mit $\hat{u}_i = y_i - \hat{y}_i$: $\sum_{i=1}^{n} y_i - \hat{y}_i = 0 \Leftrightarrow \underbrace{\frac{1}{n} \sum_{i=1}^{n} y_i}_{\bar{y}} = \frac{1}{n} \sum_{i=1}^{n} \hat{y}_i$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} \hat{y}_i = \frac{1}{n} (1, 1, \dots, 1) \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \frac{1}{n} \iota' \cdot \hat{y}$$

$$\hat{y} = X \cdot \hat{\beta} \Rightarrow \bar{y} = \frac{1}{n} \iota' X \cdot \hat{\beta} = \frac{1}{n} \iota' (x_{\cdot 0}, x_{\cdot 1}, \dots, x_{\cdot k}) \cdot \hat{\beta}$$

$$= \frac{1}{n} (\iota' x_{\cdot 0}, \iota' x_{\cdot 1}, \dots, \iota' x_{\cdot k}) \hat{\beta} = \frac{1}{n} \left(\sum_{i=1}^{n} x_{i0}, \sum_{i=1}^{n} x_{i1}, \dots, \sum_{i=1}^{n} x_{ik} \right) \hat{\beta}$$

$$= \Big(\underbrace{\frac{1}{n} \sum_{i=1}^{n} x_{i0}}_{\bar{x}_0}, \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_{i1}}_{\bar{x}_1}, \dots, \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_{ik}}_{\bar{x}_k} \Big) \cdot \hat{\beta} = \sum_{j=0}^{k} \hat{\beta}_j \cdot \bar{x}_j$$

Algebraische Eigenschaften

Die Stichprobendurchschnitte liegen auf der Regressionsgeraden:

$$\bar{y} = \sum_{j=0}^{k} \hat{\beta}_j \bar{x}_j$$

Begründung:

$$\sum_{i=1}^{n} y_i - \hat{y}_i = \sum_{i=1}^{n} \hat{u}_i = 0$$

Daraus folgt:

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i = \frac{1}{n}\boldsymbol{\iota}'\hat{\mathbf{y}} = \frac{1}{n}\boldsymbol{\iota}' X \hat{\boldsymbol{\beta}} = \frac{1}{n}\left(\boldsymbol{\iota}'\mathbf{x}_{\bullet 0}, \boldsymbol{\iota}'\mathbf{x}_{\bullet 1}, \ldots, \boldsymbol{\iota}'\mathbf{x}_{\bullet k}\right)\hat{\boldsymbol{\beta}}$$

Vergleiche hierzu für $k = 1$:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Algebraische Eigenschaften

Die Matrix $I - X(X'X)^{-1}X'$ projiziert den Vektor $\mathbf{y}$ auf das orthogonale Komplement des Spaltenraums von X.

Auch $I - X(X'X)^{-1}X'$ wird deshalb **Projektionsmatrix** genannt.

Da eine Matrix genau dann ein Projektionsmatrix ist, falls sie auch **idempotent symmetrisch** ist, lässt sich diese Eigenschaft auch einfacher nachweisen.

Eine Matrix A heißt idempotent, falls $AA = A$.

Für $X(X'X)^{-1}X'$:

$$(X(X'X)^{-1}X')\,(X(X'X)^{-1}X') = X(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{=I}X'$$

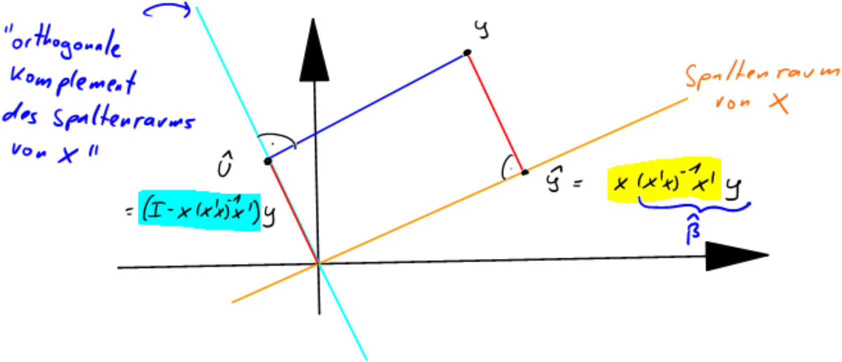
Übung: Zeige, dass auch $I - X(X'X)^{-1}X'$ idempotent ist.

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

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

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

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



Interpretation des multiplen Regressionsmodells

Wie stark ändert sich die abhängige Variable y, wenn **ceteris paribus** die j-te unabhängige Variable um eine Einheit ansteigt:

$$\beta_j = \frac{\Delta y}{\Delta x_j}$$

ceteris paribus:

Alle anderen unabhängigen Variablen bleiben konstant.

Es muss strikte Exogenität angenommen werden, d.h. die Veränderung Δx_j hat keinen Einfluss auf die unbeobachtbaren Störterme.

Beispiel: Durchschn.-Noten von College und High school

Datensatz: GPA1 (n=141 SuS)

$$\widehat{colGPA} = 1.29 + .453 hsGPA + .0094 ACT$$

$\widehat{colGPA}$: OLS-Prognose durchschn.-Note College

$hsGPA$: durchschn.-Note Highschool

ACT: Achievement test score (misst akademische Fähigkeiten)

Interpretation der Schätzer .453 und .0094:

Im Durchschnitt bringt ein $hsGPA$ etwa einen halben $colGPA$.

Im Durchschnitt bringen zehn ACT weniger als einen Zehntel $colGPA$.

Beispiel: Durchschn.-Noten von College und High school

Datensatz: GPA1 (n=141 SuS)

$$\widehat{colGPA} = 2.40 + .0271ACT$$

$\widehat{colGPA}$: OLS-Prognose Durchschn.-Note College

ACT: Achievement test score (misst akademische Fähigkeiten)

Der Einfluss von ACT auf $colGPA$ hat sich scheinbar verdreifacht, weil wir den Einfluss von $hsGPA$ vernachlässigen!

Partitionierte Regression

Das vorherige Beispiel zeigt:

Wenn wir den Einfluss eines bestimmten Regressors auf y ermitteln wollen, müssen wir alle relevanten Regressoren mit aufnehmen.

Nehmen wir als Beispiel das Modell

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \mathbf{u}$$

Falls $\mathbf{x}_1$ und $\mathbf{x}_2$ untereinander korreliert sind, würde eine Schätzung von β_1 im Modell

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \mathbf{u}$$

zu Fehlinformationen führen, da $\mathbf{x}_2$ in $\mathbf{u}$ enthalten wäre und dann $\mathbf{x}_1$ mit $\mathbf{u}$ (über $\mathbf{x}_2$) korreliert wäre, was die strikte Exogenität verletzen würde.

