TECHNISCHE UNIVERSITÄT DRESDEN
Fakultät Wirtschaftswissenschaften

Dresdner Beiträge zu
Quantitativen Verfahren

Nr. 69/17

Einführung in die Ökonometrie

von

Stefan Huschens

Herausgeber:
Die Professoren der
Fachgruppe Quantitative Verfahren

ISSN 0945-4802

Einführung in die Ökonometrie

Stefan Huschens

Fassung[1] vom 1. Februar 2017

[1] Unter https://www.stefan-huschens.de/statistik/statistische-miszellen/ können Sie sich informieren, ob es eine aktuellere Fassung gibt, und diese gegebenenfalls herunterladen.

Inhaltsverzeichnis

Vorwort

Die Kapitel 1 bis 6 im ersten Teil dieses Skriptes beruhen auf einer Vorlesung Ökonometrie I, die zuletzt im WS 2001/02 gehalten wurde, die Kapitel 7 bis 16 beruhen auf einer Vorlesung Ökonometrie II, die zuletzt im SS 2006 gehalten wurde. Das achte Kapitel enthält eine komprimierte Zusammenfassung der Ergebnisse aus dem Teil Ökonometrie I.

Ökonometrie I

Kapitel 1

Vorbemerkungen zur Ökonometrie I

1-1

1.1 Gliederung

1. Vorbemerkungen zur Ökonometrie I
2. Das lineare Einfach-Regressionsmodell
3. Verteilungen von Stichprobenfunktionen
4. Konfidenzintervalle und Hypothesentests
5. Das multiple lineare Regressionsmodell
6. Hypothesentest im multiplen linearen Modell

1-2

1.2 Literatur

Einführende Literatur zur Ökonometrie:

- Auer, L. von: Ökonometrie. Eine Einführung, 3. Aufl., Springer: Berlin, Heidelberg, New York 2005.
- Gujarati, D. N: Basic Econometrics, 4. Aufl., McGraw-Hill: New York 2003.
- Gujarati, D. N: Basic Econometrics, 3. Aufl., McGraw-Hill: New York 1995.
- Judge, G., Hill, C., Griffits, W. E., Lütkepohl, H.. Lee, T.: Introduction to the Theory and Practice of Econometrics. 2. Aufl., Wiley: New York 1988.

1-3

Weiterführende Literatur zur Ökonometrie:

- Greene, W. H.: Econometric Analysis. 5. Aufl., Prentice Hall: Upper Saddle River 2003.
- Judge, G., Griffits, W. E., Hill, C., Lütkepohl, H., Lee, T.-C.: The Theory and Practice of Econometrics. 2. Aufl., Wiley: New York 1985.

Ergänzende Literatur zur Linearen Regression:

- Kapitel 7. Lineare Regression. In:
  Mosler, K., Schmid, F.: Wahrscheinlichkeitsrechnung und schließende Statistik, 2. Aufl., Springer: Berlin, Heidelberg, New York 2006.
- Kapitel 4.4. Deskriptive lineare Regression. In:
  Huschens, S.: Statistik im Grundstudium – Teil A: Deskriptive Statistik, Vorlesungsskript, 3. Aufl., 2007.
- Kapitel 4.4. Deskriptive lineare Regression. In:
  Huschens, S.: Statistik im Bachelorstudium, Vorlesungsskript, 1. Aufl., 2008.

1-4

1.3 Gegenstand der Ökonometrie

- **Ökonometrie** (*econometrics*): Zusammensetzung aus Ökonomie (*economics*) und Metrie (Messen)
- Andere ...metrien: **Psychometrie** (Psychologie), **Biometrie** (Landwirtschaft, Medizin), **Technometrie** (Technik), ...
- Parameterschätzung für Modelle aus der Wirtschaftstheorie
  - Wirtschaftstheoretische Modelle
  - Daten
  - Statistische Methodik
- Allgemeinere Auffassung von Ökonometrie: Statistische Analyse ökonomischer Daten (auch ohne Bezug zu ökonomischen Modellen)

1-5

1.4 Vorbemerkungen zur Notation

Allgemeine Notation:

- $\overset{\text{def}}{=}$ ist das **definitorische Gleichheitszeichen**. Die linke Seite der Gleichung wird durch die rechte Seite definiert.
- **Summenzeichen**

$$\sum_{i=1}^{n} x_i \overset{\text{def}}{=} x_1 + x_2 + \ldots + x_n$$

- $\mathbb{R}$ bezeichnet die **Menge der reellen Zahlen**.
- $\mathbb{N} \overset{\text{def}}{=} \{1, 2, \ldots\}$ bezeichnet die **Menge der natürlichen Zahlen**.

1-6

Zufallsvariablen und Verteilungen

- Bei statistisch-theoretischen Darstellungen ist es üblich, Zufallsvariablen und ihre Werte (Realisationen) durch Groß- und Kleinbuchstaben zu unterscheiden. Beispielsweise bezeichnet X eine Zufallsvariable und x eine Realisation.
- In vielen Anwendungsbereichen der Statistik, so auch in der Ökonometrie, wird diese Unterscheidung in der Notation häufig nicht vorgenommen. Dies vereinfacht teilweise die Notation, erfordert aber besondere Aufmerksamkeit, weil der Zusammenhang entscheidend ist.
- Beispielweise bezeichnet $\hat{\theta}$ eine konkreten **Schätzwert** für den Parameter θ, z. B. $\hat{\theta} = 0.15$, aber auch den **Schätzer**, d. h. die Zufallsvariable, für die sich z. B. der Erwartungswert $\mathbb{E}(\hat{\theta})$ berechnen läßt.

1-7

- Bezeichnungen für Zufallsvariablen
  - **Erwartungswert** einer Zufallsvariablen x:

$$\mathbb{E}(x)$$

  - **Varianz** einer Zufallsvariablen x:

$$\mathbb{V}(x)$$

  - **Kovarianz** von zwei Zufallsvariablen x und y:

$$\mathbb{C}\mathrm{ov}(x, y)$$

  - **Standardfehler** (engl.: *standard error*) bzw. Standardabweichung (engl.: *standard deviation*) eines Schätzers $\hat{\theta}$:

$$se(\hat{\theta}) = \sqrt{\mathbb{V}(\hat{\theta})}$$

1-8

- Spezielle Verteilungen:
  - **Normalverteilung**: $N(\mu, \sigma^2)$, wobei $\mu = \mathbb{E}(x)$ und $\sigma^2 = \mathbb{V}(x)$, falls $x \sim N(\mu, \sigma^2)$.
  - t-**Verteilung** mit n Freiheitsgraden: $t_n \quad (n \in \mathbb{N})$
  - χ^2-**Verteilung** mit n Freiheitsgraden: $\chi^2_n \quad (n \in \mathbb{N})$
  - F-**Verteilung** mit Parametern m und n: $F_{m,n} \quad (m, n \in \mathbb{N})$
- α-**Fraktil** (α-**Quantil**) einer t-Verteilung mit n Freiheitsgraden: $t_{n;\alpha}$. Falls $t \sim t_n$ gilt $P(t \leq t_{n;\alpha}) = \alpha$ für $0 < \alpha < 1$.

Stichprobenmaßzahlen 1-9

- Gegebene Beobachtungen: $(x_t, y_t), \quad t = 1, \ldots, T$
- Mittelwerte:

$$\bar{x} \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T} x_t, \quad \bar{y} \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T} y_t$$

- (empirische) Varianz und Standardabweichung:

$$s_x^2 \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2, \quad s_x \stackrel{\text{def}}{=} \sqrt{s_x^2}$$

  Alternative Notation: $s_{xx} = s_x^2$

- (empirische) Kovarianz:

$$s_{xy} \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})$$

1-10

- Stichprobenkorrelationskoeffizient (Pearsonscher Korrelationskoeffizient):

$$r_{xy} \stackrel{\text{def}}{=} \frac{s_{xy}}{s_x s_y} = \frac{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2 \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2}} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sum_{t=1}^{T}(y_t - \bar{y})^2}}$$

- Es gilt

$$-1 \leq r_{xy} \leq 1\,.$$

Alternative Notationen: 1-11

- Bei der Parameterschätzung werden auch

$$s_x^{*2} \stackrel{\text{def}}{=} \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \bar{x})^2 \quad \text{und} \quad s_{xy}^{*} \stackrel{\text{def}}{=} \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})$$

  verwendet.

- Ludwig von Auer verwendet in seinem Lehrbuch die Notation

$$S_{xx} \stackrel{\text{def}}{=} \sum_{t=1}^{T}(x_t - \bar{x})^2 \quad \text{und} \quad S_{xy} \stackrel{\text{def}}{=} \sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})\,.$$

- Es gilt

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} = \frac{s_{xy}^{*}}{\sqrt{s_x^{*2} s_y^{*2}}} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

Kapitel 2

Lineares Einfach-Regressionsmodell

2-1

2. Lineares Einfach-Regressionsmodell

2-2

2.1 Methode der kleinsten Quadrate

Bemerkung 2.1

1. Die Methode der kleinsten Quadrate wird zunächst als Verfahren der deskriptiven Statistik betrachtet.
2. An Beobachtungen
$$(x_t, y_t), \quad t = 1, \ldots, T$$
mit
$$\sum_{t=1}^{T} (x_t - \bar{x})^2 > 0 \tag{2.1}$$
soll eine Gerade $y = \beta_1 + \beta_2 x$ angepaßt werden.
3. Die Bedingung (2.1) ist genau dann erfüllt, wenn nicht alle x_i gleich sind.

2-3

Bemerkung 2.2

1. Die Schätzwerte für β_1 und β_2 nach der Methode der kleinsten Quadrate sind

$$\hat{\beta}_2 = \frac{\sum\limits_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum\limits_{t=1}^{T}(x_t - \bar{x})^2} \tag{2.2}$$

und

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x}\,. \tag{2.3}$$

2. Es gilt

$$\hat{\beta}_2 = \frac{s_{xy}}{s_x^2} = r_{xy}\frac{s_y}{s_x}\,.$$

2-4

Bemerkung 2.3

1. $(\hat{\beta}_1, \hat{\beta}_2)$ ist die Optimalstelle der Minimierungsaufgabe

$$Q(\beta_1, \beta_2) \stackrel{\text{def}}{=} \sum_{t=1}^{T}(y_t - \beta_1 - \beta_2 x_t)^2 \to \min \quad \text{bzgl.} \quad \beta_1 \text{ und } \beta_2$$

d. h.

$$\sum_{t=1}^{T}(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2 = \min_{\beta_1,\beta_2}\sum_{t=1}^{T}(y_t - \beta_1 - \beta_2 x_t)^2$$

bzw.

$$Q(\hat{\beta}_1, \hat{\beta}_2) = \min_{\beta_1,\beta_2} Q(\beta_1, \beta_2)\,.$$

4. Die Optimalstelle $(\hat{\beta}_1, \hat{\beta}_2)$ erhält man durch Nullsetzen der partiellen Ableitungen der Funktion $Q(\beta_1, \beta_2)$ nach β_1 und β_2.

2-5

$$\begin{aligned}
\frac{\partial Q(\beta_1, \beta_2)}{\partial \beta_1} &= \frac{\partial \sum_{t=1}^{T}(y_t - \beta_1 - \beta_2 x_t)^2}{\partial \beta_1} = \sum_{t=1}^{T}\frac{\partial (y_t - \beta_1 - \beta_2 x_t)^2}{\partial \beta_1} \\
&= -2\sum_{t=1}^{T}(y_t - \beta_1 - \beta_2 x_t)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Q(\beta_1, \beta_2)}{\partial \beta_2} &= \frac{\partial \sum_{t=1}^{T}(y_t - \beta_1 - \beta_2 x_t)^2}{\partial \beta_2} = \sum_{t=1}^{T}\frac{\partial (y_t - \beta_1 - \beta_2 x_t)^2}{\partial \beta_2} \\
&= -2\sum_{t=1}^{T}x_t(y_t - \beta_1 - \beta_2 x_t)
\end{aligned}$$

Aus den beiden Gleichungen

2-6

$$
\begin{aligned}
-2\sum_{t=1}^{T}(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) &= 0 && (2.4)\\
-2\sum_{t=1}^{T}x_t(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) &= 0, && (2.5)
\end{aligned}
$$

die auch Normalgleichungen heißen, erhält man zusammen mit (2.1) die Formeln (2.2) und (2.3) für $\hat{\beta}_1$ und $\hat{\beta}_2$.

2-7

Bemerkung 2.4

1. Die **geschätzte Regressionsgerade** ist

$$y = \hat{\beta}_1 + \hat{\beta}_2 x\,.$$

2. Die **y-Werte auf der geschätzten Regressionsgeraden** sind

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_t, \quad t = 1, \ldots, T\,.$$

3. Die **Residuen** (geschätzten Fehler) der Regression sind

$$\hat{u}_t = y_t - \hat{y}_t, \quad t = 1, \ldots, T\,.$$

2-8

Bemerkung 2.5

1. Bei einer Regressionsgeraden der Form

$$y = \beta_1 + \beta_2 x$$

spricht man von einer linearen Regression **mit Absolutglied** oder von einer **inhomogenen** linearen Regression.

2. Bei einer Regressionsgeraden der Form

$$y = \beta x$$

spricht man von einer linearen Regression **ohne Absolutglied** oder von einer **homogenen** linearen Regression.

2-9

Bemerkung 2.6
Im Fall einer homogenen linearen Regression führt die Minimierung von

$$Q(\beta) \stackrel{\text{def}}{=} \sum_{t=1}^{T}(y_t - \beta x_t)^2$$

bzgl. β zu der Normalgleichung

$$\sum_{t=1}^{T} x_t y_t = \hat{\beta} \sum_{t=1}^{T} x_t^2$$

und im Fall $\sum_{t=1}^{T} x_t^2 > 0$ zu

$$\hat{\beta} = \frac{\sum\limits_{t=1}^{T} x_t y_t}{\sum\limits_{t=1}^{T} x_t^2}$$

als KQ-Schätzwert für β.

2-10

2.2 Bestimmtheitsmaß und Streuungszerlegung

In diesem Abschnitt ist der Fall einer Regression mit Absolutglied vorausgesetzt.
Aus Gleichung (2.4) folgt

$$\sum_{t=1}^{T} \hat{u}_t = 0$$

und daher

$$\bar{\hat{u}} = 0, \quad s_{\hat{u}}^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2, \quad \bar{\hat{y}} = \bar{y}$$

Je kleiner $s_{\hat{u}}^2$, desto besser ist die Anpassung der Regressionsgerade an die Beobachtungen.
Extremfall:
$s_{\hat{u}}^2 = 0$, d. h. $\hat{y}_t = y_t$ für $t = 1, \ldots, T$ und alle (x_t, y_t) liegen auf der Regressionsgeraden.

2-11

Bestimmtheitsmaß (Determinationskoeffizient):

$$\mathrm{R}^2 \stackrel{\text{def}}{=} \frac{\sum\limits_{t=1}^{T} (\hat{y}_t - \bar{y})^2}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2}$$

Es gilt:

$$0 \leq \mathrm{R}^2 \leq 1$$

Streuungszerlegung:

$$\sum_{t=1}^{T}(y_t - \bar{y})^2 = \sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2 + \sum_{t=1}^{T}(y_t - \hat{y}_t)^2$$

$$s_y^2 = s_{\hat{y}}^2 + s_{\hat{u}}^2$$

$$\mathrm{R}^2 = \frac{s_{\hat{y}}^2}{s_y^2} = 1 - \frac{s_{\hat{u}}^2}{s_y^2} = r_{xy}^2$$

2-12

2.3 Annahmen im Regressionsmodell

1. In der **Regressionsfunktion der Grundgesamtheit** (PRF, *population regression function*):
$$y = \beta_1 + \beta_2 x$$
sind β_1 und β_2 feste und unbekannte Parameter.

2. Die Abweichungen von der Regressionsgeraden werden stochastisch modelliert durch den Ansatz
$$y_t = \beta_1 + \beta_2 x_t + u_t,$$
wobei u_t eine Zufallsvariable ist, z. B. $u_t \sim N(0, \sigma^2)$ mit $\mathbb{E}(u_t) = 0$, $\mathbb{V}(u_t) = \sigma^2$.

2-13

3. Für festes x_t und feste Parameter β_1 und β_2 ist y_t eine Zufallsvariable mit
$$\mathbb{E}(y_t) = \beta_1 + \beta_2 x_t$$
und
$$\mathbb{V}(y_t) = \mathbb{V}(u_t) = \sigma^2 .$$

4. Damit sind $\hat{\beta}_1$ und $\hat{\beta}_2$ als Funktionen der Zufallsvariablen $y_1, \ldots, y_T$ ebenfalls Zufallsvariablen (Schätzer) und die Eigenschaften dieser Zufallsvariablen hängen von den Eigenschaften der Zufallsvariablen u_t ab.

2-14

Bemerkung 2.7 (Annahmen über das Modell)
Das Regressionsmodell ist korrekt spezifiziert, d. h. die folgenden beiden Annahmen sind erfüllt.

M1 Alle relevanten erklärenden Variablen (Regressoren) sind im Modell enthalten.

M2 Die funktionale Form des Modells ist korrekt.

2-15

Bemerkung 2.8 (Annahmen über die Regressoren)

X1 Die x_t sind nicht zufällig.

X2 $\sum_{t=1}^{T}(x_t - \bar{x})^2 > 0$, d. h. nicht alle x_t sind identisch.

2-16

Bemerkung 2.9 (Annahmen über die Fehlerterme)

U1 Die Fehlervariablen (Störterme) u_t haben den Erwartungswert Null,

$$\mathbb{E}(u_t) = 0, \quad t = 1, \ldots, T\,.$$

U2 Die Fehlervariablen u_t sind nicht autokorreliert,

$$\mathbb{C}\text{ov}(u_t, u_s) = 0, \quad t, s = 1, \ldots, T, \quad t \neq s\,.$$

U3 Homoskedastie (keine Heteroskedastie):

$$\mathbb{V}(u_t) = \sigma^2 \quad \forall t$$

U4 Die Fehlervariablen sind normalverteilt.

U5 Die Fehlervariablen sind unabhängig und identisch $N(0, \sigma^2)$-verteilt.

2-17

2.4 Eigenschaften der KQ-Schätzer

Bemerkung 2.10 (Erwartungstreue)
Aus den Annahmen M1, M2, X1, X2 und U1 folgt, daß die KQ-Schätzer $\hat{\beta}_1$ und $\hat{\beta}_2$ für die Parameter β_1 und β_2 **erwartungstreu** (unverzerrt) sind, d. h. es gilt

$$\mathbb{E}(\hat{\beta}_1) = \beta_1 \quad \text{und} \quad \mathbb{E}(\hat{\beta}_2) = \beta_2\,. \tag{2.6}$$

2-18

Bemerkung 2.11 (Varianzen und Kovarianz der KQ-Schätzer)
Mit den Annahmen M1, M2, X1, X2, U1, U2 und U3 ergibt sich

$$\mathbb{V}(\hat\beta_1) = \frac{\sum\limits_{t=1}^{T} x_t^2}{T\sum\limits_{t=1}^{T}(x_t-\bar x)^2}\sigma^2\,, \quad \mathbb{V}(\hat\beta_2) = \frac{1}{\sum\limits_{t=1}^{T}(x_t-\bar x)^2}\,\sigma^2 \tag{2.7}$$

und

$$\mathbb{C}\mathrm{ov}(\hat\beta_1,\hat\beta_2) = -\frac{\bar x}{\sum\limits_{t=1}^{T}(x_t-\bar x)^2}\,\sigma^2\,.$$

2-19

Bemerkung 2.12 (Normalverteilung der KQ-Schätzer)
Unter den Annahmen M1, M2, X1, X2 und U5 sind die KQ-Schätzer $\hat\beta_1$ und $\hat\beta_2$ jeweils **normalverteilt**,

$$\hat\beta_i \sim N\left(\mathbb{E}(\hat\beta_i), \mathbb{V}(\hat\beta_i)\right), \quad i = 1, 2,$$

wobei die Erwartungswerte der Normalverteilungen durch (2.6) und die Varianzen der Normalverteilungen durch (2.7) gegeben sind.

2-20

Bemerkung 2.13 (Theorem von Gauß-Markov)

1. *Falls die Annahmen M1, M2, X1, X2, U1, U2, U3 und U4 erfüllt sind, sind die KQ-Schätzer $\hat\beta_i$ unter allen linearen erwartungstreuen Schätzern diejenigen mit der kleinsten Varianz.*

2. *Falls die Annahmen M1, M2, X1, X2 und U5 erfüllt sind, sind die KQ-Schätzer $\hat\beta_i$ unter allen erwartungstreuen Schätzern diejenigen mit der kleinsten Varianz.*

Bemerkung 2.14 (Zu den Annahmen)

1. Aus U5 folgen die Annahmen U1, U2, U3 und U4. Allerdings implizieren die Annahmen U1, U2, U3 und U4 zusammen nicht U5.

2. Die Normalverteilungsannahme für die Fehler kann über den den zentralen Grenzwertsatz der Statistik gerechtfertigt werden.

2-21

Bemerkung 2.15 (Schätzung der Varianzen)

1. Die in (2.7) angegebenen Varianzen $\mathbb{V}(\hat{\beta}_1)$ und $\mathbb{V}(\hat{\beta}_2)$ hängen von σ^2, der in der Regel unbekannten Varianz der Fehlerterme u_t, ab.
2. Die Varianz σ^2 der Fehlerterme wird durch

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \hat{u}_t^2}{T-2}.$$

aus den Residuen $\hat{u}_t = y_t - \hat{y}_t$ erwartungstreu geschätzt, d.h. für $\hat{\sigma}^2$ gilt

$$\mathbb{E}(\hat{\sigma}^2) = \sigma^2.$$

2-22

3. Die **Standardabweichungen** (theoretischen Standardfehler) von $\hat{\beta}_1$ und $\hat{\beta}_2$ sind

$$se(\hat{\beta}_1) = \sqrt{\mathbb{V}(\hat{\beta}_1)} = \sqrt{\frac{\sum_{t=1}^{T} x_t^2}{T \sum_{t=1}^{T} (x_t - \bar{x})^2}}\, \sigma$$

und

$$se(\hat{\beta}_2) = \sqrt{\mathbb{V}(\hat{\beta}_2)} = \frac{1}{\sqrt{\sum_{t=1}^{T} (x_t - \bar{x})^2}}\, \sigma.$$

2-23

4. Die Standardabweichungen werden durch die Standardfehler

$$\widehat{se}(\hat{\beta}_1) = \sqrt{\frac{\sum_{t=1}^{T} x_t^2}{T \sum_{t=1}^{T} (x_t - \bar{x})^2}}\, \hat{\sigma} \quad \text{und} \quad \widehat{se}(\hat{\beta}_2) = \frac{1}{\sqrt{\sum_{t=1}^{T} (x_t - \bar{x})^2}}\, \hat{\sigma}$$

geschätzt.

2.5 Ergänzungen

Bemerkung 2.a (Herleitung von $\mathbb{E}(\hat{\beta}_2) = \beta_2$, Gleichung (2.6))

$$\begin{aligned}
\mathbb{E}(\hat{\beta}_2) &= \mathbb{E}\left(\frac{\sum_{t=1}^{T}(x_t-\bar{x})(y_t-\bar{y})}{\sum_{t=1}^{T}(x_t-\bar{x})^2}\right) \\
&= \frac{\sum_{t=1}^{T}(x_t-\bar{x})\mathbb{E}(y_t-\bar{y})}{\sum_{t=1}^{T}(x_t-\bar{x})^2} \\
&= \frac{\sum_{t=1}^{T}(x_t-\bar{x})\mathbb{E}\left(\beta_2(x_t-\bar{x})+u_t-\bar{u}\right)}{\sum_{t=1}^{T}(x_t-\bar{x})^2} \\
&= \frac{\sum_{t=1}^{T}(x_t-\bar{x})(\beta_2(x_t-\bar{x})+E(u_t-\bar{u}))}{\sum_{t=1}^{T}(x_t-\bar{x})^2} \\
&= \beta_2\frac{\sum_{t=1}^{T}(x_t-\bar{x})(x_t-\bar{x})}{\sum_{t=1}^{T}(x_t-\bar{x})^2} \\
&= \beta_2
\end{aligned}$$

Bemerkung 2.b (Herleitung von $\mathbb{E}(\hat{\beta}_1) = \beta_1$, Gleichung (2.6))

$$\begin{aligned}
\mathbb{E}(\hat{\beta}_1) &= \mathbb{E}(\bar{y}-\hat{\beta}_2\bar{x}) \\
&= \mathbb{E}(\bar{y})-\mathbb{E}(\hat{\beta}_2)\bar{x} \\
&= \mathbb{E}(\beta_1+\beta_2\bar{x}+\bar{u})-\beta_2\bar{x}) \\
&= \beta_1+\bar{x}\beta_2-\beta_2\bar{x} \\
&= \beta_1
\end{aligned}$$

Bemerkung 2.c (Herleitung von $\mathbb{V}(\hat{\beta}_2)$, Gleichung (2.7))

$$\begin{aligned}
\sum_{t=1}^{T}(x_t-\bar{x})(y_t-\bar{y}) &= \sum_{t=1}^{T}(x_t-\bar{x})(\beta_2(x_t-\bar{x})+u_t-\bar{u}) \\
&= \beta_2\sum_{t=1}^{T}(x_t-\bar{x})^2+\sum_{t=1}^{T}(x_t-\bar{x})(u_t-\bar{u}) \\
&= \beta_2\sum_{t=1}^{T}(x_t-\bar{x})^2+\sum_{t=1}^{T}(x_t-\bar{x})u_t
\end{aligned}$$

impliziert

$$\hat{\beta}_2 = \beta_2 + \frac{\sum_{t=1}^{T}(x_t-\bar{x})u_t}{\sum_{t=1}^{T}(x_t-\bar{x})^2}$$

und somit

$$\mathbb{V}(\hat{\beta}_2) = \frac{\mathbb{V}\left(\sum_{t=1}^{T}(x_t-\bar{x})u_t\right)}{\left(\sum_{t=1}^{T}(x_t-\bar{x})^2\right)^2} = \frac{\sum_{t=1}^{T}(x_t-\bar{x})^2}{\left(\sum_{t=1}^{T}(x_t-\bar{x})^2\right)^2}\sigma^2 = \frac{1}{\sum_{t=1}^{T}(x_t-\bar{x})^2}\sigma^2\,.$$

Bemerkung 2.d (Herleitung von $\mathbb{V}(\hat{\beta}_1)$, Gleichung (2.7))
Mit $S_{xx} \overset{\text{def}}{=} \sum\limits_{t=1}^{T}(x_t - \bar{x})^2$ gilt

$$\begin{aligned}
\mathbb{V}(\hat{\beta}_1) &= \mathbb{V}\left(\bar{y} - \hat{\beta}_2\bar{x}\right) \\
&= \mathbb{V}\left(\beta_1 + \beta_2\bar{x} + \bar{u} - \beta_2\bar{x} - \frac{\sum\limits_{t=1}^{T}(x_t - \bar{x})u_t}{S_{xx}}\bar{x}\right) \\
&= \mathbb{V}\left(\bar{u} - \frac{\sum\limits_{t=1}^{T}(x_t - \bar{x})u_t}{S_{xx}}\bar{x}\right) \\
&= \mathbb{V}\left(\sum_{t=1}^{T}\left(\frac{1}{T} - \frac{x_t - \bar{x}}{S_{xx}}\bar{x}\right)u_t\right) \\
&= \sum_{t=1}^{T}\left(\frac{1}{T} - \frac{x_t - \bar{x}}{S_{xx}}\bar{x}\right)^2\sigma^2 .
\end{aligned}$$

Weitere Umformungen führen zu

$$\sum_{t=1}^{T}\left(\frac{1}{T} - \frac{x_t - \bar{x}}{S_{xx}}\bar{x}\right)^2 = \frac{\sum\limits_{t=1}^{T}x_t^2}{TS_{xx}} .$$

Bemerkung 2.e (Zu den Annahmen)

1. Wird U2 durch die stärkere Annahme

   U2* Die Fehlervariablen u_t sind stochastisch unabhängig

   ersetzt, dann implizieren U1, U2*, U3 und U4 gemeinsam U5.
2. Wird U4 durch die stärkere Annahme

   U4* Die Fehlervariablen u_t sind multivariat normalverteilt

   ersetzt, dann implizieren U1, U2, U3 und U4* gemeinsam U5.
3. Aus den Annahmen U2* und U4 folgt U4*.
4. Aus den Annahmen U2 und U4* folgt U4.
5. Es ist also äquivalent, entweder U2* und U4 oder U2 und U4* anzunehmen.
6. Die beiden Annahmen U2 und U4 sind allerdings schwächer als U2* und U4 bzw. U2 und U4*.

Bemerkung 2.f (Stochastische Regressoren)

- In einer allgemeineren Modellklasse werden abweichend von der Annahme X1 auch die Regressoren x_t stochastisch modelliert. Dies führt zu den sogenannten **Modellen mit stochastischen Regressoren**.
- Man benötigt dann zusätzliche Annahmen, z. B. über die stochastische Unabhängigkeit der x_t und der u_s.

Kapitel 3

Verteilungen von Stichprobenfunktionen

3-1

3. Verteilungen von Stichprobenfunktionen

3-2

3.1 Linearkombinationen normalverteilter Zufallsvariablen

- Die Zufallsvariablen $x_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, seien stochastisch unabhängig, dann gilt

$$\sum_{i=1}^{n} x_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

- Die Zufallsvariablen $x_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, seien stochastisch unabhängig, dann gilt

$$\sum_{i=1}^{n} c_i x_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right).$$

3-3

3.2 Chiquadratverteilung

- Die Zufallsvariablen $x_i \sim N(0,1)$, $i = 1,\ldots,n$, seien stochastisch unabhängig, dann folgt die Zufallsvariable

$$y \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_i^2$$

einer **Chiquadratverteilung mit** n **Freiheitsgraden**.

- Notation: $y \sim \chi_n^2$.
- Es gilt

$$\mathbb{E}(y) = n \quad \text{und} \quad \mathbb{V}(y) = 2n.$$

3-4

3.3 t-Verteilung

- Die Zufallsvariablen $x \sim N(0,1)$ und $y \sim \chi_n^2$ seien stochastisch unabhängig, dann folgt die Zufallsvariable

$$t \stackrel{\text{def}}{=} \frac{x}{\sqrt{y/n}}$$

der (Student'schen) t**-Verteilung mit** n **Freiheitsgraden**.

- Notation: $t \sim t_n$.
- Für $n \to \infty$ approximiert t_n die Standardnormalverteilung. Für viele Anwendungen ergibt sich für $n \geq 30$ eine befriedigende Approximation.

3-5

3.4 F-Verteilung

- Die Zufallsvariablen $x \sim \chi_m^2$ und $y \sim \chi_n^2$ seien stochastisch unabhängig, dann folgt die Zufallsvariable

$$F \stackrel{\text{def}}{=} \frac{x/m}{y/n}$$

der (Fisher'schen) F**-Verteilung mit** m **und** n **Freiheitsgraden**.

- Notation: $F \sim F_{m,n}$.
- Für $t \sim t_n$ gilt

$$t^2 \sim F_{1,n}.$$

Kapitel 4

Konfidenzintervalle und Hypothesentests

4-1

4. Konfidenzintervalle und Hypothesentests

4.1 Standardfehler
4.2 Konfidenzintervalle
4.3 Hypothesentests

4-2

- Alle angegebenen Konfidenzintervalle und Tests in diesem Kapitel basieren auf der folgenden Annahme für die Fehlervariablen.
- **Annahme U5**:
  Die u_t sind unabhängig und identisch (i. i. d. abkürzend für *independent identically distributed*) $N(0, \sigma^2)$-verteilt.

4-3

4.1 Standardfehler

1. Der Schätzer $\hat{\beta}_2$ für β_2 ist erwartungstreu, $E(\hat{\beta}_2) = \beta_2$, und normalverteilt,
$$\hat{\beta}_2 \sim N\left(\beta_2, \mathbb{V}(\hat{\beta}_2)\right),$$
mit der Varianz
$$\mathbb{V}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{t=1}^{T}(x_t - \bar{x})^2}.$$

2. Der **Standardfehler des Schätzers** $\hat{\beta}_2$ ist
$$se(\hat{\beta}_2) = \sqrt{\mathbb{V}(\hat{\beta}_2)} = \frac{\sigma}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2}}.$$

4-4

4. Durch Standardisierung erhält man die standardnormalverteilte Zufallsvariable
$$\frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} \sim N(0,1).$$

5. Analog gilt
$$\frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \sim N(0,1)$$
mit dem Standardfehler
$$se(\hat{\beta}_1) = \sqrt{\mathbb{V}(\hat{\beta}_1)} = \sigma\sqrt{\frac{\sum_{t=1}^{T}x_t^2}{T\sum_{t=1}^{T}(x_t - \bar{x})^2}}.$$

4-5

4.2 Konfidenzintervalle

4.2.1 Konfidenzintervalle bei bekanntem σ^2

Bei **bekanntem** σ ist
$$\left[\hat{\beta}_2 - z_{1-\frac{\alpha}{2}}se(\hat{\beta}_2),\ \hat{\beta}_2 + z_{1-\frac{\alpha}{2}}se(\hat{\beta}_2)\right] \tag{4.1}$$
ein **Konfidenzintervall** zum Niveau $1-\alpha$ für β_2.
Analog ist bei **bekanntem** σ
$$\left[\hat{\beta}_1 - z_{1-\frac{\alpha}{2}}se(\hat{\beta}_1),\ \hat{\beta}_1 + z_{1-\frac{\alpha}{2}}se(\hat{\beta}_1)\right] \tag{4.2}$$
ein **Konfidenzintervall** zum Niveau $1-\alpha$ für β_1.

4-6

- In (4.1) und (4.2) bezeichnet $z_{1-\frac{\alpha}{2}}$ das $(1-\frac{\alpha}{2})$-Fraktil von $N(0,1)$.
- Wegen der Symmetrie der Dichtefunktion gilt

$$-z_{1-\frac{\alpha}{2}} = z_{\frac{\alpha}{2}}\,.$$

- $1-\alpha$ ist das **Konfidenzniveau**.
- α ist die **Irrtumswahrscheinlichkeit**.
- Für z. B. $\alpha = 0.05 = 5\%$ gilt $z_{0.975} = 1.96$.

4-7

4.2.2 Konfidenzintervalle bei unbekanntem σ^2

- Der – in der Regel – **unbekannte** Parameter σ wird durch

$$\hat{\sigma} = \sqrt{\frac{\sum_{t=1}^{T} \hat{u}_t^2}{T-2}}$$

geschätzt.

- Man ersetzt σ durch $\hat{\sigma}$ und erhält $\widehat{se}(\hat{\beta}_1)$ aus $se(\hat{\beta}_1)$ und $\widehat{se}(\hat{\beta}_2)$ aus $se(\hat{\beta}_2)$.

4-8

- Es gilt

$$\frac{\hat{\beta}_i - \beta_i}{\widehat{se}(\hat{\beta}_i)} \sim t_{T-2}, \quad i = 1,2$$

und daher

$$P\left(-t_{T-2;1-\frac{\alpha}{2}} \leq \frac{\hat{\beta}_i - \beta_i}{\widehat{se}(\hat{\beta}_i)} \leq t_{T-2;1-\frac{\alpha}{2}}\right) = 1-\alpha\,.$$

- Dabei bezeichnet $t_{T-2;1-\frac{\alpha}{2}}$ das $(1-\frac{\alpha}{2})$-Fraktil einer t-Verteilung mit $T-2$ Freiheitsgraden.
- Wegen der Symmetrie der Dichtefunktion gilt

$$-t_{T-2;1-\frac{\alpha}{2}} = t_{T-2;\frac{\alpha}{2}}\,.$$

4-9

Ein **Konfidenzintervall** zum Niveau $1-\alpha$ für β_1 ist

$$\left[\hat{\beta}_1 - t_{T-2;1-\frac{\alpha}{2}}\,\widehat{se}(\hat{\beta}_1)\,,\ \hat{\beta}_1 + t_{T-2;1-\frac{\alpha}{2}}\,\widehat{se}(\hat{\beta}_1)\right].$$

Ein **Konfidenzintervall** zum Niveau $1-\alpha$ für β_2 ist

$$\left[\hat{\beta}_2 - t_{T-2;1-\frac{\alpha}{2}}\,\widehat{se}(\hat{\beta}_2)\,,\ \hat{\beta}_2 + t_{T-2;1-\frac{\alpha}{2}}\,\widehat{se}(\hat{\beta}_2)\right].$$

Ein **Konfidenzintervall** zum Niveau $1-\alpha$ für σ^2 ist

$$\left[\frac{(T-2)\hat{\sigma}^2}{\chi^2_{T-2;1-\frac{\alpha}{2}}}\,,\ \frac{(T-2)\hat{\sigma}^2}{\chi^2_{T-2;\frac{\alpha}{2}}}\right].$$

4-10

Es gilt

$$\frac{(T-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-2}$$

und daher

$$\begin{aligned}
1-\alpha &= P\left(\chi^2_{T-2;\frac{\alpha}{2}} \le \frac{(T-2)\hat{\sigma}^2}{\sigma^2} \le \chi^2_{T-2;1-\frac{\alpha}{2}}\right) \\
&= P\left(\frac{1}{\chi^2_{T-2;\frac{\alpha}{2}}} \ge \frac{\sigma^2}{(T-2)\hat{\sigma}^2} \ge \frac{1}{\chi^2_{T-2;1-\frac{\alpha}{2}}}\right) \\
&= P\left(\frac{(T-2)\hat{\sigma}^2}{\chi^2_{T-2;1-\frac{\alpha}{2}}} \le \sigma^2 \le \frac{(T-2)\hat{\sigma}^2}{\chi^2_{T-2;\frac{\alpha}{2}}}\right).
\end{aligned}$$

4-11

4.3 Hypothesentests

- Ziel von Hypothesentests ist die Überprüfung von Thesen über die Parameter des Modells, die sich aus der ökonomischen Theorie ergeben.
- Bestätigen die Daten Aussagen der ökonomischen Theorie?

a) Formulierung von Nullhypothese und Gegenhypothese

- Hypothesen übersetzen Aussagen der ökonomischen Theorie in Aussagen über die Parameter.
- z. B. $H_0 : \beta_2 = 0$ und $H_1 : \beta_2 \neq 0$
- z. B. „*x beeinflußt y nicht*“ versus „*x beeinflußt y*“

b) Festlegung eines Signifikanzniveaus α

4-12

- unabhängig von den Beobachtungen
- üblich sind $\alpha = 1\%$, $\alpha = 5\%$, seltener $\alpha = 10\%$, $\alpha = 0.1\%$
- α fixiert die Wahrscheinlichkeit, sich gegen H_0 (und damit für H_1) zu entscheiden, obwohl H_0 richtig (und H_1 falsch) ist.
- Fehler 1. Art:
  Entscheidung gegen H_0, obwohl H_0 richtig ist
- Fehler 2. Art:
  H_0 wird nicht abgelehnt, obwohl H_0 falsch (H_1 richtig) ist

4-13

- Zwei Fehlerarten:

	H_0 richtig	H_0 falsch (H_1 richtig)
Entscheidung für H_0	kein Fehler	**Fehler 2. Art**
Entscheidung gegen H_0	**Fehler 1. Art**	kein Fehler

- Wahrscheinlichkeit des Fehlers 1. Art $= \alpha$ (bzw. $\leq \alpha$)
- Wahrscheinlichkeit des Fehlers 2. Art
  – unbekannt
  – hängt von H_1 ab (z. B. $\beta_2 \neq 0$, β_2 beeinflußt den Fehler 2. Art)
  – wird kleiner mit wachsendem Stichprobenumfang

c) Teststatistik

4-14

- **Grundidee**: Entscheidung gegen $H_0 : \beta_2 = 0$, falls der Schätzwert $\hat{\beta}_2$ für den Parameter β_2 *zu weit* von 0 entfernt ist.
- Bei Normalverteilung der u_t gilt:

$$\frac{\hat{\beta}_2 - \beta_2}{\widehat{se}(\hat{\beta}_2)} \sim t_{T-2}$$

- Falls H_0 richtig ist, gilt $\beta_2 = 0$ und somit für die Teststatistik

$$t = \frac{\hat{\beta}_2}{\widehat{se}(\hat{\beta}_2)} \sim t_{T-2}$$

d) Annahme- und Ablehnbereich 4-15

- Falls H_0 richtig ist, gilt für die Teststatistik t:

$$P\left(-t_{T-2;1-\frac{\alpha}{2}} \leq t \leq t_{T-2;1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

- **Annahmebereich** (für die Nullhypothese) ist

$$\left[-t_{T-2;1-\frac{\alpha}{2}},\, t_{T-2;1-\frac{\alpha}{2}}\right] .$$

- Der **Ablehnbereich** (kritische Bereich) ist

$$\left(-\infty, -t_{T-2;1-\frac{\alpha}{2}}\right) \cup \left(t_{T-2;1-\frac{\alpha}{2}}, \infty\right)$$

e) Testentscheidung 4-16

- Lehne H_0 ab, falls t im Ablehnbereich liegt, anderenfalls behalte H_0 bei.
- Lehne $H_0 : \beta_2 = 0$ ab, falls

$$t \in \left(-\infty, -t_{T-2;1-\frac{\alpha}{2}}\right) \cup \left(t_{T-2;1-\frac{\alpha}{2}}, \infty\right)$$

bzw.

$$|t| \geq t_{T-2;1-\frac{\alpha}{2}}$$

- $t_{T-2;1-\frac{\alpha}{2}}$ heißt in diesem Zusammenhang auch **kritischer Wert**.

Kapitel 5

Multiples lineares Regressionsmodell

5-1

5. Multiples lineares Regressionsmodell

5-2

5.1 Modellbeschreibung und Annahmen

- T Gleichungen, k Parameter, generelle Voraussetzung: $T > k$

$$\begin{aligned} y_1 &= \beta_1 x_{11} + \beta_2 x_{21} + \ldots + \beta_k x_{k1} + u_1 \\ &\vdots \\ y_t &= \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t \\ &\vdots \\ y_T &= \beta_1 x_{1T} + \beta_2 x_{2T} + \ldots + \beta_k x_{kT} + u_T \end{aligned}$$

  bzw.

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \quad t = 1, \ldots, T$$

- Häufig gilt $x_{1t} = 1$ für $t = 1, \ldots, T$, d. h. es liegt eine **Regression mit Absolutglied** oder **inhomogene Regression** vor.

5-3

In **Matrixschreibweise**:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{5.1}$$

mit

- $\mathbf{y}$ der $T \times 1$-Vektor der **abhängigen Variablen**
- $\mathbf{X}$ eine $T \times k$-Matrix, **Regressormatrix**
- $\boldsymbol{\beta}$ ein $k \times 1$-Vektor der **Parameter**
- $\mathbf{u}$ ein $T \times 1$-Vektor der **Störvariablen** (Fehlervektor)

Annahmen

5-4

- M: Das Modell (5.1) ist korrekt spezifiziert.
- X: $\mathbf{X}$ ist nichtstochastisch und $\text{rg}(\mathbf{X}) = k$
- U1: $\mathbb{E}(\mathbf{u}) = \mathbf{0}$
- U2: $\mathbb{V}(\mathbf{u}) = \sigma^2 \mathbf{I}$
- U3: $\mathbf{u}$ ist multivariat normalverteilt

In U1 ist $\mathbb{E}(\mathbf{u})$ ein $T \times 1$-Vektor von Erwartungswerten und $\mathbf{0}$ bezeichnet einen $T \times 1$-Nullvektor.
In U2 ist $\mathbb{V}(\mathbf{u})$ die (Varianz-)Kovarianzmatrix des Zufallsvektors $\mathbf{u}$ und $\mathbf{I}$ bezeichnet die $T \times T$-Einheitsmatrix.

5-5

5.2 Schätzung der Parameter

- Die Minimierung von

$$\sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{k} \beta_i x_{it} \right)^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

bezüglich β führt zum **KQ-Schätzer**

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

für den Parametervektor $\boldsymbol{\beta}$.

- Es gilt also

$$(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \min_{\boldsymbol{\beta} \in \mathbb{R}^k} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

5-6

- Der Vektor der **geschätzten y-Werte** ist

$$\hat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}.$$

- Der Vektor der **geschätzten Residuen** ist

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}.$$

5-7

5.3 Eigenschaften der Schätzer

Linearität und Erwartungstreue

- Der KQ-Schätzer $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ ist ein **linearer Schätzer**,

$$\widehat{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y} \quad \text{mit} \quad \mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,.$$

- Der KQ-Schätzer ist **erwartungstreu**, d. h.

$$\mathbb{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}.$$

- Somit gilt

$$\mathbb{E}(\hat{\mathbf{y}}) = \mathbb{E}(\mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{X}\mathbb{E}(\widehat{\boldsymbol{\beta}}) = \mathbf{X}\boldsymbol{\beta} = \mathbf{y}$$

5-8

Standardfehler

- Die **Kovarianzmatrix** von $\widehat{\boldsymbol{\beta}}$ ist

$$\mathbb{V}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \tag{5.2}$$

- Der Schätzer

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T-k}$$

für die Varianz σ^2 ist **erwartungstreu**, d. h. $\mathbb{E}(\hat{\sigma}^2) = \sigma^2$.

- Ein Schätzer für die Kovarianzmatrix (5.2) ist

$$\widehat{\mathbb{V}}(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

- Der **Standardfehler** für den Schätzer $\hat{\beta}_i$ ist

$$\widehat{se}(\hat{\beta}_i) = \hat{\sigma}\sqrt{d_{ii}},$$

wobei d_{ii} das i-te Diagonalelement der Matrix $(\mathbf{X}'\mathbf{X})^{-1}$ ist.

Gauß-Markov-Theorem 5-9

- Unter den Annahmen M, X, U1 und U2 ist der KQ-Schätzer

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

der **beste lineare unverzerrte Schätzer** (BLUE, *best linear unbiased estimator*), d. h. unter allen linearen und erwartungstreuen Schätzern ist $\widehat{\boldsymbol{\beta}}$ varianzminimal in dem Sinn, daß

$$\mathbb{V}(\mathbf{x}'\tilde{\boldsymbol{\beta}}) \geq \mathbb{V}(\mathbf{x}'\widehat{\boldsymbol{\beta}}) \text{ für alle } \mathbf{x} = (x_1, \ldots, x_k)' \in \mathbb{R}^k \tag{5.3}$$

für jeden Schätzer $\tilde{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y}$ mit $\mathbb{E}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ gilt.

- Unter den Annahmen M, X, U1, U2 und U3 ist der KQ-Schätzer $\widehat{\boldsymbol{\beta}}$ der **beste unverzerrte Schätzer** (BUE, *best unbiased estimator*), d. h. unter allen – nicht nur den linearen – erwartungstreuen Schätzern ist $\widehat{\boldsymbol{\beta}}$ varianzminimal im dem Sinn, daß (5.3) für jeden Schätzer $\tilde{\boldsymbol{\beta}}$ mit $\mathbb{E}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ gilt.

Anmerkung zu (5.3) 5-10

- Insbesondere folgt aus (5.3), daß

$$\mathbb{V}(\tilde{\beta}_i) \geq \mathbb{V}(\hat{\beta}_i), \quad i = 1, \ldots, k.$$

- Wegen

$$\mathbb{V}(\mathbf{x}'\tilde{\boldsymbol{\beta}}) = \mathbf{x}'\mathbb{V}(\tilde{\boldsymbol{\beta}})\mathbf{x}$$

ist (5.3) ist äquivalent zu

$$\mathbf{x}'\left(\mathbb{V}(\tilde{\boldsymbol{\beta}}) - \mathbb{V}(\widehat{\boldsymbol{\beta}})\right)\mathbf{x} \geq 0 \qquad \text{für alle } \mathbf{x} \in \mathbb{R}^k,$$

d. h. die Matrix

$$\mathbb{V}(\tilde{\boldsymbol{\beta}}) - \mathbb{V}(\widehat{\boldsymbol{\beta}})$$

ist positiv semidefinit.

5-11

5.4 Bestimmtheitsmaß

Bei einer Regression mit Absolutglied gilt die **Streuungszerlegung**

$$\sum_{t=1}^{T}(y_t - \bar{y})^2 = \sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2 + \sum_{t=1}^{T}\hat{u}_t^2\,. \tag{5.4}$$

Mit den Abkürzungen

$$TSS \stackrel{\text{def}}{=} \sum_{t=1}^{T}(y_t - \bar{y})^2, \quad ESS \stackrel{\text{def}}{=} \sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2, \quad RSS \stackrel{\text{def}}{=} \sum_{t=1}^{T}\hat{u}_t^2 \tag{5.5}$$

für ***t**otal*, ***e**stimated* und ***r**esidual **s**um of **s**quares* schreibt sich (5.4) als

$$TSS = ESS + RSS\,.$$

Das **Bestimmtheitsmaß** (*coefficient of determination, R-squared*) 5-12

$$R^2 \stackrel{\text{def}}{=} \frac{\sum\limits_{t=1}^{T} (\hat{y}_t - \bar{y})^2}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2}$$

dient zur Beurteilung der Anpassungsgüte der Regression (*a measure of how well the regression line fits the data*).
R^2 mißt den Anteil der durch die Regressoren erklärten Varianz an der gesamten Varianz.
Es gilt

$$R^2 = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2}$$

und

$$0 \leq R^2 \leq 1\,.$$

Mit den Abkürzungen aus (5.5) schreibt sich R^2 als 5-13

$$R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}\,.$$

Das **adjustierte** (oder korrigierte) **Bestimmtheitsmaß** (*adjusted R-squared*) lautet

$$R^2_{\text{adj.}} \stackrel{\text{def}}{=} 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(T-k)}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2/(T-1)} \qquad \leq R^2$$

$R^2_{\text{adj.}}$ kann auch negative Werte annehmen, wenn R^2 sehr nahe bei Null liegt.

5-14

5.5 Restringierter KQ-Schätzer

- **Zusatzinformation** über die Parameter
- **Lineare Restriktionen**

  $$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

  mit $\mathbf{r}$ ein $m \times 1$-Vektor, $\mathbf{R}$ eine $m \times k$-Matrix mit $\text{rg}(\mathbf{R}) = m < k$.
- KQ-Schätzung: Minimierung der Quadratsumme mit Nebenbedingung, z. B. mit Lagrange-Verfahren
- Die Lösung der Minimierung ist der **restringierte KQ-Schätzer**

  $$\widehat{\boldsymbol{\beta}}_{\text{R}} = \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})$$

- $\widehat{\boldsymbol{\beta}}_{\text{R}}$ ist erwartungstreuer Schätzer für $\boldsymbol{\beta}$,

  $$\mathbb{E}(\widehat{\boldsymbol{\beta}}_{\text{R}}) = \mathbb{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

Beispiel: Cobb-Douglas-Produktionsfunktion $Y = F(K, L)$ 5-15

- Ausgangsgleichung

$$Y = F(K, L) = e^{\beta_1} K^{\beta_2} L^{\beta_3} e^u$$

- Lineares Modell durch **logarithmieren**

$$\ln(Y) = \beta_1 + \beta_2 \ln(K) + \beta_3 \ln(L) + u$$

bzw.

$$y = \beta_1 + \beta_2 k + \beta_3 l + u$$

mit

$$y \stackrel{\text{def}}{=} \ln(Y), \quad k \stackrel{\text{def}}{=} \ln(K), \quad l \stackrel{\text{def}}{=} \ln(L).$$

5-16

- Die **Annahme konstanter Skalenerträge**

$$F(\lambda K, \lambda L) = \lambda F(K, L), \quad \lambda > 0$$

ist wegen

$$\begin{aligned} F(\lambda K, \lambda L) &= e^{\beta_1} (\lambda K)^{\beta_2} (\lambda L)^{\beta_3} e^u \\ &= \lambda^{\beta_2} \lambda^{\beta_3} e^{\beta_1} K^{\beta_2} L^{\beta_3} e^u \\ &= \lambda^{\beta_2 + \beta_3} F(\lambda K, \lambda L) \end{aligned}$$

genau dann erfüllt, wenn

$$\beta_2 + \beta_3 = 1\,.$$

- Die entsprechende lineare Restriktion für den Parametervektor $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ ist

$$0 \cdot \beta_1 + 1 \cdot \beta_2 + 1 \cdot \beta_3 = 1\,,$$

d. h. mit $m = 1$, $k = 3$, $\mathbf{R} = (0, 1, 1)$ und $\mathbf{r} = 1$ gilt $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$.