Deswegen braucht man die multiple Regression, die $\mathbf{x}_1$ und $\mathbf{x}_2$ gleichzeitig berücksichtigt.

Partitionierte Regression

Es gibt aber eine Methode, das Ergebnis der multiplen Regression

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \mathbf{u}$$

durch eine Sequenz von Einfachregressionen zu erhalten, nämlich die **partitionierte Regression:**

$\rightarrow \hat{\gamma}_0, \hat{\gamma}_1 \rightsquigarrow \hat{y} \rightsquigarrow \hat{r}_y = y - \hat{y}$

1. Schätze $\mathbf{y} = \gamma_0 + \gamma_1 \mathbf{x}_2 + \mathbf{r}_y$ und ermittle die Residuen $\widehat{\mathbf{r}}_y$.
2. Schätze $\mathbf{x}_1 = \delta_0 + \delta_1 \mathbf{x}_2 + \mathbf{r}_1$ und ermittle die Residuen $\widehat{\mathbf{r}}_1$.
3. Schätze $\widehat{\mathbf{r}}_y = \alpha_0 + \alpha_1 \widehat{\mathbf{r}}_1 + \mathbf{u}$

$\hat{\delta}_1 = \frac{S_{x_1 x_2}}{S_{x_2 x_2}}$

Dann gilt:

$$\widehat{\alpha}_1 = \widehat{\beta}_1$$

Partitionierte Regression

Diese Aussage folgt aus dem für die Regressionstheorie fundamentalen **Frisch-Waugh-Theorem**.

Das Einbeziehen von $\mathbf{x}_2$ in das Modell nimmt den Einfluss dieser Variablen auf $\mathbf{y}$ von $\mathbf{x}_1$ weg.

Mit anderen Worten: OLS **kontrolliert** für den Einfluss von $\mathbf{x}_2$ auf den Zusammenhang zwischen $\mathbf{y}$ und $\mathbf{x}_1$. Das Resultat gilt analog auch für mehrere Variablen gemeinsam (also für Gruppen von Variablen).

Deshalb bezeichnet man Regressoren auch als **Kontrollvariablen**.

Das erlaubt die ceteris-paribus - Interpretation der geschätzten Parameter einer multiplen Regression.

Anpassungsgüte (Goodness-of-Fit) – analog zu Kapitel 2

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

$\bar{y} = \bar{\hat{y}}$

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

SSR $= \sum_{i=1}^{n}(\hat{u}_i - 0)^2$
(sum of squared residuals)

Es gilt:

$$SST = SSE + SSR$$

Bestimmheitsmaß:

$$R^2 = SSE/SST = 1 - SSR/SST$$

Anpassungsgüte (Goodness-of-Fit)

Wie verändert sich die Anpassungsgüte bei multiplen Regressoren?

Man kann zeigen, dass SSE nicht sinken bzw. SSR nicht steigen kann, wenn ein weiterer Regressor hinzugefügt wird.

Intuition:

- Ist der Regressor unwichtig, so ist der Schätzer für dessen Parameter nahe null und die Prognose $\hat{y}$ verändert sich kaum.
- Ist der Regressor wichtig, so übernimmt der Schätzer für dessen Parameter einen Teil des Einflusses anderer Regressoren und die Prognose verbessert sich (SSE steigt, SSR sinkt).

Daher kann R^2 nicht sinken, wenn weitere Regressoren hinzugefügt werden.

Beispiel: Erklärung für Verhaftungen [CRIME1]

$$\widehat{narr86} = .712 - .150 pcnv - 0.034 ptime86 - .104 qemp86$$

Stichprobengröße: 2.725, $R^2 = 0.0413$

Individuen: Männer, die 1960 oder 1961 in Kalifornien geboren wurden und vor 1986 mindestens einmal verhaftet wurden.

$\widehat{narr86}$: Prognose für die Anzahl von Verhaftungen in 1986

$pcnv$: Anteil der vorangegangenen Verhaftungen, die zu einer Verurteilung führten

$ptime1986$: Monate im Gefängnis in 1986

$qemp86$:
Anzahl der Quartale von 1986, in welchen der Mann arbeitete

Beispiel: Erklärung für Verhaftungen [CRIME1]

Was passiert, wenn wir *avgsen* inkludieren?

avgsen: durchschn. Haftstrafe für vorangegangene Verurteilungen

$$\widehat{narr86} = .707 - .151 pcnv + .0074 avgsen - 0.037 ptime86 - .103 qemp86$$

Stichprobengröße: 2.725, $R^2 = 0.0422$

R^2 steigt nur sehr leicht an.

Ist der Regressor *avgsen* wirklich wertvoll?

Durch die Aufnahme irrelevanter Regressoren steigt in der Regel die Varianz der Schätzer, so dass die Inferenz unpräziser wird. Dazu später mehr.

Der Erwartungswert der OLS-Schätzer

Der Erwartungswert der OLS-Schätzer

- Die Parameterschätzer des Regressionsmodells, also die Elemente von $\hat{\boldsymbol{\beta}}$, sind selbst Zufallsvariablen.
- Das wahre $\boldsymbol{\beta}$ ist unbekannt. Jede Stichprobe wird ein etwas anderes Schätzergebnis $\hat{\boldsymbol{\beta}}$ liefern.
- Damit hat der Schätzer $\hat{\boldsymbol{\beta}}$ eine Wahrscheinlichkeitsverteilung. Diese müssen wir ermitteln, damit wir Inferenz betreiben können, d.h. von dem Stichprobenergebnis auf die Grundgesamtheit schließen können.
- Damit das gelingt, müssen einige Voraussetzungen erfüllt sein, die im Folgenden zusammengefasst sind:

Annahmen des multiplen Regressionsmodells

MLR 1 **Linear in den Parametern**

Das ökonometrische Modell lautet $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$, wobei

$X \in \mathbb{R}^{n\times(k+1)}$ bekannt mit $k < n$,

$\boldsymbol{\beta} \in \mathbb{R}^{k+1}$ unbekannt und nicht-stochastisch und

$\mathbf{u} \in \mathbb{R}^n$ unbekannt und stochastisch.

MLR 3 **Lineare Unabhängigkeit**

Die Regressormatrix X hat vollen Spaltenrang, $rk(X) = k + 1$.

MLR 4 **Bedingter Nullerwartungswert**

Die Regressoren sind strikt exogen und der Erwartungswert der Störterme ist null: $E[\mathbf{u}|X] = E[\mathbf{u}] = \mathbf{0}$.

Wir schenken der Annahme MLR 2 „Zufallsstichprobe“ keine weitere Beachtung.