5.6 Ergänzungen

Bemerkung 5.a (Notation für Zufallsvektoren)

1. Für einen Vektor $\mathbf{z} = (z_1, z_2, \ldots, z_n)'$ von Zufallsvariablen bezeichnet

$$\mathbb{E}(\mathbf{z}) = (\mathbb{E}(z_1), \mathbb{E}(z_2), \ldots, \mathbb{E}(z_n))'$$

den Vektor der Erwartungswerte. $\mathbb{E}(\mathbf{z})$ heißt **Erwartungswertvektor** (oder kurz Erwartungswert) des Zufallsvektors $\mathbf{z}$.

2. Die $n \times n$-Matrix der Kovarianzen $\mathbb{C}\mathrm{ov}(z_i, z_j)$ für $i, j = 1, \ldots, n$ wird mit

$$\mathbb{V}(\mathbf{z}) = [\mathbb{C}\mathrm{ov}(z_i, z_j)]_{i,j=1,\ldots,n}$$

bezeichnet. Diese Matrix heißt **Kovarianzmatrix** des Zufallsvektors $\mathbf{z}$.

3. Wegen $\mathbb{C}\mathrm{ov}(z_i, z_i) = \mathbb{V}(z_i)$ sind die Diagonalelemente der Kovarianzmatrix die Varianzen. Deswegen wird die Kovarianzmatrix auch **Varianz-Kovarianzmatrix** (*variance-covariance matrix*) genannt. Die Kovarianzmatrix eines Zufallsvektors $\mathbf{z}$ wird manchmal in der Literatur auch mit $\mathbb{C}\mathrm{ov}(\mathbf{z})$ bezeichnet.

Kapitel 6

Hypothesentest im multiplen linearen Modell

6-1

6. Hypothesentest im multiplen linearen Modell

6.1 F-Test
6.2 t-Test
6.3 Test linearer Restriktionen

6-2

Voraussetzungen für die folgenden Tests

- Normalverteilung der Fehlervariablen
- Erster Regressor ist konstant Eins ($x_{11} = x_{12} = \ldots = x_{1t} = 1$)

6-3

6.1 F-Test

- $H_0 : \beta_2 = \ldots = \beta_k = 0$
- H_1 : mindestens ein $\beta_j \neq 0$, $j = 2, \ldots, k$
- **Teststatistik**: Falls H_0 richtig ist, gilt

$$F = \frac{\sum_{t=1}^{T} (\hat{y}_t - \bar{y})^2/(k-1)}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(T-k)} \sim F_{k-1,T-k}, \tag{6.1}$$

  d. h. F hat eine F-Verteilung mit den Freiheitsgraden $k-1$ und $T-k$.

- Ablehnung von H_0 für große Werte der Statistik F. Der kritische Wert ist das Quantil $F_{k-1,T-k;1-\alpha}$.
- Testentscheidung: Lehne H_0 ab, falls $F > F_{k-1,T-k;1-\alpha}$.

6-4

- Die Teststatistik F aus (6.1) kann auch in der Form

$$F = \frac{ESS}{RSS} \cdot \frac{T-k}{k-1}$$

  oder

$$F = \frac{R^2}{1-R^2} \cdot \frac{T-k}{k-1}$$

  dargestellt werden.

- Für $R^2 \to 1$ gilt $F \to \infty$, für $R^2 \to 0$ gilt $F \to 0$.

6-5

6.2 t-Test

- $H_0 : \beta_i = 0$
- $H_1 : \beta_i \neq 0$
- Teststatistik:
  Falls H_0 richtig ist, gilt

$$t = \frac{\hat{\beta}_i}{\widehat{se}(\hat{\beta}_i)} \sim t_{T-k}$$

- Testentscheidung:
  Lehne H_0 ab, falls $|t| > t_{T-k;1-\alpha/2}$.

6-6

6.3 Test linearer Restriktionen

- Gegeben sind **lineare Restriktionen**

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

für die zu schätzenden Parameter $\boldsymbol{\beta}$.

- Dabei ist $\mathbf{r}$ ein $m \times 1$-Vektor und $\mathbf{R}$ ist eine $m \times k$-Matrix mit $\text{rg}(\mathbf{R}) = m < k$.

6-7

- **Hypothesen**: $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ versus $H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$
- **Testidee**: $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ ist äquivalent zu $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$. Werte von $\mathbf{R}\boldsymbol{\beta} - \mathbf{r}$ die „zu weit“ vom Nullvektor entfernt sind, sprechen gegen H_0.
- **Teststatistik**:
Falls H_0 richtig ist, gilt

$$F = \frac{(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})/m}{(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})/(T-k)} \sim F_{m,T-k}$$

- **Testentscheidung**:
Lehne H_0 ab, falls $F > F_{m,T-k;1-\alpha}$.

6-8

Ausblick auf Ökonometrie II

Weitergehende Fragestellungen

- Verletzung der Varianzhomogenität: **Heteroskedastizität**
- Korrelierte Fehlervariablen: **Autokorrelation**
- Verletzung der Rangbedingung $\text{rg}(X) = k$: **Multikollinearität**
- **Qualitative Variablen**, z. B. $x_i \in \{0, 1\}$.

Erforderlich: **Modifizierte Schätz- und Testverfahren**

Ökonometrie II

Kapitel 7

Vorbemerkungen zur Ökonometrie II

7-1

Ökonometrie II

7. Vorbemerkungen zur Ökonometrie II

7.1 Gliederung
7.2 Literatur
7.3 Notation

7-2

7.1 Gliederung

7 Vorbemerkungen zur Ökonometrie II
8 Klassisches lineares Modell
9 Allgemeines lineares Modell
10 Autokorrelation
11 Heteroskedastie
12 Multikollinearität
13 Regression mit Dummy-Regressoren
14 Regression mit dichotomen endogenen Variablen
15 Anhang: Eigenschaften von Matrizen
16 Anhang: Statistische Grundlagen

7-3

7.2 Literatur

Bemerkung 7.1 (Einführende Literatur)

- Gujarati, D. N.: Basic Econometrics, 4. Aufl., McGraw-Hill: New York 2003. Auch: 3. Aufl., 1995.
- Auer, L. von: Ökonometrie. Eine Einführung, 3. Aufl., Springer: Berlin, Heidelberg, New York 2005.
- Judge, G., Hill, C., Griffits, W. E., Lütkepohl, H., Lee, T.-C.: Introduction to the Theory and Practice of Econometrics. 2. Aufl., Wiley: New York 1988.

Bemerkung 7.2 (Weiterführende Literatur)

- Judge, G., Griffits, W. E., Hill, C., Lütkepohl, H., Lee, T.-C.: The Theory and Practice of Econometrics. 2. Aufl., Wiley: New York 1985.
- Greene, W. H.: Econometric Analysis. 5. Aufl., Prentice Hall: Upper Saddle River 2003.

7-4

Bemerkung 7.3 (Literatur zu einzelnen Themen)

- **Klassisches lineares Modell**: Kap. 4-8 und Appendix C in Gujarati (2003), Kap. 4-9 in Gujarati (1995), Kap. 5-7 in Judge (1988).
- **Allgemeines lineares Modell und verallgemeinerte KQ-Schätzung**: Kap. 11.3 in Gujarati (2003, 1995), Appendix C.11 in Gujarati (2003), Kap. 8 in Judge (1988).
- **Autokorrelation**: Kap. 12 in Gujarati (2003, 1995), Kap. 18 in Auer (2005), Kap. 9.5-9.6 in Judge (1988), Kap. 8 in Judge (1985).
- **Heteroskedastie**: Kap. 11 in Gujarati (2003, 1995), Kap. 17 in Auer (2005), Kap. 9.3-9.4 in Judge (1988), Kap. 11 in Judge (1985).
- **Multikollinearität**: Kap. 10 in Gujarati (2003, 1995), Kap. 21 in Auer (2005), Kap. 21 in Judge (1988), Kap. 22 in Judge (1985).
- **Binäre Variablen**: Kap. 9, 15 in Gujarati (2003), Kap. 15-16 in Gujarati (1995), Kap. 15.4 und 15.6 in Auer (2005), Kap. 11 in Judge (1988), Kap. 18 in Judge (1985).

7-5

7.3 Notation

Bemerkung 7.4 (Vorbemerkung zur Notation)

1. Bei wahrscheinlichkeitstheoretischen und statistisch-theoretischen Darstellungen ist es üblich, Zufallsvariablen durch Großbuchstaben und ihre Werte (Realisationen) durch Kleinbuchstaben zu unterscheiden. X bezeichnet dann z. B. eine Zufallsvariable und x eine Realisation von X.
2. In vielen Anwendungsbereichen der Statistik, so auch in der Ökonometrie, wird diese Unterscheidung in der Notation häufig nicht vorgenommen.
3. Dies vereinfacht einerseits die Notation, erfordert aber besondere Aufmerksamkeit, weil der Zusammenhang entscheidend ist.

7-6

4. Beispielsweise schreibt Gujarati (2003, 1995) eine lineare Einfachregression in der Form

$$Y_i = \beta_1 + \beta_2 X_i + u_i, \quad i = 1, \ldots, n\,.$$

Dabei sind die X_i feste Zahlen, während die u_i, damit auch die Y_i, Zufallsvariablen sind. Y_i und u_i können je nach Zusammenhang aber auch die Realisationen der Zufallsvariablen bedeuten.

7-7

5. Beispielsweise bezeichnet

$$\hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} \tag{7.1}$$

einerseits einen konkreten **Schätzwert** für den Parameter β bei der Regression

$$y_t = \beta x_t + u_t, \quad t = 1, \ldots, T, \tag{7.2}$$

z. B. den Wert $\hat{\beta} = 0.15$, falls y_t die beobachteten Werte sind. Andererseits kann $\hat{\beta}$ aus (7.1) auch den **Schätzer** für β bezeichnen, falls y_t die Zufallsvariablen bezeichnen. Dann ist $\hat{\beta}$ eine Zufallsvariable, die eine Wahrscheinlichkeitsverteilung besitzt und für die sich z. B. der Erwartungswert $\mathbb{E}(\hat{\beta})$ berechnen lässt.

Kapitel 8

Klassisches lineares Modell

8-1

8. Klassisches lineares Modell

8-2

8.1 Multiple lineare Regression in Matrixnotation

Bemerkung 8.1

1. Die Modellgleichung einer multiplen linearen Regression mit einem Regressanden und k Regressoren ist

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u\,.$$

2. Für T Beobachtungen ergibt sich das Gleichungssystem

$$\begin{aligned} y_1 &= \beta_1 x_{11} + \beta_2 x_{21} + \ldots + \beta_k x_{k1} + u_1 \\ y_2 &= \beta_1 x_{12} + \beta_2 x_{22} + \ldots + \beta_k x_{k2} + u_2 \\ &\vdots \\ y_T &= \beta_1 x_{1T} + \beta_2 x_{2T} + \ldots + \beta_k x_{kT} + u_T \end{aligned}$$

bzw.

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \quad t = 1, \ldots T\,.$$

8-3

Bemerkung 8.2

1. Falls $x_{11} = x_{12} = \ldots = x_{1T} = 1$, liegt eine **Regression mit Absolutglied** oder **inhomogene Regression** vor, anderenfalls eine **Regression ohne Absolutglied** oder **homogene Regression**.

2. Für eine Regression mit Absolutglied sind auch die Notationen

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_m x_{mt} + u_t, \quad t = 1, \ldots T$$

oder

$$y_t = \alpha + \beta_1 x_{1t} + \ldots + \beta_m x_{mt} + u_t, \quad t = 1, \ldots T$$

üblich, wobei $k = m + 1$.

8-4

Bemerkung 8.3
Mit der Matrix

$$\mathbf{X} = [x_{jt}] = \begin{bmatrix} x_{11} & x_{21} & \cdots & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & \cdots & x_{k2} \\ \vdots & \vdots & & & \vdots \\ x_{1T} & x_{2T} & \cdots & \cdots & x_{kT} \end{bmatrix} \in \mathbb{R}^{T \times k} \tag{8.1}$$

und den Vektoren

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \in \mathbb{R}^T, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \in \mathbb{R}^k \quad \text{und} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix} \in \mathbb{R}^T$$

resultiert für die T Gleichungen die kompakte **Matrixnotation**

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\,. \tag{8.2}$$

8-5

Bemerkung 8.4

1. Die Elemente von $\mathbf{X}$ in (8.1) sind nicht wie in der Matrizenrechnung üblich bezeichnet. Die Notation

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t, \quad t = 1, \ldots, T$$

ist naheliegend, wenn man die Matrixdarstellung (8.1) im Auge hat und die Elemente von $\mathbf{X}$ so bezeichnet, wie es in der Matrizenrechnung üblich ist. Sie wird z. B. in Judge (1988) verwendet.

2. Dagegen wird hier, wie z. B. auch von Auer (2005) und Gujarati (2003, 1995), die Darstellung

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \quad t = 1, \ldots, T$$

einzelner Gleichungen verwendet, wobei der Variablenindex der erste Index und der Index t für Beobachtungen als zweiter Index steht. Die Matrix $\mathbf{X} \in \mathbb{R}^{T \times k}$ enthält in diesem Fall das Element x_{jt} in *Zeile* t und *Spalte* j.

8-6

Bemerkung 8.5

1. Die häufigste Konvention in statistischen und ökonometrischen Softwarepaketen ist es, in einer Datenmatrix die **Spalten als Variablen** und die **Zeilen als Beobachtungen** (Fälle, Zeitpunkte etc.) zu definieren.

2. In der Regel erfolgt die Datenübergabe durch eine Datenmatrix, in der jede Variable, auch der Regressand y, eine Spalte bildet.

$$\begin{bmatrix} y_1 & x_{11} & x_{21} & \cdots & x_{k1} \\ y_2 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ y_T & x_{1T} & x_{2T} & \cdots & x_{kT} \end{bmatrix}$$

8-7

Bemerkung 8.6 (Zur Matrixnotation)

1. $\mathbb{R}$, $\mathbb{R}^n$ und $\mathbb{R}^{n\times m}$ bezeichnen die Menge der reellen Zahlen, die Menge der n-dimensionalen reellwertigen Spaltenvektoren und die Menge der reellwertigen Matrizen mit n Zeilen und m Spalten.

2. $\mathbf{x}'$ und $\mathbf{X}'$ bezeichnen einen transponierten Vektor bzw. eine transponierte Matrix.

3. $\mathbf{x} \in \mathbb{R}^n = \mathbb{R}^{n\times 1}$ ist ein Spaltenvektor, $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$.

4. $\mathbf{0} \stackrel{\text{def}}{=} (0, 0, \ldots, 0)'$ bezeichnet einen **Nullvektor**.

5. $\mathbf{I}_n = \mathbb{R}^{n\times n}$ bezeichnet eine **Einheitsmatrix**. Es wird die Kurzschreibweise $\mathbf{I} = \mathbf{I}_n$ verwendet, wenn die Dimension n aus dem Zusammenhang klar ist.

8-8

8.2 Annahmen im klassischen linearen Modell

Bemerkung 8.7 (Klassisches lineares Modell)

1. Die Modellstruktur ist

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \boldsymbol{\beta} \in \mathbb{R}^k\,. \tag{8.3}$$

2. $\mathbf{X} \in \mathbb{R}^{T\times k}$ mit $T \geq k$ ist nichtstochastisch und es gilt

$$\mathbf{X}'\mathbf{X} \in \mathbb{R}^{k\times k} \text{ ist invertierbar.} \tag{8.4}$$

3. Für die Fehlervariablen $\mathbf{u}$ gilt

$$\mathbb{E}(\mathbf{u}) = \mathbf{0} \tag{8.5}$$

und

$$\mathbb{V}(\mathbf{u}) = \sigma^2\mathbf{I} \quad \text{mit} \quad \sigma^2 > 0\,. \tag{8.6}$$

8-9

4. Normalverteilungsannahme:

Der Vektor der Fehlervariablen ist multivariat normalverteilt. (8.7)

8-10

8.3 Gewöhnlicher KQ-Schätzer

Bemerkung 8.8 (Existenz und Eindeutigkeit)
Aus der Annahme (8.4) folgt, dass

1. für jedes fixierte $\mathbf{y} \in \mathbb{R}^T$ das Minimum der Funktion

$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

eindeutig ist und als Lösung der Normalgleichungen gegeben ist,

2. für jedes fixierte $\mathbf{y} \in \mathbb{R}^T$ die **Normalgleichungen**

$$\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

die eindeutige Lösung $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ haben,

8-11

3. der **gewöhnliche KQ-Schätzer** (OLS = *ordinary least squares*)

$$\widehat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{8.8}$$

für $\boldsymbol{\beta}$ existiert.

Bemerkung 8.9
Im Folgenden bezeichnet

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{\text{OLS}}$$

immer den gewöhnlichen KQ-Schätzer, der dann auch kurz als KQ-Schätzer bezeichnet wird.

8-12

Bemerkung 8.10 (Zur Invertierbarkeit von $\mathbf{X}'\mathbf{X}$)
Für $\mathbf{X} \in \mathbb{R}^{T \times k}$ sind die sechs Bedingungen

1. $\mathbf{X}'\mathbf{X}$ ist invertierbar,
2. $\text{Rang}(\mathbf{X}) = k$,
3. $\text{Rang}(\mathbf{X}'\mathbf{X}) = k$,
4. $\det(\mathbf{X}'\mathbf{X}) > 0$,
5. $\mathbf{X}'\mathbf{X}$ ist positiv definit und
6. alle Eigenwerte von $\mathbf{X}'\mathbf{X}$ sind positiv

äquivalent.

8-13

Definition 8.11 (Linearer Schätzer)
Ein Schätzer $\check{\boldsymbol{\beta}}$ heißt **linearer Schätzer** für $\boldsymbol{\beta} \in \mathbb{R}^k$, falls eine Matrix $\mathbf{C} \in \mathbb{R}^{k \times T}$ existiert, so dass

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

Satz 8.12 (Linearität des KQ-Schätzers)
Der gewöhnliche KQ-Schätzer $\widehat{\boldsymbol{\beta}}$ ist ein linearer Schätzer.

Bemerkung 8.13 (Beweis der Linearität)
Mit $\mathbf{C} \stackrel{\text{def}}{=} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ ergibt sich $\widehat{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y}$ und somit, dass $\widehat{\boldsymbol{\beta}}$ ein linearer Schätzer ist.

8-14

Definition 8.14 (Erwartungstreuer Schätzer)
Ein Schätzer $\check{\boldsymbol{\beta}}$ heißt **erwartungstreuer** oder **unverzerrter** Schätzer für den Parametervektor $\boldsymbol{\beta} \in \mathbb{R}^k$, falls

$$\mathbb{E}(\check{\boldsymbol{\beta}}) = \boldsymbol{\beta}\,.$$

Satz 8.15 (Erwartungstreue des KQ-Schätzers)
Unter den Annahmen (8.3) bis (8.5) ist der gewöhnliche KQ-Schätzer $\widehat{\boldsymbol{\beta}}$ ein erwartungstreuer Schätzer für $\boldsymbol{\beta}$, d. h. $\mathbb{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.

8-15

Bemerkung 8.16 (Beweis der Erwartungstreue)
Aus $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ und $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ folgt

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

und somit

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\,. \tag{8.9}$$

Wegen $\mathbb{E}(\mathbf{u}) = \mathbf{0}$ gilt

$$\mathbb{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}(\mathbf{u}) = \boldsymbol{\beta}\,.$$

8-16

Satz 8.17 (Kovarianzmatrix des KQ-Schätzers)
Unter den Annahmen (8.3) bis (8.6) hat der Schätzer $\widehat{\boldsymbol{\beta}}$ *die Kovarianzmatrix*

$$\mathbb{V}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\,.$$

Bemerkung 8.18 (Herleitung der Kovarianzmatrix)
Aus (8.9) folgt

$$\begin{aligned}\mathbb{V}(\widehat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{V}(\mathbf{u})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.\end{aligned}$$

8-17

Definition 8.19 (BLU-Schätzer)
Ein Schätzer $\check{\boldsymbol{\beta}}$ für $\boldsymbol{\beta} \in \mathbb{R}^k$ heißt **bester linearer unverzerrter** Schätzer (BLU-Schätzer), falls

$$\mathbb{V}(\boldsymbol{\beta}^*) - \mathbb{V}(\check{\boldsymbol{\beta}})$$

für alle linearen und unverzerrten Schätzer $\boldsymbol{\beta}^*$ positiv semidefinit ist.

Bemerkung 8.20

1. Die Differenz $\mathbb{V}(\boldsymbol{\beta}^*) - \mathbb{V}(\check{\boldsymbol{\beta}})$ ist genau dann positiv semidefinit, falls

$$\mathbb{V}(\mathbf{x}'\boldsymbol{\beta}^*) \geq \mathbb{V}(\mathbf{x}'\check{\boldsymbol{\beta}}) \quad \text{für alle} \quad \mathbf{x} \in \mathbb{R}^k\,.$$

2. Insbesondere gilt dann

$$\mathbb{V}(\boldsymbol{\beta}^*_j) \geq \mathbb{V}(\check{\boldsymbol{\beta}}_j), \quad j = 1, \ldots, k\,.$$

8-18

Satz 8.21 (KQ-Schätzer ist BLU)
Unter den Annahmen (8.3) bis (8.6) ist der KQ-Schätzer $\widehat{\boldsymbol{\beta}}$ ein bester linearer unverzerrter Schätzer für $\boldsymbol{\beta}$.

Bemerkung 8.22

1. Der Satz 8.21 ist als **Gauß-Markoff-Theorem** bekannt.
2. Mit einigen Umformungen lässt sich zeigen, vgl. z. B. Judge et al. (1988, S. 204), dass für jeden linearen unverzerrten Schätzer $\boldsymbol{\beta}^* = \mathbf{Cy}$ gilt

$$\mathbb{V}(\boldsymbol{\beta}^*) - \mathbb{V}(\widehat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{DD}'$$

   mit $\mathbf{D} \overset{\text{def}}{=} \mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Da $\mathbf{DD}'$ positiv semidefinit ist, folgt die Behauptung des Satzes.

8-19

Satz 8.23 (Normalverteilung des KQ-Schätzers)
Unter den Annahmen (8.3) bis (8.7) ist der Schätzer $\widehat{\boldsymbol{\beta}}$ ***multivariat normalverteilt****,*

$$\widehat{\boldsymbol{\beta}} \sim N_k(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})\,.$$

Bemerkung 8.24 (Zur Normalverteilung)
Aus (8.9) und der Invertierbarkeit von $\mathbb{V}(\widehat{\boldsymbol{\beta}})$ folgt $\widehat{\boldsymbol{\beta}} \sim N_k(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, da $\mathbf{u} \sim N_T(\mathbf{0}, \sigma^2\mathbf{I})$.

Definition 8.25 (BU-Schätzer)
Ein Schätzer $\check{\boldsymbol{\beta}}$ für $\boldsymbol{\beta} \in \mathbb{R}^k$ heißt **bester unverzerrter** Schätzer (BU-Schätzer), falls

$$\mathbb{V}(\check{\boldsymbol{\beta}}) - \mathbb{V}(\boldsymbol{\beta}^*)$$

für alle unverzerrten Schätzer $\boldsymbol{\beta}^*$ positiv semidefinit ist.

Satz 8.26 (KQ-Schätzer ist BU)
Unter den Annahmen (8.3) bis (8.7) ist der Schätzer $\widehat{\boldsymbol{\beta}}$ ein BU-Schätzer für $\boldsymbol{\beta}$.

8-20

8.4 Fehlervariablen und Varianzschätzung

Bemerkung 8.27 (Eigenschaften der Fehlervariablen)

1. Wegen

$$\mathbb{V}(\mathbf{u}) = \mathbb{E}(\mathbf{uu}') - \mathbb{E}(\mathbf{u})\mathbb{E}(\mathbf{u})'$$

   lassen sich die Annahmen (8.5) und (8.6) gemeinsam auch als

$$\mathbb{E}(\mathbf{u}) = \mathbf{0}, \quad \mathbb{E}(\mathbf{uu}') = \sigma^2\mathbf{I} \quad \text{mit} \quad \sigma^2 > 0$$

   schreiben.
2. Aus der Annahme (8.6) folgt, dass die Zufallsvariablen $u_1, u_2, \ldots, u_T$ **unkorreliert** sind.
3. Aus den Annahmen (8.6) und (8.7) folgt, dass die Zufallsvariablen $u_1, u_2, \ldots, u_T$ nicht nur unkorreliert, sondern **stochastisch unabhängig** sind.

8-21

Bemerkung 8.28 (Zur Normalverteilungsannahme)
Aus den Annahmen (8.5), (8.6) und (8.7) folgt

$$\mathbf{u} \sim N_T(\mathbf{0}, \sigma^2\mathbf{I}),$$

$$u_t \sim N(0, \sigma^2), \quad t = 1, \ldots, T,$$

$$\mathbf{x}'\mathbf{u} \sim N(0, \sigma^2\mathbf{x}'\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^T, \quad \mathbf{x} \neq \mathbf{0}$$

und

$$\frac{\mathbf{u}'\mathbf{u}}{\sigma^2} \sim \chi^2_T\,.$$

8-22

Definition 8.29 (Residuen)
$\widehat{\boldsymbol{\beta}}$ sei der KQ-Schätzer für $\boldsymbol{\beta}$.

1. Dann ist

$$y \stackrel{\text{def}}{=} \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$$

   die **geschätzte Regression** (Regressionsgerade, Regressionsebene).

2. $$\hat{y}_t = \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \ldots + \hat{\beta}_k x_{kt}, \quad t = 1, \ldots T$$

   bzw.

$$\widehat{\mathbf{y}} \stackrel{\text{def}}{=} \mathbf{X}\widehat{\boldsymbol{\beta}}$$

   sind die **geschätzten** y-Werte.

3. $$\widehat{\mathbf{u}} \stackrel{\text{def}}{=} \mathbf{y} - \widehat{\mathbf{y}}$$

   ist der Vektor der **Regressionsresiduen** oder Residuenvektor.

8-23

Bemerkung 8.30 (Varianzschätzung)
Im Fall $T = k$ lässt sich der Parameter σ^2 nicht schätzen, da alle Beobachtungen $(y_t, x_{1t}, \ldots x_{kt})$ für $t = 1, \ldots, T$ exakt auf der geschätzten Regressionsgeraden (bzw. -ebene) liegen, es gilt dann also

$$\mathbf{y} = \widehat{\mathbf{y}} \quad \text{und} \quad \widehat{\mathbf{u}} = \mathbf{0}\,.$$

Satz 8.31 (Erwartungstreue Varianzschätzung)
Es sei $T > k$. Unter den Annahmen (8.3) bis (8.6) ist

$$\hat{\sigma}^2 \stackrel{\text{def}}{=} \frac{\widehat{\mathbf{u}}'\widehat{\mathbf{u}}}{T-k}$$

ein erwartungstreuer Schätzer für σ^2 und

$$\widehat{\mathbb{V}(\widehat{\boldsymbol{\beta}})} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

ist ein erwartungstreuer Schätzer der Kovarianzmatrix $\mathbb{V}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

8-24

Bemerkung 8.32 (Stichprobenverteilungen der Schätzer)

1. Unter den Annahmen (8.3) bis (8.7) gilt

$$\frac{\hat{\sigma}^2(T-k)}{\sigma^2} \sim \chi^2_{T-k},$$

falls $T > k$.

2. Unter den Annahmen (8.3) bis (8.7) gilt für $T > k$ außerdem, dass die Schätzer $\widehat{\boldsymbol{\beta}}$ für $\boldsymbol{\beta}$ und $\hat{\sigma}^2$ für σ^2 **stochastisch unabhängig** sind.

3. Auf den Verteilungen der Schätzer $\widehat{\boldsymbol{\beta}}$ und $\hat{\sigma}^2$ und deren stochastischer Unabhängigkeit beruhen die üblicherweise angewendeten statistischen Inferenzmethoden (t-Tests für die Parameter, F-Test, Tests über die Varianz, Konfidenzintervalle für die Parameter).

8-25

8.5 Verletzung der Standardannahmen

Bemerkung 8.33 (Verletzung von Annahme (8.3))

- Sind die Parameter $\boldsymbol{\beta}$ Zufallsvariablen, so erhält man Modelle mit **stochastischen Parametern**, siehe z. B. Kap. 19 in Judge (1985).

Bemerkung 8.34 (Verletzung von Annahme (8.4))

- Sind die $\mathbf{X}$ Zufallsvariablen, so erhält man Modelle mit **stochastischen Regressoren**.
- Aus (8.4) folgt die lineare Unabhängigkeit der Spalten von $\mathbf{X}$. Sind die Spalten stattdessen linear abhängig, so spricht man von **Multikollinearität** (*multicollinearity*).

8-26

Bemerkung 8.35 (Verletzung von Annahme (8.6))

- Aus (8.6) folgt die **Unkorreliertheit** der Fehlervariablen, d. h.

$$\mathbb{C}\text{ov}(u_t, u_s) = 0 \quad \text{für} \quad t \neq s\,.$$

Sind stattdessen mindestens zwei Fehlervariablen korreliert, so spricht man von **Autokorrelation** (*autocorrelation*).

- Aus (8.6) folgt die **Homoskedastizität** (*homoscedasticity*) der Fehlervariablen, d. h.

$$\mathbb{V}(u_t) = \mathbb{V}(u_s) \quad \text{für} \quad t \neq s\,.$$

Sind stattdessen mindestens zwei Varianzen unterschiedlich, so spricht man von **Heteroskedastizität** (*heteroscedasticity*).

8.6 Ergänzungen

Bemerkung 8.a (Spezialfall A)

1. Für
$$k = 1, \quad \boldsymbol{\beta} = \beta \in \mathbb{R} \quad \text{und} \quad \mathbf{X} = \mathbf{1} \stackrel{\text{def}}{=} (1, 1, \ldots, 1)' \in \mathbb{R}^T$$
vereinfacht sich die Modellgleichung (8.3) zu
$$\mathbf{y} = \beta\mathbf{1} + \mathbf{u}$$
bzw.
$$y_t = \beta + u_t, \quad t = 1, \ldots, T\,.$$
Es gilt
$$\mathbb{E}(y_t) = \beta, \quad \mathbb{V}(y_t) = \sigma^2, \quad t = 1, \ldots, T$$
und
$$\mathbb{C}\text{ov}(y_t, y_s) = 0, \quad t, s = 1, \ldots, T, \quad t \neq s\,.$$
2. Es handelt sich um ein Modell zur Mittelwertschätzung.
3. Der KQ-Schätzer für β ist
$$\hat{\beta} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y} = \frac{\sum\limits_{t=1}^{T} y_t}{T} = \bar{y}\,.$$
4. Es gilt
$$\mathbb{E}(\hat{\beta}) = \beta$$
und
$$\mathbb{V}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{1}'\mathbf{1})^{-1} = \frac{\sigma^2}{T}\,.$$
5. Ein erwartungstreuer Schätzer für σ^2 ist
$$\hat{\sigma}^2 = \frac{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2}{T - 1}\,.$$

Bemerkung 8.b (Spezialfall B)

1. Für $k = 1$, $\boldsymbol{\beta} = \beta \in \mathbb{R}$ und $\mathbf{X} = \mathbf{0} = (0, 0, \ldots, 0)' \in \mathbb{R}^T$ vereinfacht sich die Modellgleichung (8.3) zu
$$\mathbf{y} = \beta\mathbf{0} + \mathbf{u}$$
bzw.
$$y_t = \beta 0 + u_t, \quad t = 1, \ldots, T\,.$$
2. Es gilt $\text{Rang}(\mathbf{X}) = 0$, d. h. die Rangbedingung $\text{Rang}(\mathbf{X}) = k$ ist verletzt und $\mathbf{X}'\mathbf{X} = 0$ ist nicht invertierbar.
3. Offensichtlich lässt sich β nur vernünftig aus Beobachtungen schätzen, wenn diese von Null verschieden sind.

Bemerkung 8.c (Spezialfall C)

1. Für
$$k = 1, \quad \boldsymbol{\beta} = \beta \in \mathbb{R}, \quad \mathbf{X} = \mathbf{x} = (x_1, \ldots, x_T)' \in \mathbb{R}^T, \quad \mathbf{x} \neq \mathbf{0}$$
vereinfacht sich die Modellgleichung (8.3) zu
$$\mathbf{y} = \beta\mathbf{x} + \mathbf{u}$$
bzw.
$$y_t = \beta x_t + u_t, \quad t = 1, \ldots, T\,.$$

2. Es liegt eine **lineare Einfachregression ohne Absolutglied** vor (auch: homogene lineare Regression). Der KQ-Schätzer für β ist
$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} = \frac{\sum\limits_{t=1}^{T} x_t y_t}{\sum\limits_{t=1}^{T} x_t^2}\,.$$

3. Es gilt $\mathbb{E}(\hat{\beta}) = \beta$ und
$$\mathbb{V}(\hat{\beta}) = \sigma^2(\mathbf{x}'\mathbf{x})^{-1} = \frac{\sigma^2}{\sum\limits_{t=1}^{T} x_t^2}\,.$$

4. Die geschätzten y-Werte (Werte auf der geschätzten Regressionsgeraden $y = \hat{\beta}x$) sind
$$\hat{\mathbf{y}} = \hat{\beta}\mathbf{x}\,.$$
Es gilt
$$\mathbb{E}(\hat{\mathbf{y}}) = \beta\mathbf{x}\,.$$

5. Die geschätzten Residuen sind
$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}\,.$$
Es gilt
$$\mathbb{E}(\hat{\mathbf{u}}) = \mathbf{0}\,.$$

Bemerkung 8.d (Spezialfall D)

1. Für
$$k = 2, \quad \boldsymbol{\beta} = (\beta_1, \beta_2)' \in \mathbb{R}^2, \quad \mathbf{X} = [\mathbf{x}_1|\mathbf{x}_2] \in \mathbb{R}^{T\times 2}, \quad T \geq 2$$
mit
$$\mathbf{x}_1 = \mathbf{1} \in \mathbb{R}^T$$
und
$$\mathbf{x}_2 = \mathbf{x} = (x_1, \ldots, x_T)' \in \mathbb{R}^T$$
liegt eine **lineare Einfachregression mit Absolutglied** (oder inhomogene lineare Einfachregression) vor.

2. Die Modellgleichung
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$
steht für
$$y_t = \beta_1 + \beta_2 x_t + u_t, \quad t = 1, \ldots, T\,.$$

3. Zur Invertierbarkeit von $\mathbf{X}'\mathbf{X}$
   (a) Die Matrix $\mathbf{X}'\mathbf{X}$ ist genau dann invertierbar, wenn $\mathrm{Rang}(\mathbf{X}) = 2$.
   (b) Es gilt $\mathrm{Rang}(\mathbf{X}) = 2$ genau dann, wenn
   $$c_1\mathbf{1} + c_2\mathbf{x} = \mathbf{0} \quad \Longrightarrow \quad c_1 = c_2 = 0\,.$$
   (c) Die Rangbedingung ist für $\mathbf{x} = d\mathbf{1}$ mit $d \in \mathbb{R}$ und damit insbesondere für $\mathbf{x} = \mathbf{0}$ verletzt. Dies ist gleichbedeutend mit
   $$\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2 = 0, \quad \bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t\,.$$

4. Der KQ-Schätzer für $\boldsymbol{\beta}$ ist
$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\,.$$

5. Der übliche erwartungstreue Schätzer für σ^2 ist
$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})}{T-2}\,.$$

6. Umformungen führen zu den bekannten Formeln

$$\hat{\beta}_2 = \frac{\sum\limits_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum\limits_{t=1}^{T}(x_t - \bar{x})^2}$$

   und

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x}\,,$$

   wobei

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t \quad \text{und} \quad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t\,.$$

Bemerkung 8.e (Spezialfall E)

1. Für

$$k = 2, \quad \boldsymbol{\beta} = (\beta_1, \beta_2)' \in \mathbb{R}^2, \quad \mathbf{X} = [\mathbf{x}_1|\mathbf{x}_2] \in \mathbb{R}^{T\times 2}, \quad T \geq 2$$

   liegt eine **Regression mit zwei Regressoren** vor.

2. Die Modellgleichung

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

   steht für

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t, \quad t = 1, \ldots, T\ .$$

3. Es gilt

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{x}_1'\mathbf{x}_1 & \mathbf{x}_1'\mathbf{x}_2 \\ \mathbf{x}_2'\mathbf{x}_1 & \mathbf{x}_2'\mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \sum\limits_{t=1}^{T} x_{1t}^2 & \sum\limits_{t=1}^{T} x_{1t}x_{2t} \\ \sum\limits_{t=1}^{T} x_{1t}x_{2t} & \sum\limits_{t=1}^{T} x_{2t}^2 \end{bmatrix}.$$

Kapitel 9

Allgemeines lineares Modell

9-1

9. Allgemeines lineares Modell

9-2

9.1 Autokorrelation und Heteroskedastie

Definition 9.1 (Autokorrelation)
Falls $\mathbb{C}\mathrm{ov}(u_t, u_s) \neq 0$ für mindestens zwei Variablen u_t und u_s mit $t \neq s$ gilt, liegt **Autokorrelation** vor.

Definition 9.2 (Homo- und Heteroskedastie)

1. Falls $\mathbb{V}(u_t) = \mathbb{V}(u_s)$ für alle $t, s = 1, \ldots T$, liegt **Homoskedastie** (oder Homoskedastizität) vor.
2. Falls $\mathbb{V}(u_t) \neq \mathbb{V}(u_s)$ für mindestens zwei Variablen u_t und u_s mit $t \neq s$, liegt **Heteroskedastie** (oder Heteroskedastizität) vor.
3. Man spricht auch von **homoskedastischen** oder **heteroskedastischen** Fehlertermen oder Störgrößen.

9-3

Bemerkung 9.3 (Verletzung der Standardannahme)

1. Autokorrelation und Heteroskedastie sind mögliche Abweichungen von der Standardannahme (8.6),
$$\mathbb{V}(\mathbf{u}) = \sigma^2\mathbf{I} \quad \text{mit} \quad \sigma^2 > 0,$$
die getrennt oder gemeinsam auftreten können.
2. Im Fall der **Heteroskedastie ohne Autokorrelation** ist $\mathbb{V}(\mathbf{u})$ eine Diagonalmatrix mit positiven Diagonalelementen.
3. Im Fall der **Autokorrelation ohne Heteroskedastie** ist für $\mathbb{V}(\mathbf{u})$ eine Darstellung
$$\mathbb{V}(\mathbf{u}) = \sigma^2\mathbf{\Omega}$$
möglich, wobei $\mathbf{\Omega} = [\omega_{ts}]_{t,s=1,\ldots,T}$ eine Korrelationsmatrix mit Diagonalelementen $\omega_{tt} = 1$ für $t = 1, \ldots, T$ ist.

9-4

Bemerkung 9.4 (Allgemeine Kovarianzstruktur)

1. Die Annahme (8.6) wird zugunsten von
$$\mathbb{V}(\mathbf{u}) = \mathbf{\Sigma}, \quad \mathbf{\Sigma} \in \mathbb{R}^{T\times T} \text{ ist positiv definit}, \tag{9.1}$$
verallgemeinert.
2. Im Spezialfall $\mathbf{\Sigma} = \sigma^2\mathbf{I}$ ergibt sich die Standardannahme (8.6).
3. Falls $\mathbf{u}$ multivariat normalverteilt ist, gelten
$$\mathbf{u} \sim N_T(\mathbf{0}, \mathbf{\Sigma}) \tag{9.2}$$
und
$$\mathbf{u}'\mathbf{\Sigma}^{-1}\mathbf{u} \sim \chi^2_T\,.$$

9-5

Bemerkung 9.5 (Positiv definite Kovarianzmatrix)

1. Die Matrix $\mathbf{\Sigma}$ ist **positiv definit** genau dann, wenn
$$\mathbf{x}'\mathbf{\Sigma}\mathbf{x} > 0 \quad \text{für alle } \mathbf{x} \in \mathbb{R}^T \setminus \{\mathbf{0}\}\,.$$
2. Die Matrix $\mathbf{\Sigma} = \mathbb{V}(\mathbf{u})$ ist positiv definit genau dann, wenn
$$\mathbb{V}(\mathbf{x}'\mathbf{u}) > 0 \quad \text{für alle } \mathbf{x} \in \mathbb{R}^T \setminus \{\mathbf{0}\},$$
da
$$\mathbb{V}(\mathbf{x}'\mathbf{u}) = \mathbf{x}'\mathbb{V}(\mathbf{u})\mathbf{x}\,.$$
3. Aus der positiven Definitheit der Kovarianzmatrix $\mathbf{\Sigma} \in \mathbb{R}^{T\times T}$ folgt, dass $\mathbf{\Sigma}$ invertierbar ist, d. h., dass eine Matrix $\mathbf{\Sigma}^{-1} \in \mathbb{R}^{T\times T}$ mit
$$\mathbf{\Sigma}^{-1}\mathbf{\Sigma} = \mathbf{\Sigma}\mathbf{\Sigma}^{-1} = \mathbf{I}$$
existiert.

9-6

Bemerkung 9.6

1. Manchmal ist die Darstellung der Kovarianzmatrix von $\mathbf{u}$ als
$$\mathbb{V}(\mathbf{u}) = \sigma^2\mathbf{\Omega}, \quad \sigma^2 > 0, \quad \mathbf{\Omega} \in \mathbb{R}^{T\times T} \text{ ist positiv definit,} \tag{9.3}$$
sinnvoll.
2. Im Spezialfall $\mathbf{\Omega} = \mathbf{I}$ ergibt sich die Standardannahme (8.6).
3. Falls $\mathbf{u}$ multivariat normalverteilt ist, gelten
$$\mathbf{u} \sim N_T(\mathbf{0}, \sigma^2\mathbf{\Omega}) \quad \text{und} \quad \frac{\mathbf{u}'\mathbf{\Omega}^{-1}\mathbf{u}}{\sigma^2} \sim \chi^2_T\,. \tag{9.4}$$
4. Im Fall der Homoskedastizität können σ^2 und $\mathbf{\Omega}$ so gewählt werden, dass
$$\sigma^2 = \mathbb{V}(u_1) = \ldots = \mathbb{V}(u_T)$$
gilt und alle Diagonalelemente von $\mathbf{\Omega}$ den Wert Eins haben. $\mathbf{\Omega}$ ist dann eine **Korrelationsmatrix** und **Autokorrelation** liegt vor, falls $\mathbf{\Omega} \neq \mathbf{I}$.

9-7

9.2 Annahmen im allgemeinen linearen Modell

Bemerkung 9.7 (Allgemeine Kovarianzstruktur)

1. Die Annahmen (8.3) bis (8.5) werden beibehalten, d. h. es gilt
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$
mit $\boldsymbol{\beta} \in \mathbb{R}^k$, $\mathbf{X} \in \mathbb{R}^{T\times k}$, $T \geq k$, $\mathbf{X}'\mathbf{X}$ ist invertierbar und $\mathbb{E}(\mathbf{u}) = \mathbf{0}$.
2. Anstatt der Annahme (8.6) wird die allgemeinere Annahme (9.1) gemacht, d.h.
$$\mathbb{V}(\mathbf{u}) = \mathbf{\Sigma} \in \mathbb{R}^{T\times T}, \quad \mathbf{\Sigma} \text{ ist positiv definit.}$$
3. Die Annahmen (8.3) bis (8.5) zusammen mit (9.1) werden im Folgenden als die **Annahmen des allgemeinen linearen Modells** bezeichnet.

9-8

4. Die Bezeichnung **allgemeines** lineares Modell bezieht sich auf die **allgemeine Kovarianzstruktur** im Vergleich zur **speziellen Kovarianzstruktur** $\sigma^2\mathbf{I}$ des klassischen linearen Modells.
5. Das **allgemeine** lineare Modell (*general linear model*) muss vom **verallgemeinerten** linearen Modell (*generalized linear model*) unterschieden werden. Im allgemeinen linearen Modell gilt
$$\mathbb{E}(y_t) = \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt}, \quad t = 1, \ldots, T,$$
während im verallgemeinerten linearen Modell
$$\mathbb{E}(y_t) = g(\beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt}), \quad t = 1, \ldots, T,$$
mit einer nichtlinearen Funktion g gilt. Beispiele aus der Klasse verallgemeinerter linearer Modelle sind die im Kapitel 14 behandelten Logit- und Probitmodelle.

9-9

Bemerkung 9.8 (Eigenschaften des KQ-Schätzers)

1. Da die Annahmen (8.3) bis (8.5) beibehalten werden, gilt weiterhin, dass der gewöhnliche KQ-Schätzer
$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$
für $\boldsymbol{\beta}$ existiert und ein linearer und unverzerrter Schätzer für $\boldsymbol{\beta}$ ist.

2. Im allgemeinen linearen Modell ist die Kovarianzmatrix des gewöhnlichen KQ-Schätzers
$$\mathbb{V}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\,.$$
Im Fall $\mathbb{V}(\mathbf{u}) = \sigma^2\boldsymbol{\Omega}$ mit $\sigma^2 > 0$ gilt
$$\mathbb{V}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\,.$$

9-10

Bemerkung 9.9 (Auswirkungen auf KQ-Schätzer)

1. Da sich die Kovarianzmatrix von $\widehat{\boldsymbol{\beta}}$ geändert hat, kann nicht mehr erwartet werden, dass $\widehat{\boldsymbol{\beta}}$ ein bester Schätzer ist.

2. Die Formel $\mathbb{V}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ ist nicht mehr richtig und durch
$$\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$
wird die Kovarianzmatrix des KQ-Schätzers falsch geschätzt.

3. Wird der gewöhnliche KQ-Schätzer mit den Varianzformeln aus dem klassischen Modell verwendet, so fehlt eine Begründung für die Inferenzverfahren. Diese können zu unsinnigen Ergebnissen führen.

9-11

9.3 Verallgemeinerter KQ-Schätzer

Bemerkung 9.10
Falls die Kovarianzmatrix von $\mathbf{u}$ bekannt ist, kann ein im Vergleich zum gewöhnlichen KQ-Schätzer verbesserter Schätzer eingesetzt werden.

Definition 9.11 (Verallgemeinerter KQ-Schätzer)
Die Kovarianzmatrix von $\mathbf{u}$, $\boldsymbol{\Sigma} \overset{\text{def}}{=} \mathbb{V}(\mathbf{u})$, sei invertierbar. Dann heißt der Schätzer
$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} \overset{\text{def}}{=} (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \tag{9.5}$$
verallgemeinerter Kleinst-Quadrate-Schätzer (GLS-Schätzer, GLS = *generalized least squares*) oder **Aitken-Schätzer**. Er wird im Folgenden kurz als **VKQ-Schätzer** bezeichnet.

Bemerkung 9.12
Im Spezialfall $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ gilt
$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = \widehat{\boldsymbol{\beta}}\,.$$

9-12

Satz 9.13 (BLU-Schätzer im allgemeinen linearen Modell)
Unter den Annahmen des allgemeinen linearen Modells gilt:

1. *Der VKQ-Schätzer $\widehat{\boldsymbol{\beta}}_{\text{GLS}}$ ist bester linearer unverzerrter Schätzer für den Parametervektor $\boldsymbol{\beta} \in \mathbb{R}^k$.*
2. *Der VKQ-Schätzer $\widehat{\boldsymbol{\beta}}_{\text{GLS}}$ hat die Kovarianzmatrix*

$$\mathbb{V}(\widehat{\boldsymbol{\beta}}_{\text{GLS}}) = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\,.$$

Bemerkung 9.14

- Insbesondere gilt also $\mathbb{E}(\widehat{\boldsymbol{\beta}}_{\text{GLS}}) = \boldsymbol{\beta}$.
- Der VKQ-Schätzer ist besser (im Sinn der Effizienz) als der gewöhnliche KQ-Schätzer, d. h. die Differenz

$$\mathbb{V}(\widehat{\boldsymbol{\beta}}) - \mathbb{V}(\widehat{\boldsymbol{\beta}}_{\text{GLS}})$$

  ist eine positiv semidefinite Matrix.

9-13

Bemerkung 9.15 (Zum Beweis von Satz 9.13)

1. Da $\boldsymbol{\Sigma}$ invertierbar ist, existiert $\boldsymbol{\Sigma}^{-1/2}$. Durch Linksmultiplikation mit $\boldsymbol{\Sigma}^{-1/2}$ erhält man aus $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ die transformierte Modellgleichung

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

   mit $\mathbf{y}^* = \boldsymbol{\Sigma}^{-1/2}\mathbf{y}$, $\mathbf{X}^* = \boldsymbol{\Sigma}^{-1/2}\mathbf{X}$, $\mathbf{u}^* = \boldsymbol{\Sigma}^{-1/2}\mathbf{u}$ und $\mathbb{V}(\mathbf{u}^*) = \mathbf{I}$.

2. Der gewöhnliche KQ-Schätzer im transformierten Modell ist

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^* = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

   mit der Kovarianzmatrix

$$\mathbb{V}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\,.$$

3. Die BLU-Eigenschaft von $\widehat{\boldsymbol{\beta}}$ im transformierten klassischen Modell überträgt sich auf das allgemeine Modell.

9-14

Bemerkung 9.16 (Der Fall $\mathbb{V}(\mathbf{u}) = \sigma^2\boldsymbol{\Omega}$)

1. Für $\mathbb{V}(\mathbf{u}) = \sigma^2\boldsymbol{\Omega}$ mit $\sigma^2 > 0$ und $\boldsymbol{\Omega}$ invertierbar gilt

$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

   mit der Kovarianzmatrix

$$\mathbb{V}(\widehat{\boldsymbol{\beta}}_{\text{GLS}}) = \sigma^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\,.$$

2. Der VKQ-Schätzer kann daher eingesetzt werden, wenn $\boldsymbol{\Omega}$ bekannt, aber σ^2 unbekannt ist. $\mathbb{V}(\mathbf{u})$ ist dann bis auf einen positiven Faktor bekannt.

9-15

Bemerkung 9.17 (Schätzung von σ^2 im Fall $\mathbb{V}(\mathbf{u}) = \sigma^2\boldsymbol{\Omega}$)

1. Ein erwartungstreuer Schätzer für den unbekannten Parameter σ^2 ist

$$\hat{\sigma}^2_{\text{GLS}} \stackrel{\text{def}}{=} \frac{\hat{\mathbf{u}}'\boldsymbol{\Omega}^{-1}\hat{\mathbf{u}}}{T-k}$$

mit

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\text{GLS}}\,.$$

2. Dagegen ist der Schätzer

$$\hat{\sigma}^2_{\text{OLS}} \stackrel{\text{def}}{=} \frac{(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})}{T-k}$$

für σ^2, wobei $\widehat{\boldsymbol{\beta}}$ der gewöhnliche KQ-Schätzer ist, nicht erwartungstreu, d. h. verzerrt.