Beispiele für lineare Abhängigkeit

$$\log(cons) = \beta_0 + \beta_1 \log(inc) + \beta_2 \log(inc^2) + u$$

$$voteA = \beta_0 + \beta_1 shareA + \beta_2 shareB + u$$

Weiterhin kann eine sehr kleine Stichprobe problematisch sein:

Die Matrix X darf nicht weniger Zeilen als Spalten haben. Falls $n < k + 1$ gilt $rk(X) \leq n < k + 1$.

Beispiele für lineare Abhängigkeit

$$\log(cons) = \beta_0 + \beta_1 \log(inc) + \beta_2 \log(inc^2) + u$$

Wegen $\log(x^a) = a \log(x)$ sind $\log(inc)$ und $\log(inc^2)$ linear abhängig.

$$voteA = \beta_0 + \beta_1 shareA + \beta_2 shareB + u$$

Weiterhin kann eine sehr kleine Stichprobe problematisch sein:

Die Matrix X darf nicht weniger Zeilen als Spalten haben. Falls $n < k + 1$ gilt $rk(X) \leq n < k + 1$.

Beispiele für lineare Abhängigkeit

$$\log(cons) = \beta_0 + \beta_1 \log(inc) + \beta_2 \log(inc^2) + u$$

Wegen $\log(x^a) = a \log(x)$ sind $\log(inc)$ und $\log(inc^2)$ linear abhängig.

$$voteA = \beta_0 + \beta_1 shareA + \beta_2 shareB + u$$

Wegen $shareA + shareB = 1$ sind $shareA$ und $shareB$ und die Konstante 1 linear abhängig.

Weiterhin kann eine sehr kleine Stichprobe problematisch sein:

Die Matrix X darf nicht weniger Zeilen als Spalten haben. Falls $n < k + 1$ gilt $rk(X) \leq n < k + 1$.

Der Erwartungswert als linearer Operator

Sei $\mathbf{c} \in \mathbb{R}^p$ ein zufälliger Vektor.

Dann gilt:

$$E[\mathbf{c}] = \begin{pmatrix} E[c_1] \\ E[c_2] \\ \vdots \\ E[c_p] \end{pmatrix}$$

Seien zusätzlich $A \in \mathbb{R}^{o \times p}$ und $\mathbf{b} \in \mathbb{R}^p$ nicht-stochastisch.

Dann gilt:

$$E[A\mathbf{c} + \mathbf{b}] = AE[\mathbf{c}] + \mathbf{b}$$

Erwartungswert des OLS-Schätzers

Behauptung:

$$E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} \;\forall\; \boldsymbol{\beta} \in \mathbb{R}^{k+1}$$

Beweis:

MLR 1

$$\boxed{\hat{\beta} = (X'X)^{-1}X'y}\;,\; y = X\beta + u \qquad \boxed{\hat{\beta} = \beta + (X'X)^{-1}X'u}$$

$$= (X'X)^{-1}X'(X\beta + u) = \underbrace{(X'X)^{-1}X'X}_{I}\beta + (X'X)^{-1}X'u$$

$$E[\hat{\beta} \mid X] = E[\beta + (X'X)^{-1}X'u \mid X] = \beta + (X'X)^{-1}X'\underbrace{E[u \mid X]}_{\text{MLR 4} = 0} = \beta$$

$$E[E[\hat{\beta} \mid X]] = E[\beta] = \beta \quad \checkmark$$

Erwartungswert des OLS-Schätzers

Behauptung:

$$E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} \; \forall \; \boldsymbol{\beta} \in \mathbb{R}^{k+1}$$

Beweis:

Mit MLR 3 ($rk(X) = k + 1$) und MLR 1 ($k < n$) gilt:

$$rk(X'X) = \min\{rk(X'), rk(X)\} = rk(X) = k + 1$$

Da $X'X$ die Ordnung $(k+1) \times (k+1)$ hat, existiert $(X'X)^{-1}$ und der Schätzer $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$ ist wohldefiniert.

Da der Erwartungswert ein linearer Operator ist, gilt mit MLR 1 ($\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$, X bekannt und $\boldsymbol{\beta}$ nicht-stochastisch)

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

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

$$E[\mathbf{y}|X] = X\boldsymbol{\beta}$$

Erwartungswert des OLS-Schätzers

Behauptung:

$$E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} \ \forall \ \boldsymbol{\beta} \in \mathbb{R}^{k+1}$$

Beweis:

Mit MLR 3 ($rk(X) = k + 1$) und MLR 1 ($k < n$) gilt:

$$rk(X'X) = \min\{rk(X'), rk(X)\} = rk(X) = k + 1$$

Da $X'X$ die Ordnung $(k+1) \times (k+1)$ hat, existiert $(X'X)^{-1}$ und der Schätzer $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$ ist wohldefiniert.

Da der Erwartungswert ein linearer Operator ist, gilt mit MLR 1 ($\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$, X bekannt und $\boldsymbol{\beta}$ nicht-stochastisch)

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

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

$$E[\mathbf{y}|X] = X\boldsymbol{\beta}$$

Erwartungswert des OLS-Schätzers

Beweis (Fortsetzung):

Aufgrund der Linearität des Schätzers $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$ in $\mathbf{y}$ gilt:

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

Mit $E[\mathbf{y}|X] = X\boldsymbol{\beta}$ gilt:

$$E[\hat{\boldsymbol{\beta}}|X] = \underbrace{(X'X)^{-1}X'X}_{=I}\boldsymbol{\beta} \;\forall\; \boldsymbol{\beta} \in \mathbb{R}^{k+1}$$

Mit dem Gesetz der iterierten Erwartung und MLR 1 ($\boldsymbol{\beta}$ nicht-stochastisch) gilt:

$$E[\hat{\boldsymbol{\beta}}] = E[E[\hat{\boldsymbol{\beta}}|X]] = E[\boldsymbol{\beta}] = \boldsymbol{\beta} \;\forall\; \boldsymbol{\beta} \in \mathbb{R}^{k+1} \;\square$$

Der lineare Schätzer $\hat{\boldsymbol{\beta}}$ heißt **erwartungstreu** oder **unverzerrt**.

Annahmen

MLR 1 $\quad Y = X\beta + U$

MLR 3 $\quad rg(X) = k+1$

MLR 4 $\quad E[U|X] = 0 \in \mathbb{R}^n$

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

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

$$\hat{U} = y - \hat{y} = y - X \cdot (X'X)^{-1} \cdot X'y$$

$$= \underbrace{[I - X(X'X)^{-1}X']}_{P} y$$

("Projektionsmatrix")

Sei $\check{\beta} = A \cdot y$ (linear in y).