9-16

3. Ist $\mathbf{u}$ multivariat normalverteilt, so gilt

$$\frac{\hat{\mathbf{u}}'\boldsymbol{\Omega}^{-1}\hat{\mathbf{u}}}{\sigma^2} = \frac{(T-k)\hat{\sigma}^2_{\text{GLS}}}{\sigma^2} \sim \chi^2_{T-k}\,.$$

9-17

9.4 Praktikabler verallgemeinerter KQ-Schätzer

Bemerkung 9.18
Wenn $\boldsymbol{\Omega}$ **unbekannt** ist, kann der VKQ-Schätzer

$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

nicht eingesetzt werden. Der folgende Schätzer beruht auf einem zweiphasigen Schätzverfahren, wobei zunächst $\boldsymbol{\Omega}$ und dann $\boldsymbol{\beta}$ geschätzt wird.

Definition 9.19 (Praktikabler VKQ-Schätzer)
Wenn $\boldsymbol{\Omega}$ in einer ersten Phase zunächst durch $\widehat{\boldsymbol{\Omega}}$ (konsistent) geschätzt wird, und dann in einer zweiten Phase der Schätzer

$$\widehat{\boldsymbol{\beta}}_{\text{FGLS}} \stackrel{\text{def}}{=} (\mathbf{X}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$$

verwendet wird, spricht man vom **praktikablen verallgemeinerten KQ-Schätzer** (FGLS = *feasible generalized least squares*).

9-18

Bemerkung 9.20

1. Andere Bezeichnungen sind **geschätzter verallgemeinerter KQ-Schätzer** (EGLS-Schätzer, EGLS = *estimated generalized least squares*) und **zweistufiger Aitken-Schätzer**.

2. Da $\widehat{\boldsymbol{\Omega}}$ von den Beobachtungen abhängt, ist $\widehat{\boldsymbol{\beta}}_{\text{FGLS}}$ kein linearer und erwartungstreuer Schätzer mehr.

3. Für dieses Schätzverfahren sind nur asymptotische Eigenschaften (für $T \to \infty$) bekannt, wenn $\boldsymbol{\Omega}$ konsistent geschätzt werden kann.

4. Dazu müssen für $\boldsymbol{\Omega}$ Strukturen spezifiziert werden, die für $T \to \infty$ erhalten bleiben und welche die Anzahl der zu schätzenden Parameter einschränken. Man beachte, dass im Allgemeinen die Anzahl der Parameter in $\boldsymbol{\Omega}$ mit T quadratisch wächst.

5. Die Optimalitätseigenschaften der VKQ-Schätzung können nur für $T \to \infty$ erwartet werden.

Kapitel 10

Autokorrelation

10-1

10. Autokorrelation

10-2

10.1 Modellierung von Autokorrelation

Bemerkung 10.1 (Arten von Autokorrelation)

1. Abwesenheit von Autokorrelation bedeutet $\mathbb{C}\text{ov}(u_t, u_s) = 0$ für alle t und s mit $t \neq s$.
2. Falls $\mathbb{C}\text{ov}(u_t, u_s) \neq 0$ für mindestens zwei Variablen u_t und u_s mit $t \neq s$ gilt, liegt **Autokorrelation** vor.
3. Wichtige Spezialfälle für zeitlich geordnete Beobachtungen sind der Fall **positiver Autokorrelation**, falls

$$\mathbb{C}\text{ov}(u_t, u_{t+1}) > 0, \quad t = 1, 2, \ldots,$$

und der Fall **negativer Autokorrelation**, falls

$$\mathbb{C}\text{ov}(u_t, u_{t+1}) < 0, \quad t = 1, 2, \ldots, .$$

10-3

Bemerkung 10.2 (AR(1)-Prozess)

1. Für die u_t wird ein **autoregressiver Prozess erster Ordnung**

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{10.1}$$

mit dem **Autokorrelationskoeffizienten erster Ordnung**

$$-1 < \rho < 1$$

spezifiziert.

2. Die Modellannahmen für die Fehlervariablen ε_t sind

$$\mathbb{E}(\varepsilon_t) = 0,$$

$$\mathbb{V}(\varepsilon_t) = \sigma^2$$

und

$$\mathbb{C}\mathrm{ov}(\varepsilon_s, \varepsilon_t) = 0, \quad t \neq s\,.$$

10-4

Bemerkung 10.3 (Eigenschaften der u_t)
Bei geeignetem Startwert u_0 ist durch (10.1) ein Folge

$$u_1, u_2, \ldots, u_t, \ldots$$

festgelegt mit

- den Erwartungswerten

$$\mathbb{E}(u_t) = 0, \quad t = 1, 2, \ldots\,,$$

- den Varianzen

$$\mathbb{V}(u_t) = \frac{\sigma^2}{1-\rho^2}, \quad t = 1, 2, \ldots$$

- und den Autokovarianzen erster Ordnung

$$\mathbb{C}\mathrm{ov}(u_t, u_{t+1}) = \frac{\sigma^2 \rho}{1-\rho^2}\,.$$

10-5

Bemerkung 10.4 (Kovarianzmatrix von u)

1. Es gilt

$$\mathbb{C}\mathrm{ov}(u_t, u_s) = \frac{\sigma^2 \rho^{|t-s|}}{1-\rho^2}, \quad t, s = 1, \ldots, T$$

und

$$\mathbb{C}\mathrm{orr}(u_t, u_s) = \rho^{|t-s|}, \quad t, s = 1, \ldots, T\,.$$

2. Es resultiert

$$\mathbb{V}(\mathbf{u}) = \sigma^2 \mathbf{\Omega}$$

mit

$$\mathbf{\Omega} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{T-2} \\ & & \cdots & & & \\ & & \cdots & & & \\ \rho^{T-2} & \cdots & \rho^2 & \rho & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \cdots & \rho^2 & \rho & 1 \end{bmatrix}.$$

10-6

10.2 Schätzen bei Autokorrelation

Bemerkung 10.5 (VKQ-Schätzer bei bekanntem ρ)

1. Falls ρ **bekannt** ist, ist $\mathbf{\Omega}$ bekannt. Es kann dann der verallgemeinerte KQ-Schätzer
$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y} \tag{10.2}$$
verwendet werden.

2. Es ist
$$\mathbf{\Omega}^{-1} = \begin{bmatrix} 1 & -\rho & 0 & & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ & & \cdots & & & \\ & & \cdots & & & \\ & \cdots & 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & & \cdots & 0 & -\rho & 1 \end{bmatrix} . \tag{10.3}$$

10-7

Bemerkung 10.6 (Schätzer für ρ)

1. Falls ρ **unbekannt** ist, ist auch $\mathbf{\Omega}^{-1}$ aus (10.3) unbekannt. Der VKQ-Schätzer aus (10.2) kann dann nicht eingesetzt werden.

2. Ein Schätzer für ρ ist durch
$$\hat{\rho} = \frac{\sum\limits_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum\limits_{t=2}^{T} \hat{u}_{t-1}^2} \tag{10.4}$$
gegeben.

3. Dies ist der gewöhnliche KQ-Schätzer für ρ aus der Regression
$$\hat{u}_t = \rho \hat{u}_{t-1} + \tilde{\varepsilon}_t, \quad t = 2, \ldots, T .$$
Für diese Regression liegen die $T-1$ Beobachtungspaare $(\hat{u}_t, \hat{u}_{t-1})$ für $t = 2, \ldots, T$ vor.

10-8

Bemerkung 10.7 (PVKQ bei geschätztem ρ)
Falls ρ **unbekannt** ist, können ρ durch $\hat{\rho}$ aus (10.4),
$\mathbf{\Omega}^{-1}$ durch
$$\widehat{\mathbf{\Omega}}^{-1} = \begin{bmatrix} 1 & -\hat{\rho} & 0 & & \cdots & 0 \\ -\hat{\rho} & 1+\hat{\rho}^2 & -\hat{\rho} & 0 & \cdots & 0 \\ 0 & -\hat{\rho} & 1+\hat{\rho}^2 & -\hat{\rho} & \cdots & 0 \\ & & \cdots & & & \\ & & \cdots & & & \\ & \cdots & 0 & -\hat{\rho} & 1+\hat{\rho}^2 & -\hat{\rho} \\ 0 & & \cdots & 0 & -\hat{\rho} & 1 \end{bmatrix} \tag{10.5}$$
und $\boldsymbol{\beta}$ durch den **praktikablen verallgemeinerten KQ-Schätzer** (**PVKQ-Schätzer**)
$$\widehat{\boldsymbol{\beta}}_{\text{FGLS}} = (\mathbf{X}'\widehat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{\Omega}}^{-1}\mathbf{y}$$
geschätzt werden.

10-9

Bemerkung 10.8 (Cochran-Orcutt-Verfahren)

1. Gewöhnliche KQ-Schätzung führt zu dem Residuenvektor $\widehat{\mathbf{u}}$.
2. Der Parameter ρ wird mit $\widehat{\mathbf{u}}$ durch $\hat{\rho}$ aus (10.4) geschätzt.
3. Die Matrix $\widehat{\boldsymbol{\Omega}}^{-1}$ wird durch (10.5) unter Verwendung von $\hat{\rho}$ geschätzt.
4. Der Parametervektor $\boldsymbol{\beta}$ wird durch den PVKQ-Schätzer
$$\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}} = (\mathbf{X}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$$
geschätzt.
5. Ein **modifizierter** Residuenvektor wird berechnet,
$$\widehat{\mathbf{u}} \stackrel{\mathrm{def}}{=} \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}}\,.$$

Die Schritte 2. bis 5. werden **iterativ** so lange durchlaufen, bis sich die Schätzungen nicht mehr ändern.

10-10

10.3 Tests auf Autokorrelation

10.3.1 Ein asymptotischer Test

Bemerkung 10.9 (Ein asymptotischer Test für ρ)

- Voraussetzung: Die Fehlervariablen folgen einem AR(1)-Prozess.
- Unter $H_0 : \rho = 0$ ist
$$\sqrt{T}\hat{\rho}$$
asymptotisch (für $T \to \infty$) standardnormalverteilt. Dabei ist $\hat{\rho}$ der Schätzer aus (10.4).
- Auf dem 5%-Signifikanzniveau wird H_0 zugunsten von $H_1 : \rho \neq 0$ verworfen, falls
$$|\sqrt{T}\hat{\rho}| > 1.96\,.$$

10.3.2 Durbin-Watson-Test

10-11

Bemerkung 10.10 (Testidee)

1. Der Durbin-Watson-Test (DW-Test) setzt einen AR(1)-Prozess
$$u_t = \rho u_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T$$
mit **normalverteilten** Fehlervariablen ε_t voraus.
2. Unter $H_0 : \rho = 0$ ist keine Autokorrelation vorhanden.
3. Durbin und Watson entwickelten eine Teststatistik
$$d \approx 2(1 - \hat{\rho})$$
mit folgenden Eigenschaften
$$d \approx 2, \text{ falls } \rho \approx 0,$$
$$d \approx 0, \text{ falls } \rho \approx +1$$
und
$$d \approx 4, \text{ falls } \rho \approx -1.$$

10-12

Bemerkung 10.11 (Voraussetzungen des DW-Testes)

- Ist ein sehr **spezielles** Autokorrelationsmuster **vorausgesetzt**.
- Die Daten sind zeitlich geordnet **ohne fehlende Zwischenwerte**.
- Es wird eine **Regression mit Absolutglied** (*intercept term*) geschätzt.
- Die erklärte Variable y_t wird nicht in verzögerter Form als Regressor verwendet.
- Der Prozess für die Fehlervariablen ist ein AR(1)-Prozess.
- Die Fehlervariablen sind normalverteilt.

10-13

Bemerkung 10.12 (Testdurchführung)

1. Die Teststatistik (Prüfgröße) ist

$$d = \frac{\sum\limits_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum\limits_{t=1}^{T} \hat{u}_t^2} \, ,$$

   wobei $\hat{u}_t$ die Residuen einer gewöhnlichen KQ-Regression sind.

2. Die Verteilung der Teststatistik d unter H_0 hängt von $\mathbf{X}$ ab.
3. Es gibt aber obere und untere **Abschätzungen**, die nicht von $\mathbf{X}$ abhängen, und die zu zwei **Tabellenwerten** $0 < d_L < d_U < 2$ in Abhängigkeit vom vorgegebenen Signifikanzniveau α und von der Anzahl der Regressoren führen.
4. Es gilt

$$0 < d_L < d_U < 2 < 4 - d_U < 4 - d_L < 4 \, .$$

10-14

Bemerkung 10.13 (Tabellen)

- In Gujarati (2003) und Judge et al. (1988) sind Tabellen von d_U und d_L für $\alpha = 0.01$ und $\alpha = 0.05$, jeweils für $k = 2(1)21$ und für $n = 6(1)40(5)100(50)200$ enthalten.
- Ökonometrische Software kann häufig die exakte Verteilung von d unter H_0, die von $\mathbf{X}$ abhängt, durch Simulation bestimmen und entsprechende p-Werte ausgeben.

10-15

Bemerkung 10.14 (DW-Test auf positive Autokorrelation)

- Getestet wird auf **positive Autokorrelation** erster Ordnung.
- Die Hypothesen sind $H_0 : \rho = 0$ versus $H_1 : \rho > 0$.
- H_0 wird zugunsten von H_1 verworfen, falls $d < d_L$.
- H_0 wird **nicht** zugunsten von H_1 verworfen, falls $d > d_U$.
- Es ist **keine Entscheidung** möglich, falls $d_L \leq d \leq d_U$.
- Der Test hat das **Niveau** α, falls α das Signifikanzniveau der Tabelle ist.

10-16

Bemerkung 10.15 (DW-Test auf negative Autokorrelation)

- Getestet wird auf **negative Autokorrelation** erster Ordnung.
- Die Hypothesen sind $H_0 : \rho = 0$ versus $H_1 : \rho < 0$.
- H_0 wird zugunsten von H_1 verworfen, falls $d > 4 - d_L$.
- H_0 wird **nicht** zugunsten von H_1 verworfen, falls $d < 4 - d_U$.
- Es ist **keine Entscheidung** möglich, falls $4 - d_U \leq d \leq 4 - d_L$.
- Der Test hat das **Niveau** α, falls α das Signifikanzniveau der Tabelle ist.

10-17

Bemerkung 10.16 (DW-Test auf Autokorrelation)

- Getestet wird auf **positive oder negative Autokorrelation** erster Ordnung.
- Die Hypothesen sind $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$.
- H_0 wird zugunsten von H_1 verworfen, falls $d < d_L$ oder $d > 4 - d_L$.
- H_0 wird **nicht** zugunsten von H_1 verworfen, falls $d_U < d < 4 - d_U$.
- Es ist **keine Entscheidung** möglich, falls $d_L \leq d \leq d_U$ oder $4 - d_U \leq d \leq 4 - d_L$.
- Der Test hat das **Niveau** 2α, falls α das Signifikanzniveau der Tabelle ist.

10.3.3 Test für allgemeinere Autokorrelationsmuster

10-18

Bemerkung 10.17 (Breusch-Godfrey-Test)

1. Für die Residualvariablen wird der AR(p)-Prozess

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t$$

   unterstellt.

2. Die Ordnung p muss vor der Testdurchführung und unabhängig von den Daten spezifiziert werden.

3. Die Nullhypothese ist

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_p = 0.$$

4. Die Gegenhypothese ist

$$H_1 : \quad \text{Es gibt mindestens ein } j \in \{1, \ldots, p\} \text{ mit } \rho_j \neq 0.$$

10-19

Bemerkung 10.18 (Testdurchführung)
Die Testdurchführung wird am Beispiel des Modells

$$y_t = \beta_1 + \beta_2 x_t + u_t, \quad t = 1, \ldots, T$$

illustriert.

1. Die Residuen $\hat{u}_t$ aus einer gewöhnlichen KQ-Schätzung werden berechnet.

2. Die R^2-Maßzahl der Regression

$$\hat{u}_t = \alpha_1 + \alpha_2 x_t + \rho_1 \hat{u}_{t-1} + \ldots + \rho_p \hat{u}_{t-p} + \eta_t$$

   wird berechnet.

3. Falls H_0 richtig ist, ist die Teststatistik

$$(T - p)R^2$$

   für $T \to \infty$ asymptotisch chiquadratverteilt mit p Freiheitsgraden.

10-20

Bemerkung 10.19
Für $p = 1$ ist dieser Test als der **Test von Durbin** bekannt.

Kapitel 11

Heteroskedastie

11-1

11. Heteroskedastie

11-2

11.1 Folgen der Heteroskedastie

Bemerkung 11.1

1. **Heteroskedastie** oder Heteroskedastizität (*heteroscedasticity*) liegt vor, falls $\mathbb{V}(u_t) \neq \mathbb{V}(u_s)$ für mindestens zwei Variablen u_t und u_s mit $t \neq s$. Man spricht dann auch von **heteroskedastischen Störgrößen**.

2. Im Folgenden wird nur der **reine Fall der Heteroskedastie ohne Autokorrelation** betrachtet, in welchem die Kovarianzmatrix $\mathbb{V}(\mathbf{u})$ eine Diagonalmatrix mit positiven Diagonalelementen

$$\sigma_t^2 = \mathbb{V}(u_t), \quad t = 1, \ldots, T \tag{11.1}$$

ist und

$$\mathbb{C}\mathrm{ov}(u_t, u_s) = 0, \quad t \neq s$$

gilt.

11-3

Beispiel 11.2 (Fehlschätzung der Varianz)

1. Im Modell
$$y_t = \beta_1 + \beta_2 x_t + u_t, \quad t = 1, \ldots, T$$
ist
$$\hat{\beta}_2 = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$
der gewöhnliche KQ-Schätzer für den Parameter β_2.

2. Die Varianz von $\hat{\beta}_2$ ist
$$\mathbb{V}(\hat{\beta}_2) = \frac{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sigma_t^2}{\left(\sum_{t=1}^{T}(x_t - \bar{x})^2\right)^2} . \tag{11.2}$$

11-4

3. Diese vereinfacht sich im Fall der Homoskedastie, d. h.
$$\sigma_t^2 = \sigma^2, \quad t = 1, \ldots, T ,$$
zu
$$\mathbb{V}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{t=1}^{T}(x_t - \bar{x})^2} . \tag{11.3}$$

4. Die Verwendung der Formel (11.3) anstatt (11.2), falls tatsächlich Heteroskedastie vorliegt, führt zur Fehlschätzung der Varianz.

5. Die Fehlschätzung der Varianz kann in einer Über- oder einer Unterschätzung bestehen.

11-5

Bemerkung 11.3 (Fehlschätzung der Kovarianzmatrix)

1. Die Kovarianzmatrix des **gewöhnlichen** KQ-Schätzers bei Vorliegen von Heteroskedastie ist
$$\mathbb{V}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Delta}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \tag{11.4}$$
mit der Diagonalmatrix
$$\boldsymbol{\Delta} = \mathbb{V}(\mathbf{u}).$$

2. Diese wird durch den Schätzer der Kovarianzmatrix bei Homoskedastie,
$$\widehat{\mathbb{V}(\widehat{\boldsymbol{\beta}})} = \widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}, \tag{11.5}$$
fehlgeschätzt.

3. Statistische Inferenzprozeduren, wie die Interpretation von R^2, der t-Test für Parameter, der F-Test und die Bestimmung von Konfidenzintervallen für die Parameter, können systematisch verfälscht sein, falls sie auf der Schätzung (11.5) basieren.

11-6

11.2 Schätzen bei Heteroskedastie

11.2.1 Bekannte Varianzen oder Varianzverhältnisse

Bemerkung 11.4 (Bekannte Varianzen)

1. Bei **bekannten Varianzen** kann der VKQ-Schätzer
$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$
mit der Kovarianzmatrix
$$\mathbb{V}(\mathbf{u}) = \sigma^2\boldsymbol{\Omega}$$
eingesetzt werden.
2. Dabei ist z. B. $\sigma^2 = 1$ und $\boldsymbol{\Omega}$ ist eine Diagonalmatrix mit den Diagonalelementen $\sigma_t^2 > 0$ für $t = 1, \ldots, T$.

11-7

Bemerkung 11.5 (Bekannte Varianzverhältnisse)

1. Der VKQ-Schätzer kann auch eingesetzt werden, wenn σ^2 zwar **unbekannt** ist, aber Proportionalitätskonstanten oder Gewichte $k_1, \ldots, k_T$ mit
$$\sigma_t^2 = k_t\sigma^2, \quad t = 1, \ldots, T$$
spezifiziert sind.
2. Für die Kovarianzmatrix $\mathbb{V}(\mathbf{u}) = \sigma^2\boldsymbol{\Omega}$ ist dann $\boldsymbol{\Omega}$ eine **bekannte** Diagonalmatrix mit den Diagonalelementen k_t.

11-8

Bemerkung 11.6

1. Wenn $\mathbb{V}(\mathbf{u})$ eine Diagonalmatrix mit den Diagonalelementen $\sigma_t^2 > 0$ ist, gilt
$$\mathbb{V}(\mathbf{u})^{-1} = \mathbf{W},$$
wobei $\mathbf{W}$ eine Diagonalmatrix mit den Diagonalelementen
$$w_t \stackrel{\text{def}}{=} \frac{1}{\sigma_t^2}, \quad t = 1, \ldots, T$$
ist.
2. Damit spezialisiert sich der verallgemeinerte KQ-Schätzer zu
$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}\,.$$

11-9

Bemerkung 11.7 (Gewichteter KQ-Schätzer)

1. Mit
$$\mathbf{X}^* \stackrel{\text{def}}{=} \mathbf{W}^{1/2}\mathbf{X} = \left[\frac{x_{jt}}{\sigma_t}\right] \quad \text{und} \quad \mathbf{y}^* \stackrel{\text{def}}{=} \mathbf{W}^{1/2}\mathbf{y} = \left(\frac{y_1}{\sigma_1}, \frac{y_2}{\sigma_2}, \ldots, \frac{y_T}{\sigma_T}\right)'$$
ergibt sich
$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^* . \tag{11.6}$$

2. Der Schätzer $\widehat{\boldsymbol{\beta}}_{\text{GLS}}$ minimiert die gewichtete Quadratsumme
$$Q(\boldsymbol{\beta}) = \mathbf{u}'\mathbf{W}\mathbf{u} = \sum_{t=1}^{T} u_t^2 w_t = (\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta})'(\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta}) .$$

3. Daher wird der Schätzer aus (11.6) auch als **gewichteter** oder gewogener **KQ-Schätzer** bezeichnet (*weighted least squares estimator*).

11-10

4. Der verallgemeinerte KQ-Schätzer hat die Kovarianzmatrix
$$\mathbb{V}(\widehat{\boldsymbol{\beta}}_{\text{GLS}}) = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} .$$

11-11

11.2.2 Unbekannte Varianzen

Bemerkung 11.8 (Kovarianzmatrix des KQ-Schätzers)

1. Die Kovarianzmatrix des **gewöhnlichen** KQ-Schätzers bei Hetero-skedastie ist
$$\mathbb{V}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Delta}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \tag{11.7}$$
mit der Diagonalmatrix
$$\boldsymbol{\Delta} = \mathbb{V}(\mathbf{u}).$$

2. Die Diagonalmatrix $\boldsymbol{\Delta}$ lässt sich nicht konsistent schätzen, aber eine asymptotisch begründete Approximation ist
$$\mathbf{X}'\boldsymbol{\Delta}\mathbf{X} \approx \mathbf{X}'\widehat{\boldsymbol{\Delta}}\mathbf{X}, \tag{11.8}$$
wobei $\widehat{\boldsymbol{\Delta}}$ eine Diagonalmatrix mit den Diagonalelementen $\hat{u}_t^2$ ist und die $\hat{u}_t$ die geschätzten Residuen aus einer gewöhnlichen KQ-Schätzung sind.

11-12

Bemerkung 11.9 (Verfahren von White)

1. Das Verfahren, die Kovarianzmatrix des gewöhnlichen KQ-Schätzers bei Heteroskedastie dadurch zu schätzen, dass die Kovarianzmatrix (11.7) mit der Approximation (11.8) verwendet wird, ist als **Verfahren von White** bekannt.

2. Die resultierenden Schätzer
$$\widehat{\mathbb{V}(\widehat{\boldsymbol{\beta}})} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\boldsymbol{\Delta}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$
für die Kovarianzmatrix des gewöhnlichen KQ-Schätzers heißen auch **Heteroskedastie-konsistente Schätzer für die Kovarianzmatrix** (*heteroscedasticity-consistent covariance matrix estimators*) oder Heteroskedastie-robuste Schätzer für die Kovarianzmatrix.

3. Es handelt sich nicht um eine verbesserte Schätzung von $\boldsymbol{\beta}$, sondern um eine **verbesserte Schätzung der Kovarianzmatrix** für den gewöhnlichen KQ-Schätzer $\hat{\boldsymbol{\beta}}$.

11-13

11.3 Tests auf Heteroskedastie

Bemerkung 11.10 (Graphische Analyse)
Hinweise auf das Vorliegen von Heteroskedastie ergeben sich in der Regel durch

- eine **graphische Analyse der Originaldaten** und
- eine **graphische Analyse der KQ-Residuen**.

11-14

11.3.1 Park-Test

Bemerkung 11.11 (Testidee)

1. Die Testidee geht von dem Heteroskedastiemuster
$$\ln(\sigma_t^2) = \gamma_1 + \gamma_2 \ln(x_t), \quad t = 1, \ldots, T$$
aus.