Unter welcher Bedingung gilt $E[\check{\beta}] = \beta$.

$$E[A \cdot y] = E[A(X\beta + u)] \overset{\text{MLR 1}}{=} E[\underbrace{A \cdot X \cdot \beta}_{\text{fest}} + \underbrace{A}_{\text{fest}} \cdot \overset{\text{zufällig}}{u}]$$

$$= A \cdot X \cdot \beta + A \cdot \underbrace{E[u]}_{\substack{=0 \\ \text{MLR 4}}}$$

$$= A \cdot X \cdot \beta = \beta \quad \text{für alle } \beta \in \mathbb{R}^{k+1}$$

$$\Leftrightarrow \quad A \cdot X = I$$

Irrelevante Regressoren (Überspezifikation)

Was passiert, wenn wir einen Regressor inkludieren, dessen wahrer Parameter $\beta_j = 0$ ist?

Wir hatten die Erwartungstreue des OLS-Schätzers für alle möglichen Vektoren von Parametern $\boldsymbol{\beta} \in \mathbb{R}^{k+1}$ gezeigt.

Gilt also $\beta_j = 0$ für einen dieser Parameter, so gilt unter den getroffenen Annahmen auch

- $E[\hat{\beta}_j] = 0$ und
- $E[\hat{\beta}_s] = \beta_s$ für alle $s \neq j$.

Die Erwartungstreue der anderen Schätzer bleibt also unberührt.

Ausgelassene Regressoren (omitted variable bias)

Wir schauen uns nun das Problem ausgelassener Regressoren genauer an.

Das wahre Modell sei gegeben durch

$$\mathbf{y} = X_1\boldsymbol{\beta} + \mathbf{x}_2\gamma + \mathbf{u} \ ,$$

wobei

- X_1 eine $n \times (k+1)$-Regressormatrix und
- $\mathbf{x}_2$ ein weiterer Regressor ist.
- $\boldsymbol{\beta}$ ist ein Vektor mit $k+1$ unbekannten Parametern und
- γ ist ein einzelner unbekannter Parameter.

Es gelte die Annahme strikter Exogenität für $\mathbf{u}$, also $E[\mathbf{u}|X_1, \mathbf{x}_2] = E[\mathbf{u}] = \mathbf{0}$.

Ausgelassene Regressoren (omitted variable bias)

Geschätzt werde aber irrtümlich das Modell

$$\mathbf{y} = X_1\boldsymbol{\beta} + \tilde{\mathbf{u}} \ ,$$

wobei $\tilde{\mathbf{u}} = \mathbf{x}_2\gamma + \mathbf{u}$ den unbeobachtbaren Störterm bezeichne.

$$\Rightarrow \tilde{\boldsymbol{\beta}} = \left(X_1'X_1\right)^{-1} X_1'\mathbf{y}$$

Wenn nun aber $\mathbf{y}$ tatsächlich durch das Modell

$$\mathbf{y} = X_1\boldsymbol{\beta} + \mathbf{x}_2\gamma + \mathbf{u}$$

definiert ist, ergibt sich

$$\tilde{\boldsymbol{\beta}} = \left(X_1'X_1\right)^{-1} X_1'\left(X_1\boldsymbol{\beta} + \mathbf{x}_2\gamma + \mathbf{u}\right) = \boldsymbol{\beta} + \left(X_1'X_1\right)^{-1} X_1'\left(\mathbf{x}_2\gamma + \mathbf{u}\right)$$

$$= \underbrace{(X_1'X_1)^{-1} X_1'X_1}_{I}\beta + (X_1'X_1)^{-1}X_1'x_2\gamma + (X_1'X_1)^{-1}X_1' \cdot u$$

Ausgelassene Regressoren (omitted variable bias)

Ist der Schätzer $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + (X_1'X_1)^{-1} X_1' (\mathbf{x}_2\gamma + \mathbf{u})$ erwartungstreu?

Wegen $E[\mathbf{u}|X_1, \mathbf{x}_2] = \mathbf{0}$ folgt

$$E[\tilde{\boldsymbol{\beta}}|X_1, \mathbf{x}_2] = \boldsymbol{\beta} + (X_1'X_1)^{-1} \boxed{X_1'\mathbf{x}_2}\gamma$$

Der Fehler, der durch den ausgelassenen Regressor $\mathbf{x}_2$ entsteht ist also genau dann klein, wenn

- keine oder nur wenig Korrelation zwischen den bestehenden Regressoren in X_1 und dem ausgelassenen Regressor $\mathbf{x}_2$ besteht, also $\boxed{X_1'\mathbf{x}_2}$ klein ist, oder
- der wahre Einfluss von $\mathbf{x}_2$ auf $\mathbf{y}$ klein ist, also wenn γ klein ist.

Der Ausdruck $(X_1'X_1)^{-1}X_1'\mathbf{x}_2$ ist übrigens der OLS-Schätzer für $\boldsymbol{\delta}$ im Modell

$$\mathbf{x}_2 = X_1\boldsymbol{\delta} + \mathbf{v}$$

Ausgelassene Regressoren

Falls die Regressormatrix X_1 nur aus einem echten Regressor besteht, falls also $X_1 = (\boldsymbol{\iota}, \mathbf{x}_1)$, können wir das Vorzeichen des *omitted variable bias* $(X_1'X_1)^{-1} X_1'\mathbf{x}_2\gamma$ für den Schätzer $\tilde{\beta}_1$ wie folgt kategorisieren:

	$Cov(\mathbf{x}_1, \mathbf{x}_2) > 0$	$Cov(\mathbf{x}_1, \mathbf{x}_2) = 0$	$Cov(\mathbf{x}_1, \mathbf{x}_2) < 0$
$\gamma > 0$	positive Verzerrung	keine Verzerrung	negative Verzerrung
$\gamma = 0$	keine Verzerrung	keine Verzerrung	keine Verzerrung
$\gamma < 0$	negative Verzerrung	keine Verzerrung	positive Verzerrung

Die bedingte Varianz des OLS-Schätzers

Zufallsvektor $y \in \mathbb{R}^n$

$$\Sigma(y) = E\Big[\underbrace{\underset{n\times 1}{(y - E[y])}\,\underset{1\times n}{(y - E[y])'}}_{n\times n}\Big]$$

$$\begin{aligned}
\Sigma(A\cdot y + b) &= E\Big[\big(A\cdot y + b - E[A\cdot y + b]\big)\big(A\cdot y + b - E[A\cdot y + b]\big)'\Big] \\
&= E\Big[\big(A\cdot y + \cancel{b} - AE[y] - \cancel{b}\big)\big(A\cdot y + \cancel{b} - AE[y] - \cancel{b}\big)'\Big] \\
&= E\Big[\big(A\cdot y - AE[y]\big)\big(A\cdot y - AE[y]\big)'\Big] \\
&= E\Big[A\,(y - E[y])\big(A\,(y - E[y])\big)'\Big] \\
&= E\Big[A\,(y - E[y])(y - E[y])'A'\Big]
\end{aligned}$$

$$= A \underbrace{E[(y - E[y])(y - E[y])']}_{\Sigma(y)} \cdot A'$$

Also $\boxed{\Sigma(A \cdot y + b) = A \, \Sigma(y) \, A'}$

Die bedingte Varianz des OLS-Schätzers

Um die bedingte Varianz des OLS-Schätzers herzuleiten, müssen wir die bedingte Verteilung des zufälligen Störterms **u** und damit auch die bedingte Verteilung der erklärten Variablen **y** genauer beschreiben.

Wir hatten bereits gezeigt, dass aus $E[\mathbf{u}|X] = E[\mathbf{u}]$ (strikt exogene Regressoren) $Cov(x_{ij}, u_i) = 0$ für alle $j = 0, 1, \ldots, k$ folgt.

Zusätzlich fordern wir nun:

MLR 5 **Homoskedastizität & serielle Unkorreliertheit**

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

Rechenregeln für Varianz-Kovarianz-Matrizen

Seien $\mathbf{c}, \mathbf{d} \in \mathbb{R}^p$ zufällige Vektoren.

Es seien zusätzlich nicht-stochastisch:

- Matrizen: $A_1, A_2 \in \mathbb{R}^{o \times p}$
- Vektoren: $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^o$

Dann gilt:

$$Cov(A_1\mathbf{c} + \mathbf{b}_1, A_2\mathbf{d} + \mathbf{b}_2) = A_1 Cov(\mathbf{c}, \mathbf{d}) A_2'$$

und

$$\Sigma(A_1\mathbf{c} + \mathbf{b}_1) = A_1 \Sigma(\mathbf{c}) A_1'$$

$$\text{MLR1, 3, 4}: \quad \hat{\beta} = (X'X)^{-1}X'y$$

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}(\underbrace{(X'X)^{-1}X'}_{A}\, y)$$

$$= \underbrace{(X'X)^{-1}X'}_{A}\ \mathrm{Var}(y)\ \underbrace{((X'X)^{-1}X')'}_{A'}$$

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

$$= X\,(X'X'')^{-1} = \underbrace{X(X'X)^{-1}}_{A'}$$

- $\mathrm{Var}(y) \overset{\text{MLR1}}{=} \mathrm{Var}(X\beta + u) = \mathrm{Var}(u) \overset{\text{MLR5}}{=} \sigma^2 I$

$$\Rightarrow \mathrm{Var}(\hat{\beta}) = \underbrace{(X'X)^{-1}X'}_{A}\ \sigma^2 I \cdot \underbrace{X(X'X)^{-1}}_{A'} = \sigma^2 (X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I}$$

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

Im Vergleich: $\check{\beta} = A \cdot y$ (mit $A \cdot X = I$)

$$\Sigma(\check{\beta}) = \Sigma(A \cdot y) = A \cdot \Sigma(y) \cdot A' = A \sigma^2 I \cdot A'$$

$$= \sigma^2 A \cdot A'$$

Gauß-Markov:

$$\Sigma(\hat{\beta}) = \sigma^2 (X'X)^{-1} \text{ "}\leq\text{" } \sigma^2 A \cdot A' = \Sigma(\check{\beta})$$

für alle A mit $A \cdot X = I$

Die bedingte Varianz-Kovarianz-Matrix von $\hat{\boldsymbol{\beta}}$

Für $\mathbf{y}$ gilt:

$$
\begin{aligned}
\mathrm{Var}(\mathbf{y}|X) &= \mathrm{Var}(X\boldsymbol{\beta} + \mathbf{u}|X) && \text{(MLR 1: } \mathbf{y} = X\boldsymbol{\beta} + \mathbf{u} \;\&\; X, \boldsymbol{\beta} \text{ fest)} \\
&= \mathrm{Var}(\mathbf{u}|X) = \sigma^2 I_n && \text{(MLR 5)}
\end{aligned}
$$

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

$$
\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}|X) &= \mathrm{Var}((X'X)^{-1}X'\mathbf{y}|X) \\
&= (X'X)^{-1}X'\mathrm{Var}(\mathbf{y}|X)\left((X'X)^{-1}X'\right)'
\end{aligned}
$$

Wegen $(AB)' = B'A'$ und $\left((X'X)^{-1}\right)' = (X'X)^{-1}$ (Symmetrie) gilt:

$$
\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}|X) &\overset{\text{MLR 5}}{=} (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}
\end{aligned}
$$

Komponenten der OLS-Varianz-Kovarianz-Matrix

Man kann zeigen, dass die Varianz eines einzelnen Schätzers $\hat{\beta}_j$, also die (j, j)-Komponente von $\sigma^2 \left(X'X\right)^{-1}$, für $j = 1, \ldots, \cancel{j}\, k$ gegeben ist durch:

$$Var(\hat{\beta}_j | X) = \sigma^2 \cdot \frac{1}{n} \cdot \frac{1}{\overline{x_j x_j} - \bar{x}_j^2} \cdot \frac{1}{1 - R_j^2}$$

Hierbei haben die Komponenten folgende Bedeutung:

- σ^2: Varianz des unbeobachtbaren und homoskedastischen Störterms $\mathbf{u}$
- n: Stichprobengröße
- $\overline{x_j x_j} - \bar{x}_j^2$: Stichprobenvarianz des j-ten Regressors
- R_j^2: Bestimmtheitsmaß der Regression von $\mathbf{x}_j$ auf alle anderen Regressoren

Komponenten der OLS-Varianz-Kovarianz-Matrix

Die Varianz von $\hat{\beta}_j$ ist also klein, falls σ^2 und R_j^2 klein sind und falls n und $\overline{x_j x_j} - \bar{x}_j^2$ groß sind.

$R_j^2 \in [0, 1)$ misst, wie gut der Regressor j durch die restlichen Regressoren erklärt wird.

Im besten Fall, $R_j^2 = 0$, wäre x_j unkorreliert mit den restlichen Regressoren.

Falls $R_j^2 = 1$, wäre x_j als Linearkombination der restlichen Regressoren darstellbar (was durch $MLR3$ ausgeschlossen wird).

$\rightarrow$ **Zielkonflikt**:

Weitere Regressoren erhöhen R^2, das Bestimmtheitsmaß für die Erklärung der Variation von **y**, und gleichzeitig erhöhen sie potentiell R_j^2 und schwächen damit die Präzision der einzelnen Schätzer.