2. Für
$$H_0 : \gamma_2 = 0$$
liegt Homoskedastie vor, da dann
$$\sigma_t^2 = e^{\gamma_1}, \quad t = 1, \ldots, T$$
gilt.

3. σ_t^2 ist nicht beobachtbar und wird durch $\hat{u}_t^2$ ersetzt, wobei $\widehat{\mathbf{u}}$ der Residuenvektor einer gewöhnlichen KQ-Regression ist.

11-15

Bemerkung 11.12 (Testdurchführung)

1. Der Park-Test überprüft im linearen Modell
$$\ln(\hat{u}_t^2) = \gamma_1 + \gamma_2 \ln(x_t) + v_t, \quad t = 1, \dots, T$$
mit der Annahme
$$\mathbf{v} \sim N_T(\mathbf{0}, \sigma_v \mathbf{I}) \tag{11.9}$$
die Hypothesen
$$H_0 : \gamma_2 = 0$$
versus
$$H_1 : \gamma_2 \neq 0.$$
2. Die Teststatistik
$$t = \frac{\hat{\gamma}_2}{\widehat{se}(\hat{\gamma}_2)}$$
ist unter H_0 t-verteilt mit $T-2$ Freiheitsgraden, dabei ist $\widehat{se}(\hat{\gamma}_2)$ die geschätzte Standardabweichung von γ_2.

11-16

Bemerkung 11.13 (Problematik des Park-Tests)

1. Der Test ist problematisch, da die Annahme 11.9 problematisch ist.
2. Die Abweichungen
$$v_t = \ln(\hat{u}_t^2) - \ln(\sigma_t^2)$$
sind selbst heteroskedastisch und autokorreliert.

11.3.2 Goldfeld-Quandt-Test

11-17

Bemerkung 11.14 (Testidee)

1. Die Testidee geht von dem Heteroskedastiemuster
$$\sigma_t^2 = \sigma^2 g(x_{jt}), \quad t = 1, \dots, T$$
aus. Dabei ist g eine fixierte Funktion, z. B.
$$g(x) = x, \quad g(x) = |x| \quad \text{oder} \quad g(x) = x^2,$$
und x_j ist **ein** Regressor.
2. Man bildet jeweils eine Gruppe von Beobachtungen mit kleinen und mit großen Varianzen.
3. In den beiden Gruppen werden getrennte Regressionen berechnet, die Varianzen der Fehlerterme geschätzt und diese Varianzen auf Gleichheit getestet.

11-18

Bemerkung 11.15 (Testdurchführung)

1. Ordne die Beobachtungen bzgl. der Größe der Varianzen.
2. Bilde zwei Gruppen unter Vernachlässigung der mittleren Beobachtungen.
3. In den beiden Gruppen werden getrennte Regressionen berechnet und die Varianzen geschätzt.
4. Mit einem F-Test werden die Varianzen auf Gleichheit getestet.

11.3.3 Breusch-Pagan-Test

11-19

Bemerkung 11.16 (Testidee)

1. Die Testidee geht von dem Heteroskedastiemuster

$$\sigma_t^2 = f(\alpha_1 + \alpha_2 z_{2t} + ... + \alpha_m z_{mt}), \quad t = 1, \ldots, T$$

   aus. Dabei sind die Variablen $z_2, \ldots, z_m$ eine Teilmenge der Variablen $x_2, \ldots, x_k$ mit $k \leq m$.
2. Für

$$H_0 : \alpha_2 = \ldots = \alpha_m = 0$$

   liegt Homoskedastie mit

$$\sigma_t^2 = f(\alpha_1), \quad t = 1, \ldots, T$$

   vor.
3. Die Funktion f muss nicht spezifiziert werden, sondern wird im Verfahren implizit geschätzt.

11-20

Bemerkung 11.17 (Testdurchführung)

1. Berechne eine gewöhnliche KQ-Regression und bilde daraus

$$\tilde{\sigma}^2 \stackrel{\text{def}}{=} \frac{\widehat{\mathbf{u}}\widehat{\mathbf{u}}'}{T} \quad \text{und} \quad p_t \stackrel{\text{def}}{=} \frac{\hat{u}_t^2}{\tilde{\sigma}^2}, \quad t = 1, \ldots, T\,.$$

2. Schätze die Hilfsregression

$$p_t = \alpha_1 + \alpha_2 z_{2t} + ... + \alpha_m z_{mt} + v_t\,.$$

3. Die Teststatistik

$$l = \frac{\sum_{t=1}^{T} (\hat{p}_t - \bar{p})^2}{2}$$

   ist asymptotisch chiquadratverteilt mit $m-1$ Freiheitsgraden.
4. Lehne H_0 ab, falls $l \geq \chi^2_{m-1,1-\alpha}$.
5. Große Schwankungen der $\hat{p}_t$-Werte weisen auf Heteroskedastie hin.

11.4 Ergänzungen

Bemerkung 11.a (Zur Asymptotik)

1. Vergleiche zu den folgenden Herleitungen Bemerkung 11.9.
2. Unter den Annahmen
$$\lim_{T\to\infty}\frac{1}{T}\mathbf{X}'\mathbf{X} = \mathbf{A} \in \mathbb{R}^{k\times k}, \quad \lim_{T\to\infty}\frac{1}{T}\mathbf{X}'\mathbf{\Delta}\mathbf{X} = \mathbf{B} \in \mathbb{R}^{k\times k}$$
folgt
$$\lim_{T\to\infty} T\mathbb{V}(\widehat{\boldsymbol{\beta}}) = \lim_{T\to\infty} T(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Delta}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$$
3. Für den KQ-Schätzer gilt die Konvergenz in Verteilung für $T \to \infty$
$$\mathbb{V}(\widehat{\boldsymbol{\beta}})^{-1/2}\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N_k(\mathbf{0}, \mathbf{I})$$
und die Approximation
$$\hat{\boldsymbol{\beta}} \approx N_k(\boldsymbol{\beta}, \mathbb{V}(\widehat{\boldsymbol{\beta}}))$$
für endliches T.
4. Unter den üblichen Annahmen ist
$$\frac{1}{T}\mathbf{X}'\widehat{\mathbf{\Delta}}\mathbf{X}\,.$$
ein konsistenter Schätzer für $\mathbf{B}$.
Daher ist $T(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{\Delta}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ ein konsistenter Schätzer für $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$. Mit
$$\widehat{\mathbb{V}(\widehat{\boldsymbol{\beta}})} \stackrel{\text{def}}{=} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{\Delta}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$
resultiert
$$\widehat{\mathbb{V}(\widehat{\boldsymbol{\beta}})}^{-1/2}\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N_k(\mathbf{0}, \mathbf{I})$$
und die Approximation
$$\hat{\boldsymbol{\beta}} \approx N_k(\boldsymbol{\beta}, \widehat{\mathbb{V}(\widehat{\boldsymbol{\beta}})})$$
für endliches T.

Kapitel 12

Multikollinearität

12-1

12. Multikollinearität

12-2

Bemerkung 12.1 (Multikollinearität)

1. Eine Annahme des Regressionsmodells ist
$$\mathbf{X}'\mathbf{X} \text{ ist invertierbar.}$$
Gibt es Probleme bezüglich dieser Annahme, so spricht man vom **Multikollinearitätsproblem**.

2. Wenn die Matrix $\mathbf{X}'\mathbf{X}$ nicht invertierbar ist, liegt **perfekte Multikollinearität** und damit lineare Abhängigkeit der Regressoren vor.

3. Wenn die Matrix $\mathbf{X}'\mathbf{X}$ zwar invertierbar ist, aber eine beinahe lineare Abhängigkeit der Regressoren vorliegt, spricht man von **imperfekter Multikollinearität**.

12-3

12.1 Perfekte Multikollinearität

Bemerkung 12.2
Die folgenden Aussagen sind für die Matrix $\mathbf{X} \in \mathbb{R}^{T \times k}$ äquivalent:

1. Die Matrix $\mathbf{X}'\mathbf{X}$ ist nicht invertierbar.
2. $\det(\mathbf{X}'\mathbf{X}) = 0$
3. $\text{Rang}(\mathbf{X}) < k$
4. Die Spalten der Matrix $\mathbf{X}$ sind linear abhängig.
5. Für die Eigenwerte $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k \geq 0$ der Matrix $\mathbf{X}'\mathbf{X}$ gilt $\lambda_k = 0$.

Bemerkung 12.3
Da die Spalten der Matrix $\mathbf{X}$ den Werten der Regressorvariablen entsprechen, sagt man im Fall der perfekten Multikollinearität auch:
Die Regressoren sind linear abhängig.

12-4

Bemerkung 12.4 (Eigenwerte von $\mathbf{X}'\mathbf{X}$)

1. Die Matrix $\mathbf{X}'\mathbf{X} \in \mathbb{R}^{k \times k}$ ist **symmetrisch**, d. h. es gilt
$$(\mathbf{X}'\mathbf{X})' = \mathbf{X}'\mathbf{X},$$
und **positiv semidefinit**, d. h. es gilt
$$\mathbf{x}'\mathbf{X}'\mathbf{X}\mathbf{x} \geq 0 \quad \text{für alle } \mathbf{x} \in \mathbb{R}^k,$$
daher sind alle Eigenwerte reellwertig und nichtnegativ.
2. Es gilt
$$\det(\mathbf{X}'\mathbf{X}) = \prod_{j=1}^{k} \lambda_j \quad \text{und} \quad \text{tr}(\mathbf{X}'\mathbf{X}) = \sum_{j=1}^{k} \lambda_j,$$
wobei die λ_j die Eigenwerte von $\mathbf{X}'\mathbf{X}$ bezeichnen.

12-5

Bemerkung 12.5 (Eigenwerte von $(\mathbf{X}'\mathbf{X})^{-1}$)

1. $\mathbf{X}'\mathbf{X}$ ist genau dann invertierbar und positiv definit, wenn alle Eigenwerte λ_j von $\mathbf{X}'\mathbf{X}$ positiv sind.
2. Wenn $\mathbf{X}'\mathbf{X}$ positiv definit ist, dann ist die inverse Matrix $(\mathbf{X}'\mathbf{X})^{-1}$ ebenfalls positiv definit und besitzt die Eigenwerte
$$\frac{1}{\lambda_j} > 0, \quad j = 1, \ldots, k$$
und es gilt
$$\det((\mathbf{X}'\mathbf{X})^{-1}) = \prod_{j=1}^{k} \frac{1}{\lambda_j}$$
und
$$\text{tr}((\mathbf{X}'\mathbf{X})^{-1}) = \sum_{j=1}^{k} \frac{1}{\lambda_j}\,.$$

12-6

Bemerkung 12.6 (Multikollinearität und KQ-Methode)
Im Fall perfekter Multikollinearität gilt:

1. Die Funktion
$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$
besitzt keine eindeutige Minimalstelle.
2. Die Minimalstellen sind genau durch die Lösungen der Normalgleichungen
$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$
beschrieben, die nicht eindeutig nach $\boldsymbol{\beta}$ auflösbar sind, da $\mathbf{X}'\mathbf{X}$ nicht invertierbar ist.
3. Der Parametervektor $\boldsymbol{\beta}$ ist mit der KQ-Methode nicht eindeutig schätzbar.

12-7

Bemerkung 12.7 (Nichtidentifizierbarkeit)

- Bei perfekter Multikollinearität gibt es bei gegebener Matrix $\mathbf{X}$ unterschiedliche Parametervektoren $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$, die zu demselben Vektor
$$\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2$$
und damit zu derselben Wahrscheinlichkeitsverteilung der Beobachtungen
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$
führen. Man sagt auch, diese Parametervektoren seien **beobachtungsäquivalent**.
- In diesem Fall spricht man von der **Nichtidentifizierbarkeit von Parametern**.

12-8

Beispiel 12.8 (Nichtidentifizierbarkeit)

1. Es sei
$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t, \quad t = 1, \ldots, T$$
mit
$$x_{1t} = x_{2t} \quad \text{für} \quad t = 1, \ldots, T.$$
2. In diesem Fall lassen sich der Parameter α und die Summe $\beta = \beta_1 + \beta_2$ durch Minimierung von
$$\sum_{t=1}^{T} (y_t - (\alpha + \beta_1 x_{1t} + \beta_2 x_{2t}))^2$$
schätzen. Jede Aufteilung von $\hat{\beta}$ mit $\hat{\beta} = \hat{\beta}_1 + \hat{\beta}_2$ minimiert die Quadratsumme, so dass die Parameter β_1 und β_2 nicht identifizierbar sind.
3. Alle Parametervektoren mit derselben Summe $\beta_1 + \beta_2$ sind beobachtungsäquivalent.

12-9

12.2 Imperfekte Multikollinearität

Bemerkung 12.9
Mit imperfekter Multikollinearität bezeichnet man Fälle, in denen eine **beinahe lineare Abhängigkeit** der Regressoren vorliegt. Die möglichen Folgen sind

1. **numerische Instabilität** bei der Invertierung von $\mathbf{X}'\mathbf{X}$, da
$$\det(\mathbf{X}'\mathbf{X}) \approx 0,$$
2. **große Standardfehler** bei der Parameterschätzung.

12-10

Definition 12.10 (Hilfsregressionen)
Gegeben sei eine Matrix $\mathbf{X}$ mit $\mathbf{X}'\mathbf{X}$ invertierbar. Die k Regressionen

$$\begin{aligned}
x_{1t} &= \alpha_{12}x_{2t} + \ldots + \alpha_{1k}x_{kt} + u_{1t}, \quad t = 1, \ldots, T \\
x_{2t} &= \alpha_{21}x_{1t} + \alpha_{23}x_{3t} + \ldots + \alpha_{2k}x_{kt} + u_{2t}, \quad t = 1, \ldots, T \\
&\ldots \\
x_{kt} &= \alpha_{k1}x_{1t} + \alpha_{k2}x_{2t} + \ldots + \alpha_{k,k-1}x_{k-1,t} + u_{kt}, \quad t = 1, \ldots, T
\end{aligned}$$

mit jeweils $k-1$ Regressoren heißen **Hilfsregressionen**.

Bemerkung 12.11
Die k Hilfsregressionen helfen bei der Aufdeckung beinahe linearer Abhängigkeiten zwischen den Regressoren.

12-11

Definition 12.12 (Multikollinearitätsindex)
Die Matrix $\mathbf{X}'\mathbf{X}$ sei invertierbar mit den Eigenwerten $\lambda_1, \ldots, \lambda_k$. Dann heißt

$$\mathrm{MKI}(\mathbf{X}'\mathbf{X}) \stackrel{\text{def}}{=} \frac{\lambda_{\max}}{\lambda_{\min}}$$

mit

$$\lambda_{\min} = \min_{j=1}^{k} \lambda_j \quad \text{und} \quad \lambda_{\max} = \max_{j=1}^{k} \lambda_j$$

der **Multikollinearitätsindex** von $\mathbf{X}'\mathbf{X}$.

Bemerkung 12.13

- Es gilt $\mathrm{MKI} \geq 1$.
- $\mathrm{MKI} = 1$ genau dann, wenn alle Eigenwerte gleich sind.
- Falls $\mathrm{MKI} > 1000$, spricht man von **starker Multikollinearität**, falls $100 \leq \mathrm{MKI} \leq 1000$, spricht man von **moderater Multikollinearität**.

12-12

12.3 Schätzen bei Multikollinearität

12.3.1 Ridge-Regression

Bemerkung 12.14

1. Eine Grundidee für die Schätzung bei imperfekter Multikollinearität ist die Verbesserung im Sinn eines kleineren mittleren quadratischen Fehlers.
2. Diese Verbesserung wird erreicht durch eine **kleinere Varianz** des Schätzers, im Vergleich zum gewöhnlichen KQ-Schätzer, wobei zugleich eine (geringe) **Verzerrung** zugelassen wird.
3. Die bekannteste Methodik mit dieser Idee ist die **Ridge-Regression**.

Definition 12.15 (Ridge-Schätzer)
Für $\kappa \geq 0$ heißt

$$\tilde{\boldsymbol{\beta}}(\kappa) = (\mathbf{X}'\mathbf{X} + \kappa\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$$

Ridge-Schätzer für den Parametervektor $\boldsymbol{\beta} \in \mathbb{R}^k$.

12-13

Bemerkung 12.16 (κ und k)

- Die hier mit dem griechischen Kleinbuchstaben Kappa bezeichnete Konstante wird häufig auch mit k bezeichnet.
- Dies wird hier vermieden, um eine Verwechselung mit der Dimension k des Parametervektors, d. h. der Anzahl der Regressoren, zu vermeiden.

Bemerkung 12.17

1. Für $\kappa = 0$ und eine invertierbare Matrix $\mathbf{X}'\mathbf{X}$ ergibt sich der gewöhnliche KQ-Schätzer als Spezialfall, d. h.
$$\tilde{\boldsymbol{\beta}}(0) = \widehat{\boldsymbol{\beta}}.$$
2. Für $\kappa = 0$ und eine nicht invertierbare Matrix $\mathbf{X}'\mathbf{X}$ ist $\tilde{\boldsymbol{\beta}}(\kappa)$ nicht definiert.
3. Die Matrix $\mathbf{X}'\mathbf{X} + \kappa\mathbf{I}$ mit $\kappa > 0$ ist auch dann invertierbar, wenn $\mathbf{X}'\mathbf{X}$ nicht invertierbar ist.

12-14

Satz 12.18 (Beziehung zum KQ-Schätzer)
Falls $\mathbf{X}'\mathbf{X}$ *invertierbar ist, gilt*

$$\tilde{\boldsymbol{\beta}}(\kappa) = (\mathbf{I} + \kappa(\mathbf{X}'\mathbf{X})^{-1})^{-1}\widehat{\boldsymbol{\beta}}\,.$$

Definition 12.19 (Verzerrung)
$\boldsymbol{\beta}^\star$ sei ein Schätzer für $\boldsymbol{\beta} \in \mathbb{R}^k$, dann heißt

$$\mathrm{Bias}(\boldsymbol{\beta}^\star) \stackrel{\text{def}}{=} \mathbb{E}(\boldsymbol{\beta}^\star) - \boldsymbol{\beta}$$

die **Verzerrung** (*bias*) des Schätzers $\boldsymbol{\beta}^\star$.

Satz 12.20 (Verzerrung des Ridge-Schätzers)
Es gilt

$$\mathbb{E}(\tilde{\boldsymbol{\beta}}(\kappa)) = (\mathbf{X}'\mathbf{X} + \kappa\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

und

$$\mathrm{Bias}(\tilde{\boldsymbol{\beta}}(\kappa)) = -\kappa(\mathbf{X}'\mathbf{X} + \kappa\mathbf{I})^{-1}\boldsymbol{\beta}\,.$$

12-15

Satz 12.21 (Kovarianzmatrix des Ridge-Schätzers)
Der Ridge-Schätzer $\tilde{\boldsymbol{\beta}}(\kappa)$ für $\boldsymbol{\beta} \in \mathbb{R}^k$ hat die Kovarianzmatrix

$$\mathbb{V}(\tilde{\boldsymbol{\beta}}(\kappa)) = \sigma^2(\mathbf{X}'\mathbf{X} + \kappa\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + \kappa\mathbf{I})^{-1}\,.$$

Definition 12.22 (Totalvarianz)
$\boldsymbol{\beta}^\star$ sei ein Schätzer für $\boldsymbol{\beta} \in \mathbb{R}^k$, dann heißt

$$\mathbb{V}_{tot}(\boldsymbol{\beta}^\star) \stackrel{\text{def}}{=} \operatorname{tr}(\mathbb{V}(\boldsymbol{\beta}^\star))$$

die **Totalvarianz** (oder totale Variation) von $\boldsymbol{\beta}^\star$.

12-16

Satz 12.23 (Totalvarianz des KQ- und des Ridge-Schätzers)
Die Eigenwerte der Matrix $\mathbf{X}'\mathbf{X}$ seien $\lambda_1, \ldots, \lambda_k$.

1. *Der Ridge-Schätzer $\tilde{\boldsymbol{\beta}}(\kappa)$ hat die Totalvarianz*

$$\mathbb{V}_{tot}(\tilde{\boldsymbol{\beta}}(\kappa)) = \sigma^2 \sum_{i=1}^{k} \frac{\lambda_i}{(\lambda_i + \kappa)^2}\,.$$

2. *Es gilt*

$$\lim_{\kappa\to\infty} \mathbb{V}_{tot}(\tilde{\boldsymbol{\beta}}(\kappa)) = 0\,.$$

3. *Falls $\mathbf{X}'\mathbf{X}$ invertierbar ist, sind alle λ_i positiv, der KQ-Schätzer hat die Totalvarianz*

$$\mathbb{V}_{tot}(\widehat{\boldsymbol{\beta}}) = \mathbb{V}_{tot}(\tilde{\boldsymbol{\beta}}(0)) = \sigma^2 \sum_{i=1}^{k} \frac{1}{\lambda_i}$$

   und für $\kappa > 0$ gilt

$$\mathbb{V}_{tot}(\tilde{\boldsymbol{\beta}}(\kappa)) < \mathbb{V}_{tot}(\tilde{\boldsymbol{\beta}}(0))\,.$$

12-17

Definition 12.24 (MSE)
$\beta_j^\star$ sei ein Schätzer für $\beta_j \in \mathbb{R}$, dann heißt

$$\operatorname{MSE}(\beta_j^\star) \stackrel{\text{def}}{=} \mathbb{E}[(\beta_j^\star - \beta_j)^2]$$

mittlerer quadratischer Fehler oder **MSE** (*mean squared error*) des Schätzers $\beta_j^\star$ für β_j.

Satz 12.25 (MSE-Zerlegung)
Es gilt

$$\operatorname{MSE}(\beta_j^\star) = \mathbb{V}(\beta_j^\star) + (\operatorname{Bias}(\beta_j^\star))^2\,.$$

12-18

Definition 12.26 (Matrizieller MSE)
$\boldsymbol{\beta}^\star$ sei ein Schätzer für den Parametervektor $\boldsymbol{\beta} \in \mathbb{R}^k$, dann heißt

$$\mathrm{MMSE}(\boldsymbol{\beta}^\star) \stackrel{\text{def}}{=} \mathbb{E}[(\boldsymbol{\beta}^\star - \boldsymbol{\beta})(\boldsymbol{\beta}^\star - \boldsymbol{\beta})']$$

matrizieller MSE von $\boldsymbol{\beta}^\star$ und

$$\mathrm{SMSE}(\boldsymbol{\beta}^\star) \stackrel{\text{def}}{=} \mathrm{tr}(\mathrm{MMSE}(\boldsymbol{\beta}^\star))$$

heißt **skalarer MSE** von $\boldsymbol{\beta}^\star$.

Satz 12.27 (Zerlegungen des MMSE und SMSE)
Es gilt

$$\mathrm{MMSE}(\boldsymbol{\beta}^\star) = \mathbb{V}(\boldsymbol{\beta}^\star) + \mathrm{Bias}(\boldsymbol{\beta}^\star)\mathrm{Bias}(\boldsymbol{\beta}^\star)'$$

und

$$\mathrm{SMSE}(\boldsymbol{\beta}^\star) = \mathbb{V}_{tot}(\boldsymbol{\beta}^\star) + \mathrm{Bias}(\boldsymbol{\beta}^\star)'\mathrm{Bias}(\boldsymbol{\beta}^\star).$$

12-19

Satz 12.28 (Zur Wahl von κ)
$\tilde{\boldsymbol{\beta}}(\kappa)$ *bezeichne den Ridge-Schätzer.*

1. *Dann gilt*

$$\frac{\mathrm{d}\mathbb{V}_{tot}(\tilde{\boldsymbol{\beta}}(\kappa))}{\mathrm{d}\kappa} < 0$$

   und

$$\frac{\mathrm{d}\mathrm{Bias}(\tilde{\boldsymbol{\beta}}(\kappa))'\mathrm{Bias}(\tilde{\boldsymbol{\beta}}(\kappa))}{\mathrm{d}\kappa} > 0\,.$$

2. *Falls* $\mathbf{X}'\mathbf{X}$ *invertierbar ist, existiert* $\kappa^* > 0$ *mit*

$$\mathrm{SMSE}(\tilde{\boldsymbol{\beta}}(\kappa)) < \mathrm{SMSE}(\widehat{\boldsymbol{\beta}}), \quad 0 < \kappa \le \kappa^*\,.$$

12-20

Bemerkung 12.29

1. Die praktische Anwendung des Ergebnisses aus Satz 12.28 ist deswegen nicht einfach, weil κ^* von den unbekannten Parametern $\boldsymbol{\beta}$ und σ^2 abhängt.

2. Es gibt Vorschläge, in einem zweistufigen Verfahren erst κ^* aus den Beobachtungen zu schätzen und dann mit diesem geschätzten κ^* einen Ridge-Schätzer zu berechnen, vgl. Judge (1988, S. 880-882). Diese Schätzer werden **adaptive Ridge-Schätzer** (*adaptive ridge estimators*) genannt.

3. Im Satz 12.28 ist κ^* nicht stochastisch. Wird κ^* aus den Daten geschätzt, so ist ein Gewinn im Sinn eines kleineren SMSE nicht mehr garantiert, vgl. Judge (1988, S. 880).

12-21

Bemerkung 12.30 (Ridge-Trace)
Eine graphische Darstellung von $\tilde{\beta}_j(\kappa)$ und der zugehörigen Summe der quadrierten Residuen für $\kappa \geq 0$ nennt man **Ridge-Trace** für den Parameter β_j. Diese veranschaulicht die Änderung des Schätzwertes für β_j in Abhängigkeit von κ.

12-22

Bemerkung 12.31 (Länge eines Vektors)

1. Die Länge eines Vektors $\mathbf{x} \in \mathbb{R}^k$ ist $\sqrt{\mathbf{x}'\mathbf{x}} \geq 0$.
2. $\sqrt{\mathbf{x}'\mathbf{x}} \leq \sqrt{\mathbf{y}'\mathbf{y}}$ ist äquivalent zu $\mathbf{x}'\mathbf{x} \leq \mathbf{y}'\mathbf{y}$.