Multikollinearität

Sind die Bestimmtheitsmaße R_j^2 für die einzelnen Regressoren hoch, sprechen wir von **Multikollinearität**.

Dies tritt auf, wenn zwei oder mehr Regressoren eine betragsmäßig hohe Stichprobenkovarianz im Verhältnis zu ihrer Stichprobenvarianz aufweisen.

Ein Maß für Multikollinearität ist der „variance inflation factor“, VIF:

$$VIF_j = \frac{1}{1 - R_j^2}$$

Je größer VIF_j, desto größer die Multikollinearität.

Daumenregel: VIF_j sollte nicht größer als 10 sein bzw. R_j^2 sollte nicht größer als 0,9 sein.

Ein Beispiel für Multikollinearität

$$avgscore = \beta_0 + \beta_1 teacherexp + \beta_2 matexp + \beta_3 othexp + \ldots$$

avgscore: standartisierter Notenschnitt der Schule

teachexp: Ausgaben für Lehrer:innen

matexp: Ausgaben für Lehrmaterialien

othexp: andere Ausgaben

Die Ausgaben-Regressoren werden stark korreliert sein, da eine Schule mit vielen Ressourcen hohe Ausgaben in allen Bereichen haben wird.

Die einzelnen Schätzer werden stark streuen.

Omitted Variable Bias vs. Multikollinearität

Sei das wahre Modell gegeben durch:

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \mathbf{u}$$

Geschätztes Modell 1: $\hat{\boldsymbol{y}} = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{x}_1 + \hat{\beta}_2 \mathbf{x}_2$

$$\begin{aligned} \Rightarrow E[\hat{\beta}_1 | \mathbf{x}_1, \mathbf{x}_2] &= \beta_1, \\ Var(\hat{\beta}_1 | \mathbf{x}_1, \mathbf{x}_2) &= \sigma^2 / \left(n(\overline{x_1 x_1} - \bar{x}_1^2)(1 - R_1^2) \right) \end{aligned}$$

Geschätztes Modell 2: $\tilde{\boldsymbol{y}} = \tilde{\beta}_0 + \tilde{\beta}_1 \mathbf{x}_1$

$$\begin{aligned} \Rightarrow \; E[\tilde{\beta}_1 | \mathbf{x}_1, \mathbf{x}_2] &= \beta_1 + \frac{\overline{x_1 x_2} - \bar{x}_1 \bar{x}_2}{\overline{x_1 x_1} - \bar{x}_1^2} \beta_2 \neq \beta_1, \\ Var(\tilde{\beta}_1 | \mathbf{x}_1, \mathbf{x}_2) &= \sigma^2 / \left(n(\overline{x_1 x_1} - \bar{x}_1^2) \right) < Var(\hat{\beta}_1 | \mathbf{x}_1, \mathbf{x}_2) \end{aligned}$$

$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$ ist BLUE (Gauss-Markov-Theorem)

BLUE: best linear unbiased estimator

$\hat{\boldsymbol{\beta}}$ ist ein Schätzer (estimator) ✓

$\hat{\boldsymbol{\beta}}$ ist linear (in $\mathbf{y}$) ✓

$\hat{\boldsymbol{\beta}}$ ist unverzerrt (unbiased) ✓

$\hat{\boldsymbol{\beta}}$ ist varianzminimal (best) ?

Varianzminimal bedeutet, dass die bedingte Varianz-Kovarianz-Matrix $\Sigma\!\!\!\!\!\!\;\,|\,(\hat{\boldsymbol{\beta}}|X)$ „kleiner" ist, als die aller anderen linearen und unverzerrten Schätzer für $\boldsymbol{\beta}$.

Was bedeutet „kleiner" für Matrizen?

Lineare Algebra: Positiv-(Semi)Definitheit

Definition:

Eine $(p \times p)$-Matrix A heißt **positiv-semidefinit (psd)**, falls

$$\mathbf{x}' A \mathbf{x} \geq 0 \;\forall\; \mathbf{x} \in \mathbb{R}^p$$

Die Matrix A heißt **positiv definit (pd)**, falls $\mathbf{x}' A \mathbf{x} > 0 \;\forall\; \mathbf{x} \neq \mathbf{0}$.

Ist die Matrix zusätzlich **symmetrisch**, lauten die Abkürzungen **psds** bzw. **pds**.

Varianz-Kovarianz-Matrizen sind grundsätzlich psds

Sei $\mathbf{z} \in \mathbb{R}^n$ ein Zufallsvektor und sei $\mathbf{x} \in \mathbb{R}$ ein beliebiger nicht-stochastischer Vektor.

Dann gilt:

$$\mathbf{x}'\Sigma(\mathbf{z})\mathbf{x} = \Sigma(\underbrace{\mathbf{x}'\mathbf{z}}_{1\times 1}) = Var(\mathbf{x}'\mathbf{z}) \geq 0$$

Symmetrie folgt aus $Cov(z_i, z_j) = Cov(z_j, z_i) \; \forall \; i, j$.

Löwner Ordnung für Matrizen

Seien A und B zwei quadratische Matrizen der Ordnung $p \times p$.

Es gilt $A \geq B$, falls $A - B$ psd ist, falls also gilt:

$$\mathbf{x}'\,(A - B)\,\mathbf{x} \geq 0 \Leftrightarrow \mathbf{x}' A \mathbf{x} \geq \mathbf{x}' B \mathbf{x} \;\forall\; \mathbf{x} \in \mathbb{R}^p$$

Diese Ordnung ist partiell, da für manche Paare C und D von Matrizen weder $C - D$ psds, noch $D - C$ psds gilt.

Mit dieser Ordnung bestimmen wir die Eigenschaft varianzminimal (best).

Beweisstruktur BLUE

Es bleibt zu Zeigen, dass der kleinste Quadrate Schätzer varianzminimal im Sinne der Löwner-Ordnung für seine Varianz-Kovarianz-Matrix ist („best“, **B**) gegenüber allen Schätzern, welche

- L linear in den Daten $\mathbf{y}$ sind und $\rightsquigarrow \tilde{\beta} = A \cdot y$
- U unverzerrt die unbekannten Parameter $\boldsymbol{\beta}$ $\rightsquigarrow A \cdot X = I$
- E schätzen (to estimate).

$\Sigma(\tilde{\beta}) = \sigma^2 \cdot A \cdot A' \geq \sigma^2 (X'X)^{-1} = \Sigma(\hat{\beta})$

Wir heben uns den Beweis aber auf, bis wir allgemeinere Varianz-Kovarianz-Matrizen des Störterms $\mathbf{u}$ zulassen.