Satz 12.32 (Schrumpfungseigenschaft)
$\mathbf{X}'\mathbf{X}$ *sei invertierbar.* $\tilde{\boldsymbol{\beta}}(\kappa)$ *bezeichne den Ridge-Schätzer und* $\widehat{\boldsymbol{\beta}}$ *den gewöhnlichen KQ-Schätzer für den Parametervektor* $\boldsymbol{\beta} \in \mathbb{R}^k$.

1. *Für* $\kappa > 0$ *und* $\widehat{\boldsymbol{\beta}} \neq \mathbf{0}$ *gilt*
$$0 < \tilde{\boldsymbol{\beta}}(\kappa)'\tilde{\boldsymbol{\beta}}(\kappa) < \widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}}\,.$$
2. *Es gilt*
$$\lim_{\kappa\to\infty} \tilde{\boldsymbol{\beta}}(\kappa)'\tilde{\boldsymbol{\beta}}(\kappa) = 0\,.$$

12-23

Bemerkung 12.33

1. Diese Eigenschaften motivieren die Bezeichnung des Ridge-Schätzers als **Schrumpfschätzer** (*shrinkage estimator*).
2. Die einfachste Form eines Schrumpfschätzers ist bei invertierbarer Matrix $\mathbf{X}'\mathbf{X}$ der Schätzer $\lambda\widehat{\boldsymbol{\beta}}$ mit einer Konstanten $0 < \lambda < 1$. Es gilt dann
$$(\lambda\widehat{\boldsymbol{\beta}})'\lambda\widehat{\boldsymbol{\beta}} = \lambda^2\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}} < \widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}}\,.$$

12.3.2 Hauptkomponentenschätzer

12-24

Bemerkung 12.34
Wenn $\mathbf{X}'\mathbf{X}$ nicht invertierbar ist (perfekte Multikollinearität), so kann anstelle von $(\mathbf{X}'\mathbf{X})^{-1}$ die **Moore-Penrose-Inverse** (oder **eindeutige verallgemeinerte Inverse**) $(\mathbf{X}'\mathbf{X})^{+}$ und anstelle des gewöhnlichen KQ-Schätzers

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

der Schätzer

$$\widehat{\boldsymbol{\beta}}_{\mathrm{HK}} \stackrel{\mathrm{def}}{=} (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{y}$$

gebildet werden.

12-25

Satz 12.35
$\widehat{\boldsymbol{\beta}}_{\mathrm{HK}}$ *ist eine spezielle Lösung der Normalgleichungen*

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

und eine Minimalstelle von

$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\,.$$

12-26

Bemerkung 12.36

1. Eine Spektralzerlegung

$$\mathbf{X}'\mathbf{X} = \sum_{j=1}^{k} \lambda_j \mathbf{q}_j \mathbf{q}_j'$$

   mit fallend geordneten Eigenwerten $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq 0$ und normierten und orthogonalen Eigenvektoren $\mathbf{q}_1, \ldots, \mathbf{q}_k$ sei gegeben.

2. Wenn $\mathbf{X}'\mathbf{X}$ invertierbar ist, gilt

$$(\mathbf{X}'\mathbf{X})^{-1} = \sum_{j=1}^{k} \lambda_j^{-1} \mathbf{q}_j \mathbf{q}_j'\,.$$

12-27

3. Wenn $\mathbf{X}'\mathbf{X}$ nicht invertierbar ist, ist

$$(\mathbf{X}'\mathbf{X})^{+} = \sum_{j=1}^{r} \lambda_j^{-1} \mathbf{q}_j \mathbf{q}_j'$$

die Moore-Penrose-Inverse von $\mathbf{X}'\mathbf{X}$, wobei λ_r der kleinste positive Eigenwert von $\mathbf{X}'\mathbf{X}$ ist.

12-28

Definition 12.37 (Hauptkomponentenschätzer)
Es sei eine Spektralzerlegung

$$\mathbf{X}'\mathbf{X} = \sum_{j=1}^{k} \lambda_j \mathbf{q}_j \mathbf{q}_j'$$

mit fallend geordneten Eigenwerten $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq 0$ und normierten und orthogonalen Eigenvektoren $\mathbf{q}_j$ gegeben. Für jedes r mit $1 \leq r \leq k$ und $\lambda_r > 0$ heißt der Schätzer

$$\widehat{\boldsymbol{\beta}}_r \stackrel{\text{def}}{=} \mathbf{A}_r^{+} \mathbf{X}'\mathbf{y}$$

mit

$$\mathbf{A}_r^{+} \stackrel{\text{def}}{=} \sum_{j=1}^{r} \lambda_j^{-1} \mathbf{q}_j \mathbf{q}_j'$$

Hauptkomponentenschätzer von $\boldsymbol{\beta}$.

12-29

Bemerkung 12.38

1. Es gilt

$$\mathbf{X}'\mathbf{X} = \mathbf{A}_r + \sum_{j=r+1}^{k} \lambda_j \mathbf{q}_j \mathbf{q}_j'$$

mit

$$\mathbf{A}_r \stackrel{\text{def}}{=} \sum_{j=1}^{r} \lambda_j \mathbf{q}_j \mathbf{q}_j' .$$

2. $\mathbf{A}_r^{+}$ ist die Moore-Penrose-Inverse von $\mathbf{A}_r$.
3. Wenn λ_r der kleinste positive Eigenwert ist, gilt $\widehat{\boldsymbol{\beta}}_r = \widehat{\boldsymbol{\beta}}_{\text{HK}}$.
4. Falls $\mathbf{X}'\mathbf{X}$ invertierbar ist, gilt $\mathbf{A}_k^{+} = (\mathbf{X}'\mathbf{X})^{-1}$ und $\widehat{\boldsymbol{\beta}}_k = \widehat{\boldsymbol{\beta}}_{\text{HK}} = \widehat{\boldsymbol{\beta}}$.
5. Für $r < k$ ist $\widehat{\boldsymbol{\beta}}_r$ verzerrt.
6. Falls $\mathbf{X}'\mathbf{X}$ invertierbar ist, gilt $\widehat{\boldsymbol{\beta}}_r'\widehat{\boldsymbol{\beta}}_r < \widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}}$ für $r < k$.

12.4 Ergänzungen

Bemerkung 12.a (Beweis von Satz 12.18)

$$\begin{aligned}\tilde{\boldsymbol{\beta}}(\kappa) &= (\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}\\ &= \left(\mathbf{X}'\mathbf{X}(\mathbf{I}+\kappa(\mathbf{X}'\mathbf{X})^{-1})\right)^{-1}\mathbf{X}'\mathbf{y}\\ &= (\mathbf{I}+\kappa(\mathbf{X}'\mathbf{X})^{-1})^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\\ &= (\mathbf{I}+\kappa(\mathbf{X}'\mathbf{X})^{-1})^{-1}\widehat{\boldsymbol{\beta}}\end{aligned}$$

Bemerkung 12.b (Beweis von Satz 12.20)
Es gilt

$$\mathrm{Bias}(\tilde{\boldsymbol{\beta}}(\kappa)) = \mathbb{E}(\tilde{\boldsymbol{\beta}}(\kappa)) - \boldsymbol{\beta} = [(\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}-\mathbf{I}]\boldsymbol{\beta}\,.$$

Aus

$$(\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X}+\kappa\mathbf{I}) = \mathbf{I}$$

erhält man

$$(\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}-\mathbf{I} = -\kappa(\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}$$

und somit die angegebene Verzerrung.

Bemerkung 12.c (Beweis von Satz 12.21)

$$\begin{aligned}\mathbb{V}(\tilde{\boldsymbol{\beta}}(\kappa)) &= \mathbb{V}((\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}\mathbf{X}'\mathbf{y})\\ &= (\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}\mathbf{X}'\mathbb{V}(\mathbf{y})((\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}\mathbf{X}')'\\ &= (\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}\\ &= \sigma^2(\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X}+\kappa\mathbf{I})^{-1}\end{aligned}$$

Bemerkung 12.d
Der Satz 12.32 ergibt sich als Spezialfall des folgenden allgemeineren Satzes.

Satz 12.e (Längenschrumpfung)
Es sei $\mathbf{A} \in \mathbb{R}^{n\times n}$ *eine positiv definite Matrix,* $\mathbf{x} \in \mathbb{R}^n$ *und* $\kappa > 0$*. Für*

$$\mathbf{y} = (\mathbf{I}_n + \kappa\mathbf{A})^{-1}\mathbf{x}$$

gilt

$$\mathbf{y}'\mathbf{y} \le \mathbf{x}'\mathbf{x},$$

$$\mathbf{y}'\mathbf{y} < \mathbf{x}'\mathbf{x}, \quad \textit{falls} \quad \mathbf{x} \neq \mathbf{0}$$

und

$$\lim_{\kappa\to\infty} \mathbf{y}'\mathbf{y} = 0\,.$$

Bemerkung 12.f (Beweis von Satz 12.e)

1. Aus

   $$\mathbf{y} = (\mathbf{I}_n + \kappa\mathbf{A})^{-1}\mathbf{x}$$

   folgt

   $$\mathbf{x} = (\mathbf{I}_n + \kappa\mathbf{A})\mathbf{y}$$

   und daraus

   $$\mathbf{x}'\mathbf{x} = \mathbf{y}'(\mathbf{I}_n+\kappa\mathbf{A})'(\mathbf{I}_n+\kappa\mathbf{A})\mathbf{y} = \mathbf{y}'\mathbf{y} + 2\kappa\mathbf{y}'\mathbf{A}\mathbf{y} + \kappa^2(\mathbf{A}\mathbf{y})'\mathbf{A}\mathbf{y}\,.$$

   Die letzten beiden Summanden sind nichtnegativ, also gilt $\mathbf{y}'\mathbf{y} \le \mathbf{x}'\mathbf{x}$.

2. Da $\mathbf{A}$ invertierbar ist, sind die letzten beiden Summanden genau dann gleich Null, wenn $\mathbf{y} = \mathbf{0}$. Es ist $\mathbf{y} = \mathbf{0}$ genau dann, wenn $\mathbf{x} = \mathbf{0}$.

3. Die Matrix
$$\mathbf{B} = (\mathbf{I}_n + \kappa \mathbf{A})^{-1}$$
ist positiv definit mit den Eigenwerten
$$\lambda_i = \frac{1}{1 + \kappa \lambda_i^A},$$
wenn λ_i^A die Eigenwerte von $\mathbf{A}$ sind. Aus der Spektralzerlegung $\mathbf{B} = \mathbf{P\Lambda P}'$ folgt
$$\mathbf{y}'\mathbf{y} = \mathbf{x}'\mathbf{B}^2\mathbf{x} = \mathbf{x}^{*\prime}\mathbf{\Lambda}^2\mathbf{x}^* = \sum_{i=1}^{n} x_i^{*2} \frac{1}{(1 + \kappa \lambda_i^A)^2}$$
mit $\mathbf{x}^* = \mathbf{P}'\mathbf{x}$. Da $\lambda_i^A > 0$ für $i = 1, \ldots, n$ ergibt sich für $\kappa \to \infty$ die letzte Behauptung des Satzes.

Bemerkung 12.g (Beweis von Satz 12.35)

1. Das System der Normalgleichungen $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ besitzt mindestens eine Lösung. Damit ist $\boldsymbol{\beta}^+$ eine Lösung der Normalgleichungen, vgl. Satz 15.32.
2. Die Funktion $Q(\boldsymbol{\beta})$ ist konvex in $\boldsymbol{\beta}$, so dass die Erfüllung der sogenannten Normalgleichungen
$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_j} = 0, \quad j = 1, \ldots, k,$$
notwendig und hinreichend für das Vorliegen eines Minimums ist.

Bemerkung 12.h (Zur KQ-Schätzung)

1. Wenn $\mathbf{X}'\mathbf{X}$ invertierbar ist, dann besitzen die Normalgleichungen $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ genau eine Lösung und $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ ist diese Lösung.
2. Wenn $\mathbf{X}'\mathbf{X}$ nicht invertierbar ist, dann besitzen die Normalgleichungen $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ mindestens eine Lösung und $\widehat{\boldsymbol{\beta}}_{\mathrm{HK}} = \mathbf{X}^+\mathbf{y} = (\mathbf{X}'\mathbf{X})^+\mathbf{X}'\mathbf{y}$ ist diejenige Lösung mit der kürzesten Länge $(\mathbf{x}'\mathbf{x})^{1/2}$.

Kapitel 13

Regression mit Dummy-Regressoren

13-1

13. Regression mit Dummy-Regressoren

13-2

13.1 Dummy-Regressoren

Bemerkung 13.1 (Dummy-Regressoren)

- Wenn eine Variable nur zwei Werte (in der Regel 0 oder 1) annehmen kann, spricht man von **binären Variablen**, **dichotomen Variablen**, **0-1-Variablen** oder **Dummy-Variablen**.
- Beispiel: männlich / weiblich
- Allgemeiner: **kategoriale Variable** mit K Ausprägungen $x_{jt} \in \{1, 2, \ldots, K\}$
- Erklärende Variablen (Regressoren) von der Form

$$x_{jt} \in \{0, 1\}$$

werden im Folgenden als **Dummy-Regressoren** bezeichnet.

13.1.1 Ein Dummy-Regressor

13-3

Bemerkung 13.2

- Es folgen fünf Fälle für Verknüpfungen von einer erklärten Variablen (y), einem Regressor (x) und einem Dummy-Regressor (D):

  1. $y_t = \alpha_1 + \beta_1 x_t + u_t$
  2. $y_t = \alpha_1 + \alpha_2 D_t + u_t$
  3. $y_t = \alpha_1 + \beta_1 x_t + \alpha_2 D_t + u_t$
  4. $y_t = \alpha_1 + \beta_1 x_t + \beta_2 D_t x_t + u_t$
  5. $y_t = \alpha_1 + \beta_1 x_t + \alpha_2 D_t + \beta_2 D_t x_t + u_t$

- y_t steht für Sparen (gesparter Teil des Einkommens)
- x_t steht für Einkommen
- der **Dummy-Regressor** D_t steht für die Variable Geschlecht:
  $D_t = 0$, falls die Beobachtungseinheit t männlich ist,
  $D_t = 1$, falls die Beobachtungseinheit t weiblich ist.

13-4

Bemerkung 13.3 (Fall 1)
Es sind drei lineare Regressionen von Sparen (y) auf Einkommen (x) der Art

$$y_t = \alpha_1 + \beta_1 x_t + u_t$$

denkbar:

- in der Gruppe der männlichen Beobachtungseinheiten ($D_t = 0$),
- in der Gruppe der weiblichen Beobachtungseinheiten ($D_t = 1$),
- für alle Beobachtungseinheiten ($D_t = 0$ oder $D_t = 1$).

13-5

Bemerkung 13.4 (Fall 2)

- Regression von Sparen (y) auf Dummy-Regressor Geschlecht (D)
- Regressionsmodell:
$$y_t = \alpha_1 + \alpha_2 D_t + u_t$$
- Regressionsschätzung:
$$\hat{y}_t = \hat{\alpha}_1 + \hat{\alpha}_2 D_t$$
- Interpretation der Regressionsschätzung:
  - $\bar{y}$ ist der mittlere Sparbetrag aller Beobachtungseinheiten.
  - $\hat{\alpha}_1$ ist der mittlere Sparbetrag, falls $D_t = 0$.
  - $\hat{\alpha}_1 + \hat{\alpha}_2$ ist der mittlere Sparbetrag, falls $D_t = 1$.

13-6

Bemerkung 13.5 (Fall 3)

- Regression von Sparen (y) auf Einkommen (x) und Geschlecht (D)

$$y_t = \alpha_1 + \beta_1 x_t + \alpha_2 D_t + u_t$$

- Regressionsschätzung:

$$\hat{y}_t = \hat{\alpha}_1 + \hat{\beta}_1 x_t + \hat{\alpha}_2 D_t$$

- Interpretation der Regressionsschätzung:

$$\hat{y}_t = \hat{\alpha}_1 + \hat{\beta}_1 x_t, \quad \text{falls} \quad D_t = 0$$

$$\hat{y}_t = (\hat{\alpha}_1 + \hat{\alpha}_2) + \hat{\beta}_1 x_t, \quad \text{falls} \quad D_t = 1$$

- Der Dummy-Regressor D_t kann in diesem Fall Niveauunterschiede, aber keine unterschiedlichen Steigungen erklären.
- $\hat{\alpha}_2$ heißt **Differential-Absolutglied** (*differential intercept*).

13-7

Bemerkung 13.6 (Fall 4)

- Regression von Sparen (y) auf Einkommen (x) mit **Interaktionsterm** (xD)
$$y_t = \alpha_1 + \beta_1 x_t + \beta_2 x_t D_t + u_t$$
- Regressionsschätzung:

$$y_t = \hat{\alpha}_1 + \hat{\beta}_1 x_t + \hat{\beta}_2 x_t D_t$$

- Interpretation der Regressionsschätzung:

$$\hat{y}_t = \hat{\alpha}_1 + \hat{\beta}_1 x_t, \quad \text{falls} \quad D_t = 0$$

$$\hat{y}_t = \hat{\alpha}_1 + (\hat{\beta}_1 + \hat{\beta}_2) x_t, \quad \text{falls} \quad D_t = 1$$

- Der Dummy-Regressor D_t kann in diesem Fall zwar Steigungsunterschiede, aber keine unterschiedlichen Niveauparameter erklären.
- $\hat{\beta}_2$ heißt **Differential-Steigung** (*differential slope coefficient*).

13-8

Bemerkung 13.7 (Fall 5)

- Regression von Sparen (y) auf Einkommen (x) und Geschlecht (D) mit **Interaktionsterm** (xD)

$$y_t = \alpha_1 + \beta_1 x_t + \alpha_2 D_t + \beta_2 x_t D_t + u_t$$

- Regressionsschätzung:

$$y_t = \hat{\alpha}_1 + \hat{\beta}_1 x_t + \hat{\alpha}_2 D_t + \hat{\beta}_2 x_t D_t$$

- Interpretation der Regressionsschätzung:

$$\hat{y}_t = \hat{\alpha}_1 + \hat{\beta}_1 x_t, \quad \text{falls} \quad D_t = 0$$

$$\hat{y}_t = (\hat{\alpha}_1 + \hat{\alpha}_2) + (\hat{\beta}_1 + \hat{\beta}_2) x_t, \quad \text{falls} \quad D_t = 1$$

- Der Dummy-Regressor D_t kann in diesem Fall Niveau- **und** Steigungsunterschiede erklären.

13.1.2 Ein kategorialer Regressor

13-9

Bemerkung 13.8
Eine Variable mit m Kategorien wird nicht durch m, sondern durch $m-1$ binäre Variablen kodiert.

13-10

Beispiel 13.9
Die Variable Religion sei mit den $m = 3$ Kategorien

christlich, mohammedanisch, sonstige

erschöpfend beschrieben. Es werden **zwei** binäre Variablen gebildet

- $D_{1t} = 1$, falls t christlich ist, $D_{1t} = 0$, falls t nicht christlich ist
- $D_{2t} = 1$, falls t mohammedanisch ist, $D_{2t} = 0$, falls t nicht mohammedanisch ist

Die zwei Dummy-Regressoren D_1 und D_2 kodieren **drei** Kategorien

- $(D_{1t}, D_{2t}) = (1, 0)$, falls t christlich
- $(D_{1t}, D_{2t}) = (0, 1)$, falls t mohammedanisch
- $(D_{1t}, D_{2t}) = (0, 0)$, falls t weder christlich noch mohammedanisch

13-11

Bemerkung 13.10 (Multikollinearitätsfalle)

1. Bei einer Kodierung durch drei binäre Variablen
   - $(D_{1t}, D_{2t}, D_{3t}) = (1, 0, 0)$, falls t christlich
   - $(D_{1t}, D_{2t}, D_{3t}) = (0, 1, 0)$, falls t mohammedanisch
   - $(D_{1t}, D_{2t}, D_{3t}) = (0, 0, 1)$, falls t weder christlich noch mohammedanisch

   gilt
   $$\sum_{j=1}^{3} D_{jt} = 1, \quad t = 1, \ldots, T\,.$$
   Damit liegt **perfekte Multikollinearität** vor, falls eine Regression mit Absolutglied berechnet wird.
2. Zur **Vermeidung der Multikollinearitätsfalle** ist eine Kodierung durch zwei anstatt durch drei Dummy-Regressoren oder die Schätzung einer Regression **ohne** Absolutglied erforderlich.

13.1.3 Mehrere Dummy-Regressoren

13-12

Bemerkung 13.11

1. Wenn zwei Dummy-Regressoren für zwei Merkmale stehen, die kombiniert auftreten können, können **multiplikative Interaktionsterme** berücksichtigt werden.

2. Z. B.
$$y_t = \alpha_1 + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{2t} D_{3t} + \beta_1 x_t + u_t,$$
wobei $D_{2t}D_{3t} = 1$ genau dann, wenn $D_{2t} = D_{3t} = 1$.

3. Der Parameter α_4 misst den Niveaueffekt, der durch die Kombination der Merkmalswerte $D_{2t} = 1$ und $D_{3t} = 1$ auftritt.

4. Analog können Steigungseffekte durch Interaktionsterme der Form $D_{2t}D_{3t}x_t$ berücksichtigt werden.

13-13

13.2 Strukturbrüche und Saisonmuster

Bemerkung 13.12 (Strukturbruch)

1. Ein Dummy-Regressor kann zur Modellierung eines **Strukturbruches** (*structural change*) in den Parametern verwendet werden.

2. Ein Strukturbruch beim Übergang von Jahr t_0 zu $t_0 + 1$ liegt vor, wenn z. B. im Fall der linearen Einfachregression
$$y_t = \gamma_1 + \gamma_2 x_t + u_t, \quad t = 1, \ldots, t_0$$
und
$$y_t = \delta_1 + \delta_2 x_t + u_t, \quad t = t_0 + 1, \ldots, T$$
mit unterschiedlichen Parameterwerten gilt.

13-14

3. Mit
$$D_t = 0 \quad \text{für} \quad t = 1, \ldots, t_0$$
und
$$D_t = 1 \quad \text{für} \quad t = t_0, t_0 + 1 \ldots, T$$
kann ein Strukturbruch an der bekannten Stelle t_0 durch den Regressionsansatz
$$y_t = \alpha_1 + \alpha_2 D_t + \beta_1 x_t + \beta_2 x_t D_t + u_t$$
abgebildet werden. Dabei gilt
$$\alpha_1 = \gamma_1, \quad \alpha_1 + \alpha_2 = \delta_1, \quad \beta_1 = \gamma_2, \quad \beta_1 + \beta_2 = \delta_2.$$

4. Man kann **testen**, **ob ein Strukturbruch** an der Stelle t_0 **vorliegt**, indem man testet, ob die Parameter α_2 und β_2 signifikant von Null verschieden sind.

5. Ein alternativer Test auf das Vorliegen eines Strukturbruches ist der **Chow-Test**, vgl. Gujarati (S. 273, 2003).

13-15

Bemerkung 13.13 (Chow-Test: Grundidee)

1. Der Chow-Test geht von homoskedastischen normalverteilten Residuen über den gesamten Beobachtungszeitraum aus.

2. Es werden drei Regressionen berechnet, aus denen sich drei Summen quadrierter Residuen ergeben, $\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1$ und $\hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2$ für zwei Gruppen von Beobachtungen vor und nach der Stelle, an der ein Strukturbruch vermutet wird, sowie $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ aus allen T Beobachtungen.

3. Es gilt immer
$$\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2 \leq \hat{\mathbf{u}}'\hat{\mathbf{u}},$$
da sich zwei Regressionen (mit insgesamt vier Parametern) besser an die Daten anpassen, als eine Regression mit zwei Parametern.

13-16

4. Ohne Strukturbruch ist
$$\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2 \approx \hat{\mathbf{u}}'\hat{\mathbf{u}}$$
zu erwarten. Mit Strukturbruch ist
$$\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2 < \hat{\mathbf{u}}'\hat{\mathbf{u}}$$
zu erwarten.

13-17

Bemerkung 13.14 (Chow-Test: Testdurchführung)

1. Es werden drei Regressionen berechnet, aus denen sich drei Summen quadrierter Residuen ergeben, $\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1$ und $\hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2$ für die beiden Gruppen von Beobachtungen vor und nach dem Strukturbruch, sowie $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ aus allen T Beobachtungen.

2. Die Nullhypothese ist H_0 : Es liegt kein Strukturbruch vor.

3. Die Teststatistik
$$F = \frac{\dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}} - (\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2)}{k}}{\dfrac{\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2}{T-2k}}$$
ist, falls H_0 richtig ist, F-verteilt mit k und $T-2k$ Freiheitsgraden.

4. H_0 wird für große Werte von F abgelehnt.

5. Eine kritische Voraussetzung des Chow-Tests ist, dass die Varianzen in beiden Beobachtungsphasen gleich sind.

13-18

Bemerkung 13.15 (Saisonmuster)

1. Dummy-Regressoren können verwendet werden, um saisonale, sich wiederholende Muster in den Daten abzubilden.

2. Dies können z. B. bei Monatsdaten **sich** jährlich **wiederholende Ereignisse** sein (Weihnachtsverkauf, Umsatzeinbruch in Ferienzeit usw.), die durch einen zusätzlichen Dummy-Regressor für einen bestimmten Monat abgebildet werden.