MLR 5

Bisher: $\Sigma(\mathbf{u}) = \sigma^2 I_n$ (Homoskedastizität) $\rightarrow$ OLS

Später: $\Sigma(\mathbf{u}) = \Omega$, beliebig $n \times n$ pds $\rightarrow$ GLS

Schätzung der Varianz der Störterme

Behauptung:

Unter den Annahmen $MLR1$, $MLR3$, $MLR4$ und $MLR5$ ist der Schätzer

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} = \frac{1}{n-k-1}\overbrace{\sum_{i=1}^{n}\hat{u}_i^2}^{SSR}$$

erwartungstreu für σ^2:

$$E[\hat{\sigma}^2] = \sigma^2 \ \forall \ \sigma^2 > 0$$

Bemerkung:
Wieder reduzieren wir n, die Zahl der Modellgleichungen, um $k+1$, die Zahl der Definitionsgleichungen der OLS-Schätzer, da wir die OLS-Schätzer benutzen um die Residuen $\hat{\boldsymbol{u}}$ zu berechnen.

$$\hat{u} = \underbrace{[I - X(X'X)^{-1}X']}_{P} y$$

$$\hat{\sigma}^2 = \frac{1}{n-k-1} \cdot \hat{u}' \cdot \hat{u} = \frac{1}{n-k-1} (Py)'(Py)$$

$$= \frac{1}{n-k-1} y' \cdot P' \cdot P \cdot y$$

zu Zeigen $E[\hat{\sigma}^2] = \sigma^2$

$$= E\left[\frac{1}{n-k-1} y' P' P y\right] = \sigma^2$$

$$\Leftrightarrow \underline{E[y' P' P y] = \sigma^2 (n-k-1)}$$

Eigenschaften von $P = I - X(X'X)^{-1}X'$

- $\boxed{P' = P}$ (symmetrie)

$$\left(I - X(X'X)^{-1}X'\right)' = I' - \left(X(X'X)^{-1}X'\right)'$$
$$= I - (X')'\left((X'X)^{-1}\right)'X'$$
$$= I - X(X'X)^{-1}\cdot X' = P \checkmark$$

- $\boxed{P \cdot P = P}$ (Idempotenz)

$$= [I - X(X'X)^{-1}X'][I - X(X'X)^{-1}X']$$
$$= I \cdot I - I X(X'X)^{-1}X' - X(X'X)^{-1}X' \cdot I$$
$$+ X(X'X)^{-1}\cancel{X'X}\cancel{(X'X)^{-1}}X'$$
$$= I - 2X(X'X)^{-1}X' + X(X'X)^{-1}X' = P \checkmark$$

∘ $P \cdot X = \mathcal{N}$ (Nullmatrix) $\Rightarrow$ $X' \cdot P = \mathcal{N}$

$$[I - X(X'X)^{-1}X']X = I \cdot X - X\underline{(X'X)^{-1}}\,\underline{X' \cdot X}$$

$$= X - X = \mathcal{N} \checkmark$$

$$(P \cdot X)' = X' \cdot P' = X' \cdot P = \mathcal{N}' = \mathcal{N}$$

$$E[\hat{u}'\hat{u} \mid X] = E[y' \underbrace{P'}_{=P} P y \mid X]$$

$$= E[y' \cdot \underbrace{P \cdot P}_{=P} \cdot y \mid X]$$

$$= E[y' \cdot P \cdot y \mid X]$$

MLR 1 $\quad y = X\beta + u$

$$= E\left[(X\beta + u)' \cdot P \cdot (X\beta + u) \mid X\right]$$

$$= E\left[(X\beta)' \cdot P (X\beta + u) + u' \underbrace{P X}_{=N} \beta + u' P u \mid X\right]$$

$$= E\left[\beta' \underbrace{X' P}_{N} (X\beta + u) + u' P u \mid X\right] = E\left[\underbrace{\underset{1\times n}{u'} \underset{n\times n}{P} \underset{n\times 1}{u}}_{1\times 1} \mid X\right]$$

trace

Die Spur einer Matrix A, $tr(A)$, ist die Summe aller Komponenten auf der Diagonalen von A.

Es gilt
- $tr(E[A]) = E[tr(A)]$
- $tr(ABC) = tr(CAB) = tr(BCA)$

Da $u'Pu \in \mathbb{R} \Rightarrow$

$$u'Pu = tr(u'Pu) = tr(uu'P)$$

$$\Rightarrow E[u'Pu \mid X] = E[tr(u'Pu) \mid X]$$

$$= E[tr(uu'P) \mid X]$$

$$= tr\left(E[uu'P \mid X]\right)$$

$$= tr\left(E[uu' \mid X] \cdot P\right)$$

$$= tr\left(\underbrace{E[uu']}_{\Sigma(u) = \sigma^2 \cdot I} \cdot P\right)$$

$$= tr(\sigma^2 P) = \sigma^2 \cdot tr(P)$$

$$\Rightarrow E[\hat{u}'\hat{u}] = \sigma^2 \cdot tr(P)$$

$$= \sigma^2 \cdot tr(I_n - X(X'X)^{-1}X')$$

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

$$= \sigma^2 (n - tr(\underbrace{X'X(X'X)^{-1}}_{I_{k+1}}))$$

$$= \sigma^2 (n - (k+1)) \qquad \blacksquare$$

$$E[\hat{\sigma}^2] = \sigma^2$$

Vorüberlegungen

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

Die Residuen $\hat{\boldsymbol{u}}$ sind eine Projektion von $\mathbf{y}$ auf das orthogonale Komplement des Spaltenraumes der Regressormatrix X:

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

wobei

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

die idempotent symmetrische Projektionsmatrix ist. Es gilt:

$$\left[I - X(X'X)^{-1}X'\right]X = N$$

(Jede Spalte von X wird auf den Ursprung $\mathbf{0}$ projiziert.)

Beweis $E[\hat{\sigma}^2] = \sigma^2$

Mit

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

gilt

$$\begin{aligned}
(n-k-1)\hat{\sigma}^2 &= (X\boldsymbol{\beta} + \mathbf{u})' \left[I - X(X'X)^{-1}X'\right](X\boldsymbol{\beta} + \mathbf{u}) \\
&= (X\boldsymbol{\beta})' \left[I - X(X'X)^{-1}X'\right] X\boldsymbol{\beta} && (= N) \\
&+ \mathbf{u}' \left[I - X(X'X)^{-1}X'\right] X\boldsymbol{\beta} && (= N) \\
&+ (X\boldsymbol{\beta})' \left[I - X(X'X)^{-1}X'\right] \mathbf{u} && (= N) \\
&+ \mathbf{u}' \left[I - X(X'X)^{-1}X'\right] \mathbf{u}
\end{aligned}$$

Hilfssatz (zyklische Vertauschungen):

Für drei Matrizen A, B, C, für welche ABC und CAB existieren, gilt:

$$tr(ABC) = tr(CAB)$$

Da $\underbrace{\mathbf{u}'}_{1\times n}\underbrace{\left[I - X(X'X)^{-1}X'\right]}_{n\times n}\underbrace{\mathbf{u}}_{n\times 1}$ ein Skalar ist, gilt:

$$\mathbf{u}'\left[I - X(X'X)^{-1}X'\right]\mathbf{u} = tr(\mathbf{u}'\left[I - X(X'X)^{-1}X'\right]\mathbf{u})$$

Beweis $E[\hat{\sigma}^2] = \sigma^2$ (Forts.)

Damit gilt

$$
\begin{aligned}
E[\mathbf{u}'\left[I - X(X'X)^{-1}X'\right]\mathbf{u}|X] &= E[tr(\mathbf{u}'\left[I - X(X'X)^{-1}X'\right]\mathbf{u})|X] \\
&= E[tr(\mathbf{u}\mathbf{u}'\left[I - X(X'X)^{-1}X'\right])|X] \\
&= tr\left(E[\mathbf{u}\mathbf{u}'\left[I - X(X'X)^{-1}X'\right]|X]\right) \\
&= tr\left(E[\mathbf{u}\mathbf{u}'|X]\left[I - X(X'X)^{-1}X'\right]\right) \\
&= \sigma^2 tr(\left[I - X(X'X)^{-1}X'\right]) \\
&= \sigma^2\left(tr(I) - tr\left(X(X'X)^{-1}X'\right)\right) \\
&= \sigma^2\left(n - tr\left(X'X(X'X)^{-1}\right)\right) \\
&= \sigma^2\left(n - (k+1)\right)
\end{aligned}
$$

Daraus folgt

$$
E[\hat{\sigma}^2|X] = \frac{E[\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}|X]}{n-k-1} = \frac{1}{n-k-1}E[\mathbf{u}'\left[I - X(X'X)^{-1}X'\right]\mathbf{u}|X] = \sigma^2
$$

Bemerkung zu $E[\hat{\sigma}^2] = \sigma^2$

$$\frac{1}{n-k-1} \underbrace{\hat{u}'\hat{u}}_{SSR}$$

Die Umkehrung der Behauptung lässt sich unter den getroffenen Annahmen ebenfalls zeigen:

Falls ein Schätzer

$$\tilde{\sigma}^2 = \mathbf{y}' D \mathbf{y}$$ "quadratische Form"

mit D psds erwartungstreu für σ^2 ist,

dann gilt $\tilde{\sigma}^2 = \hat{\sigma}^2$.

Wir heben uns den Beweis für diese Behauptung aber ebenfalls für den allgemeinen Fall

$$\Sigma(\mathbf{u}) = \Omega$$ "GLS"

auf.

($\Sigma(u) = \sigma^2 \cdot I$: "OLS")

Standard Error of Regression SER

Der Standardfehler der Regression ist die positive Wurzel des OLS-Schätzers $\hat{\sigma}^2$ für σ^2:

$$SER = \sqrt{\hat{\sigma}^2} = \hat{\sigma}$$

Hiermit können wir nun die Standardfehler der OLS-Schätzer der einzelnen Parameter berechnen:

$$se(\hat{\beta}_j) = \sqrt{\widehat{Var}(\hat{\beta}_j)} = \hat{\sigma}\sqrt{(X'X)^{-1}_{(j,j)}} = \hat{\sigma}/\sqrt{n(\overline{x_j x_j} - \bar{x}_j^2)(1 - R_j^2)}$$

$$Var(\hat{\beta}_j) = \sigma^2 \cdot \frac{1}{n} \frac{1}{\overline{x_j x_j} - \bar{x}_j \cdot \bar{x}_j} \cdot \frac{1}{1 - R_j^2}$$

Anwendung der Konzepte auf konkrete Daten

Wir wenden das bisher Gelernte nun auf einen konkreten Datensatz an:

wage1.xls aus dem Datensatz von Wooldridge.

1. Öffne die Datei wage1.xls mit dem Programm Gretl.
2. Benenne falls gewünscht die Variablen gemäß WAGE1_description.txt um.
3. Erstelle ein KQ-Modell (untere Leiste „$\hat{\beta}$“).
   Wähle hierbei v22/lwage als abhängige Variable und v2/educ und v3/exper als Regressoren.

Der erzeugte Output sollte dann der Tabelle der nächsten Folie entsprechen:

Anwendung auf *wage1.xls*

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

$H_0: \beta_0 = 0$

$\frac{\hat\beta_0 - 0}{se(\hat\beta_0)}$

$$\underbrace{\ln(wage)}_{y} = \beta_0 + \beta_1 \underbrace{educ}_{x_1} + \beta_2 \underbrace{exper}_{x_2} + \mathbf{u}$$

Stichprobengröße: $n = 526$

$\sqrt{\widehat{Var}(\hat\beta_0)}$

Chapt. 4

$\hat\beta_0$	Koeffizient	Std. Fehler	t-Quotient	p-Wert
const	0,216854	0,108595	1,997	0,0464
educ $\hat\beta_1$	0,0979356	0,00762240	12,85	0,0000
exper $\hat\beta_2$	0,0103469	0,00155514	6,653	0,0000

Mittel abhängige Var. $\bar{y}$	1,623268	Stdabw. abhängige Var.	0,531538
Summe quad. Residuen SSR $\hat u'\hat u$	111,3447	Stdfehler Regression $\hat\sigma$	0,461407
$R^2 = 1 - \frac{SSR}{SST}$	0,249343	Korrigiertes R^2	0,246473
Chpt. 4 $F(2, 523)$	86,86167	P-Wert(F) Chpt. 4	2,68e–33
Log-Likelihood	−338,0094	Akaike-Kriterium	682,0187
Schwarz-Kriterium	694,8146	Hannan–Quinn	687,0289

Zusammenfassung

- Matrixnotation, Regressormatrix
- OLS-Schätzer $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$
- OLS-Residuen $\hat{\boldsymbol{u}} = [I - X(X'X)^{-1}X']\mathbf{y}$
- Partitionierte Regression (Frisch-Waugh)
- Anpassungsgüte, Bestimmtheitsmaß R^2
- $MLR1$, $MLR3$, $MLR4$, BLUE
- Überspezifikation & Ausgelassene Regressoren
- Bedingte Varianz des OLS-Schätzers
- Multikollinearität
- $+MLR5$: $\hat{\sigma}^2$ erwartungstreu
- Standardfehler der Regression, SER