3. Eine andere Anwendung ist die Modellierung **sich wiederholender Saisonfiguren**. Beispielsweise bei Quartalsdaten durch den Ansatz

   $$y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + u_t, \quad t = 1, \ldots, T\,,$$

   wobei $D_{jt} = 1$ im j-ten Quartal und $D_{kt} = 0$ in den übrigen Quartalen ($k \neq j$). Der Parameter α_j ist der Erwartungswert der Daten im jeweils j-ten Quartal,

   $$\alpha_j = \mathbb{E}[y_t | D_{jt} = 1]\,.$$

Kapitel 14

Regression mit dichotomen endogenen Variablen

14-1

14. Regression mit dichotomen endogenen Variablen

14-2

14.1 Dichotome endogene Variablen

Bemerkung 14.1

1. Betrachtet wird eine **dichotome** oder binäre **endogene Variable** (dichotomer Regressand) mit $y_t \in \{0, 1\}$.
2. Es gilt
$$P(y_t = 0) + P(y_t = 1) = 1\,.$$
3. Mit
$$p_t \stackrel{\text{def}}{=} P(y_t = 1)$$
gilt
$$P(y_t = 0) = 1 - p_t.$$
4. Die Variable y_t hat den Erwartungswert
$$\mathbb{E}(y_t) = 0 \cdot P(y_t = 0) + 1 \cdot P(y_t = 1) = p_t\,.$$

14-3

14.2 Lineares Wahrscheinlichkeitsmodell

Bemerkung 14.2

1. Im Standardmodell der linearen Regression
$$y_t = \beta_1 + \beta_2 x_t + u_t$$
wird durch die Regressoren der Erwartungswert der endogenen Variablen erklärt,
$$\mathbb{E}(y_t) = \beta_1 + \beta_2 x_t \,.$$
2. Im **linearen Wahrscheinlichkeitsmodell** für eine dichotome endogene Variable wird ebenfalls der Erwartungswert von y_t, d. h. $p_t = \mathbb{E}(y_t)$, linear erklärt,
$$p_t = \beta_1 + \beta_2 x_t \,.$$

14-4

Bemerkung 14.3 (Probleme bei diesem Ansatz)

1. Es sind Werte $p_t > 1$ und $p_t < 0$ möglich.
2. Damit
$$y_t = p_t + u_t \in \{0, 1\}$$
mit
$$P(y_t = 1) = p_t = 1 - P(y_t = 0)$$
erfüllt sein kann, muss u_t eine binäre Variable sein.
3. Die Wahrscheinlichkeitsverteilung von u_t ist
$$P(u_t = 1 - p_t) = p_t, \quad P(u_t = -p_t) = 1 - p_t \,.$$
4. u_t kann nicht normalverteilt sein.
5. u_t ist heteroskedastisch, da $\mathbb{V}(u_t) = p_t(1 - p_t)$.
6. Die Verteilung von u_t hängt von $\beta_1 + \beta_2 x_t$ ab.

14-5

14.3 Probit-Modell

Bemerkung 14.4

1. Im linearen Regressionsmodell wird durch die Regressoren der Erwartungswert der endogenen Variablen erklärt, z. B.
$$\mathbb{E}(y_t) = \beta_1 + \beta_2 x_t \,.$$
2. Im Probit-Modell wird ebenfalls der Erwartungswert von y_t erklärt,
$$\mathbb{E}(y_t) = \Phi(\beta_1 + \beta_2 x_t) \,.$$
3. Dabei bezeichnet $\Phi : \mathbb{R} \to]0, 1[$ die Verteilungsfunktion der Standardnormalverteilung.

14-6

4. Wegen $\mathbb{E}(y_t) = P(y_t = 1) = p_t$ gilt
$$p_t = \Phi(\beta_1 + \beta_2 x_t)$$
bzw.
$$\Phi^{-1}(p_t) = \beta_1 + \beta_2 x_t\,.$$

Definition 14.5 (Probit)
Für $0 < p < 1$ heißt $\Phi^{-1}(p)$ **Probit** der Wahrscheinlichkeit p.

14-7

14.4 Logit-Modell

Bemerkung 14.6

- Die Erklärung des Erwartungswertes erfolgt durch den Ansatz
$$\mathbb{E}(y_t) = \frac{1}{1 + \exp(-\beta_1 - \beta_2 x_t)}\,.$$
- Alternative Darstellungen sind
$$p_t \stackrel{\text{def}}{=} P(y_t = 1) = \frac{\exp(\beta_1 + \beta_2 x_t)}{1 + \exp(\beta_1 + \beta_2 x_t)}$$
und
$$\ln\left(\frac{p_t}{1 - p_t}\right) = \beta_1 + \beta_2 x_t\,.$$

Definition 14.7 (Logit)
Für $0 < p < 1$ heißt $\ln\left(\frac{p}{1-p}\right)$ **Logit** der Wahrscheinlichkeit p.

14-8

Bemerkung 14.8

- Mit der sogenannten **logistischen Verteilungsfunktion**
$$F : \mathbb{R} \to]0,1[\,, \quad F(x) = \frac{1}{1 + e^{-x}}$$
erhält man
$$\mathbb{E}(y_t) = p_t = F(\beta_1 + \beta_2 x_t)$$
und
$$F^{-1}(p_t) = \beta_1 + \beta_2 x_t\,.$$
- Im Vergleich dazu ist das Probit-Modell
$$\mathbb{E}(y_t) = p_t = \Phi(\beta_1 + \beta_2 x_t)$$
und
$$\Phi^{-1}(p_t) = \beta_1 + \beta_2 x_t\,.$$

14-9

14.5 Schätzung im Logit- und Probit-Modell

Bemerkung 14.9 (ML-Schätzung)

1. Die Parameterschätzung erfolgt meistens mithilfe der Maximum-Likelihood-Methode.
2. Die – in der Regel numerische – Maximierung der Likelihoodfunktion

$$\prod_{t=1}^{T} p_t^{y_t}(1-p_t)^{1-y_t}$$

bzw. der Log-Likelihoodfunktion

$$\sum_{t=1}^{T}(y_t \ln(p_t) + (1-y_t)\ln(1-p_t))$$

bezüglich der beiden Parameter β_1 und β_2 ist erforderlich.

14-10

4. Dabei gilt im Fall des Logit-Modells

$$p_t = F(\beta_1 + \beta_2 x_t)$$

und im Fall des Probit-Modells

$$p_t = \Phi(\beta_1 + \beta_2 x_t)\,.$$

14-11

Bemerkung 14.10 (KQ-Schätzung)

- Bei einer Darstellung mit einer Residualvariablen u_t gilt

$$y_t = p_t + u_t$$

mit

$$\mathbb{E}(u_t) = 0$$

und

$$\mathbb{V}(u_t) = p_t(1-p_t)\,.$$

- Da die Fehlerterme heteroskedastisch sind, ist die KQ-Methodik nur in der Form einer verallgemeinerten gewogenen KQ-Schätzung sinnvoll anwendbar. Z. B. ist im Fall des Logit-Modells die Funktion

$$Q(\beta_1, \beta_2) = \sum_{t=1}^{T} \frac{(y_t - F(\beta_1 + \beta_2 x_t))^2}{F(\beta_1 + \beta_2 x_t)(1 - F(\beta_1 + \beta_2 x_t))}$$

bzgl. β_1 und β_2 zu minimieren.

Kapitel 15

Anhang: Eigenschaften von Matrizen

15-1

15. Anhang: Eigenschaften von Matrizen

15-2

15.1 Grundbegriffe

Definition 15.1 (Skalarprodukt)
Es sei $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

1. Dann heißt
$$\mathbf{x}'\mathbf{y} \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_i y_i \in \mathbb{R}$$
das **Skalarprodukt** (oder das **skalare Produkt**) von $\mathbf{x}$ und $\mathbf{y}$.

2. $\mathbf{x}$ und $\mathbf{y}$ heißen **orthogonal** (zueinander), falls $\mathbf{x}'\mathbf{y} = 0$.

3. Ein Vektor $\mathbf{x} \in \mathbb{R}^n$ heißt **normiert**, falls $\mathbf{x}'\mathbf{x} = 1$.

4. Zwei orthogonale Vektoren $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ heißen **orthonormal**, falls $\mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{y} = 1$.

15-3

Definition 15.2 (Quadratische Matrizen)

1. Eine Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ heißt **quadratisch**, falls $m = n$.
2. Eine quadratische Matrix heißt **symmetrisch**, falls $\mathbf{A}' = \mathbf{A}$.
3. Eine Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ mit den Diagonalelementen $a_{ii} = 1$ für $i = 1, \ldots, n$ und mit $a_{ij} = 0$ für $i \neq j$ heißt **Einheitsmatrix** und wird mit $\mathbf{I}_n$ bezeichnet. Es wird die Kurzschreibweise $\mathbf{I} = \mathbf{I}_n$ verwendet, wenn die Dimension n aus dem Zusammenhang klar ist.
4. Eine quadratische Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ heißt **orthogonal**, falls $\mathbf{A}'\mathbf{A} = \mathbf{I}_n$.

15-4

Bemerkung 15.3

1. Matrizen der Form $\mathbf{X}'\mathbf{X}$ und $\mathbf{X}\mathbf{X}'$ mit $\mathbf{X} \in \mathbb{R}^{m \times n}$ sind symmetrisch.
2. Wenn $\mathbf{A}$ orthogonal ist, sind die Spalten von $\mathbf{A}$ normiert und jeweils zwei verschiedene Spalten orthogonal. Man sagt auch die Spalten bilden ein **normiertes Orthogonalsystem** oder ein **Orthonormalsystem**.
3. Wenn $\mathbf{A}$ orthogonal ist, gilt $\mathbf{A}\mathbf{A}' = \mathbf{I}_n$.

Definition 15.4 (Spur)
Für $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ heißt

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

die **Spur** (*trace*) der Matrix.

15-5

Bemerkung 15.5 (Rechenregeln)

1. Es gilt
$$(\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}' .$$
2. Für zwei quadratische Matrizen $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ gilt
$$\operatorname{tr}(\mathbf{A} + \mathbf{B}) = \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{B}), \quad \operatorname{tr}(\mathbf{A}\mathbf{B}) = \operatorname{tr}(\mathbf{B}\mathbf{A})$$
und
$$\det(\mathbf{A}\mathbf{B}) = \det(\mathbf{A}) \det(\mathbf{B}) .$$
3. Für $\mathbf{A} \in \mathbb{R}^{n \times n}$ gilt
$$\det(\mathbf{A}) = \det(\mathbf{A}') .$$

15-6

15.2 Lineare Unabhängigkeit und Rang

Definition 15.6 (Lineare Unabhängigkeit)
Es sei $k, n \in \mathbb{N}$. Die k Vektoren $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ aus $\mathbb{R}^n$ sind genau dann **linear unabhängig**, wenn

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_k\mathbf{x}_k = \mathbf{0} \implies c_1 = c_2 = \ldots = c_k = 0.$$

Definition 15.7 (Linearkombination)

1. Es sei $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ aus $\mathbb{R}^n$ und $c_1, \ldots, c_k \in \mathbb{R}$, dann heißt

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_k\mathbf{x}_k$$

   Linearkombination der Vektoren $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$.

2. Die Linearkombination heißt **nichttrivial**, falls $c_i \neq 0$ für mindestens ein i.

15-7

Bemerkung 15.8

1. k Vektoren aus $\mathbb{R}^n$ sind genau dann linear unabhängig, wenn der Nullvektor nicht als nichttriviale Linearkombination der k Vektoren dargestellt werden kann.

2. Ein einzelner Vektor ist somit genau dann linear unabhängig, wenn er nicht der Nullvektor ist.

Definition 15.9 (Rang)
Es sei $n, m \in \mathbb{N}$. Der **Rang** (*rank*) einer Matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ ist die maximale Anzahl linear unabhängiger Spalten der Matrix.

15-8

Bemerkung 15.10

1. Für $\mathbf{A} \in \mathbb{R}^{n \times m}$ gilt

$$\mathrm{Rang}(\mathbf{A}) \in \{0, 1, \ldots, \min(n, m)\}$$

   und

$$\mathrm{Rang}(\mathbf{A}) = \mathrm{Rang}(\mathbf{A}').$$

2. Die maximale Anzahl unabhängiger Spalten ist auch die maximale Anzahl unabhängiger Zeilen.

3. Für $\mathbf{x} \in \mathbb{R}^n$ gilt

   - $\mathrm{Rang}(\mathbf{x}) = 0 \quad \iff \quad \mathbf{x} = \mathbf{0}$,
   - $\mathrm{Rang}(\mathbf{x}) = 1 \quad \iff \quad \mathbf{x} \neq \mathbf{0}$.

15-9

4. Es sei $\mathbf{X} = [\mathbf{x}_1|\mathbf{x}_2] \in \mathbb{R}^{n\times 2}$. Dann gilt $\text{Rang}(\mathbf{X}) = 2$ genau dann, wenn
$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = \mathbf{0} \quad \Longrightarrow \quad c_1 = c_2 = 0\,.$$
5. Es sei $\mathbf{X} = [\mathbf{x}_1|\mathbf{x}_2|\cdots|\mathbf{x}_m] \in \mathbb{R}^{n\times m}$ und $\text{Rang}(\mathbf{X}) < m$. Dann existiert ein Koeffizientenvektor $\mathbf{c} \neq \mathbf{0}$, so dass
$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_m\mathbf{x}_m = \mathbf{0}\,.$$
Damit ist mindestens ein Vektor x_i eine Linearkombination der anderen Vektoren x_j, $j \neq i$.

15-10

15.3 Invertierbarkeit einer reellen Matrix

Definition 15.11 (Inverse)
Es sei $n \in \mathbb{N}$. Eine quadratische Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ heißt **invertierbar**, regulär oder nicht singulär (*nonsingular*), falls eine Matrix $\mathbf{A}^{-1} \in \mathbb{R}^{n\times n}$ existiert, so dass
$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n\,.$$
Die Matrix $\mathbf{A}^{-1}$ heißt **inverse Matrix** oder die **Inverse** von $\mathbf{A}$.

Bemerkung 15.12

1. $\mathbf{A} \in \mathbb{R}^{n\times n}$ ist genau dann invertierbar, wenn $\text{Rang}(\mathbf{A}) = n$.
2. $\mathbf{A} \in \mathbb{R}^{n\times n}$ ist genau dann invertierbar, wenn $\det(\mathbf{A}) \neq 0$.
3. Eine Zahl $x \in \mathbb{R}$ ist genau dann invertierbar, wenn $x \neq 0$. Wenn $x \in \mathbb{R}$ invertierbar ist, gilt $x^{-1} = \frac{1}{x}$.
4. Für eine Diagonalmatrix $\mathbf{\Delta} \in \mathbb{R}^{n\times n}$ mit Diagonalelementen $\delta_{ii} > 0$ ist $\mathbf{\Delta}^{-1}$ die Diagonalmatrix mit den Diagonalelementen δ_{ii}^{-1}.

15-11

Bemerkung 15.13

1. Es gilt
$$(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}\,.$$
2. Es gilt
$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}\,.$$
3. Wenn $\mathbf{A}$ orthogonal ist, gilt $\mathbf{A}' = \mathbf{A}^{-1}$.

Bemerkung 15.14 (Wurzel einer Diagonalmatrix)

1. Für eine Diagonalmatrix $\mathbf{\Delta}$ mit den Diagonalelementen $\delta_{ii} \geq 0$ wird $\mathbf{\Delta}^{1/2}$ als die Diagonalmatrix mit den Elementen $\sqrt{\delta_{ii}}$ definiert.
2. Für eine Diagonalmatrix $\mathbf{\Delta}$ mit Diagonalelementen $\delta_{ii} > 0$ wird $\mathbf{\Delta}^{-1/2}$ als die Diagonalmatrix mit den Elementen $(\delta_{ii})^{-1/2}$ definiert.
3. Falls die Diagonalmatrix $\mathbf{\Delta}$ invertierbar ist, ist auch $\mathbf{\Delta}^{1/2}$ invertierbar und es gilt
$$(\mathbf{\Delta}^{1/2})^{-1} = \mathbf{\Delta}^{-1/2}\,.$$

15-12

15.4 Positive Definitheit einer reellen Matrix

Definition 15.15 (Positiv definite Matrix)
Es sei $n \in \mathbb{N}$.

1. Eine symmetrische quadratische Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ heißt **positiv definit**, falls
$$\mathbf{x}'\mathbf{A}\mathbf{x} > 0 \quad \text{für alle } \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}\,. \tag{15.1}$$

2. Eine symmetrische quadratische Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ heißt **positiv semidefinit**, falls
$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \quad \text{für alle } \mathbf{x} \in \mathbb{R}^n\,. \tag{15.2}$$

15-13

Bemerkung 15.16

1. Jede positiv definite Matrix ist invertierbar.

2. Die Einheitsmatrix $\mathbf{I}_n$ ist eine positiv definite Matrix, da
$$\mathbf{x}'\mathbf{I}_n\mathbf{x} = \mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2 > 0 \quad \text{für alle } \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}\,.$$

3. Es sei $\mathbf{X} \in \mathbb{R}^{n \times m}$, dann sind die Matrizen $\mathbf{X}'\mathbf{X}$ und $\mathbf{X}\mathbf{X}'$ positiv semidefinit.

4. Es sei $\mathbf{X} \in \mathbb{R}^{n \times m}$ mit $\text{Rang}(\mathbf{X}) = n$, dann gilt
$$\text{Rang}(\mathbf{X}'\mathbf{X}) = n$$
und $\mathbf{X}'\mathbf{X} \in \mathbb{R}^{n \times n}$ ist positiv definit und damit invertierbar.

15-14

Bemerkung 15.17

1. Die Bezeichnung einer Matrix mit der Eigenschaft (15.2) als **positiv semidefinit** ist in der ökonometrischen und statistischen Literatur verbreitet. Mit dieser Definition ist jede positiv definite Matrix auch positiv semidefinit.

2. Abweichend davon werden Matrizen mit der Eigenschaft (15.2) auch als **nichtnegativ definit** bezeichnet und der Begriff positiv semidefinit wird dann für Matrizen verwendet, die zwar nichtnegativ definit sind, aber nicht positiv definit sind. Bei dieser Terminologie ist nichtnegativ definit der Oberbegriff für die positiv definiten und positiv semidefiniten Matrizen.

15-15

15.5 Eigenwertzerlegung einer positiv definiten Matrix

Satz 15.18 (Eigenwertzerlegung)
$\mathbf{A} \in \mathbb{R}^{n\times n}$ sei eine positiv definite Matrix. Dann gibt es eine Darstellung

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}',$$

wobei $\mathbf{\Lambda} \in \mathbb{R}^{n\times n}$ eine Diagonalmatrix mit den Eigenwerten von $\mathbf{A}$ als den Diagonalelementen und $\mathbf{Q} \in \mathbb{R}^{n\times n}$ eine orthogonale Matrix mit $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_n$ ist.

Bemerkung 15.19

1. $\mathbf{Q}$ ist invertierbar und es gilt $\mathbf{Q}' = \mathbf{Q}^{-1}$ und $\mathbf{Q}\mathbf{Q}' = \mathbf{I}_n$.
2. Es gilt $\mathbf{\Lambda} = \mathbf{Q}'\mathbf{A}\mathbf{Q}$.
3. Es gilt $\mathbf{A}\mathbf{Q} = \mathbf{Q}\mathbf{\Lambda}$, d. h. die n Spalten $\mathbf{q}_i$ von $\mathbf{Q} = [\mathbf{q}_1|\mathbf{q}_2|\dots|\mathbf{q}_n]$ sind ein System orthogonaler Eigenvektoren mit $\mathbf{A}\mathbf{q}_i = \lambda_i\mathbf{q}_i$ für $i = 1,\dots,n$. Das Paar λ_i und $\mathbf{q}_i$ sind ein Eigenwert und ein zugehöriger Eigenvektor der Matrix $\mathbf{A}$.

15-16

Bemerkung 15.20 (Spektraldarstellung)

1. Die Eigenwertzerlegung heißt auch **Spektralzerlegung**.
2. Die Darstellung
$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{q}_i \mathbf{q}_i'$$
heißt auch **Spektraldarstellung** von $\mathbf{A}$.
3. Der Vektor $(\lambda_1, \lambda_2, \dots, \lambda_n)$ (meist aufsteigend oder absteigend geordnet) heißt auch **Spektrum** der Matrix $\mathbf{A}$.

Satz 15.21 (Eigenwerte einer positiv definiten Matrix)
Alle Eigenwerte einer positiv definiten Matrix sind positiv.

Bemerkung 15.22
$\mathbf{A} \in \mathbb{R}^{n\times n}$ sei eine positiv definite Matrix mit der Eigenwertzerlegung $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$. Dann gilt

$$\mathbf{A}^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}'.$$

15-17

Satz 15.23
$\mathbf{\Omega} \in \mathbb{R}^{n\times n}$ sei eine positiv definite Matrix. Dann existiert eine Matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ mit $\mathbf{P}\mathbf{\Omega}\mathbf{P}' = \mathbf{I}_n$ und $\mathbf{\Omega}^{-1} = \mathbf{P}\mathbf{P}'$.

Bemerkung 15.24
Mit der Eigenwertzerlegung $\mathbf{\Omega} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$ gilt $\mathbf{\Omega}^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}' = \mathbf{P}\mathbf{P}'$ mit $\mathbf{P} \stackrel{\text{def}}{=} \mathbf{Q}\mathbf{\Lambda}^{-1/2}$.

Definition 15.25 (Wurzel einer positiv definiten Matrix)
Für eine positiv definite Matrix $\mathbf{\Omega}$ mit der Eigenwertzerlegung $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$ wird $\mathbf{\Omega}^{1/2}$ durch

$$\mathbf{\Omega}^{1/2} \stackrel{\text{def}}{=} \mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}'$$

definiert.

15-18

15.6 Eigenwertzerlegung einer positiv semidefiniten Matrix

Satz 15.26 (Eigenwertzerlegung)
$\mathbf{A} \in \mathbb{R}^{n \times n}$ *sei eine positiv semidefinite Matrix mit* $\mathrm{Rang}(\mathbf{A}) = r$ *und* $1 \leq r \leq n$. *Dann gibt es eine Darstellung*

$$\mathbf{A} = \mathbf{Q}_r \mathbf{\Lambda}_r \mathbf{Q}_r', \tag{15.3}$$

wobei $\mathbf{\Lambda}_r \in \mathbb{R}^{r \times r}$ *eine Diagonalmatrix mit den* r *positiven Eigenwerten von* $\mathbf{A}$ *als Diagonalelementen und* $\mathbf{Q}_r \in \mathbb{R}^{n \times r}$ *eine Matrix mit* $\mathbf{Q}_r'\mathbf{Q}_r = \mathbf{I}_r$ *ist.*

Bemerkung 15.27

1. Die r Spalten $\mathbf{q}_i$ von $\mathbf{Q}_r = [\mathbf{q}_1|\mathbf{q}_2|\ldots|\mathbf{q}_r]$ sind ein System orthogonaler ($\mathbf{q}_i'\mathbf{q}_j = 0$ für $i \neq j$) und normierter ($\mathbf{q}_i'\mathbf{q}_i = 1$) Eigenvektoren mit $\mathbf{A}\mathbf{q}_i = \lambda_i \mathbf{q}_i$ für $i = 1, \ldots, r$, wobei die λ_i die positiven Eigenwerte von $\mathbf{A}$ sind.
2. Falls eine Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ den Rang Null hat, ist $\mathbf{A}$ die Nullmatrix und es gilt $\lambda_1 = \ldots = \lambda_n = 0$.

15-19

Definition 15.28 (Hauptkomponentenzerlegung)
$\mathbf{A} \in \mathbb{R}^{n \times n}$ sei eine positiv semidefinite Matrix mit $\mathrm{Rang}(\mathbf{A}) = r$ und $1 \leq r \leq n$.

1. Die Spektraldarstellung
$$\mathbf{A} = \sum_{i=1}^{r} \lambda_i \mathbf{q}_i \mathbf{q}_i'$$
mit absteigend geordneten positiven Eigenwerten
$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$$
heißt **Hauptkomponentenzerlegung** der Matrix $\mathbf{A}$.
2. Für $1 \leq k \leq r$ ist
$$\mathbf{A}_k \stackrel{\text{def}}{=} \sum_{i=1}^{k} \lambda_i \mathbf{q}_i \mathbf{q}_i'$$
die auf den ersten k Hauptkomponenten basierende Approximation von $\mathbf{A}$.

15-20

Bemerkung 15.29

1. Es gilt $\mathrm{Rang}(\mathbf{A}_k) = k$ für $k = 1, \ldots, r$.
2. Die eindeutige verallgemeinerte Inverse (siehe Definition 15.33) von $\mathbf{A}_k$ ist
$$\mathbf{A}_k^+ = \sum_{i=1}^{k} \lambda_i^{-1} \mathbf{q}_i \mathbf{q}_i'.$$
3. Die eindeutige verallgemeinerte Inverse von $\mathbf{A}$ ist
$$\mathbf{A}^+ = \mathbf{A}_r^+ = \mathbf{Q}_r \mathbf{\Lambda}_r^{-1} \mathbf{Q}_r'.$$

15-21

15.7 Verallgemeinerte Inversen

Definition 15.30 (Verallgemeinerte Inverse)
Für $\mathbf{A} \in \mathbb{R}^{m\times n}$ heißt eine Matrix $\mathbf{A}^{-} \in \mathbb{R}^{n\times m}$ **verallgemeinerte Inverse** (*generalized inverse*) oder g-Inverse (*g-inverse*) der Matrix $\mathbf{A}$, falls

$$\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}.$$

Bemerkung 15.31

1. Anstelle von $\mathbf{A}^{-}$ sind auch die Bezeichnungen $\mathbf{A}_g$ oder $\mathbf{A}^g$ üblich.
2. Zu jeder Matrix $\mathbf{A}$ existiert mindestens eine verallgemeinerte Inverse.
3. $\mathbf{A}^{-}$ ist genau dann eindeutig, wenn $m = n$ gilt und $\mathbf{A}$ invertierbar ist. In diesem Fall gilt $\mathbf{A}^{-} = \mathbf{A}^{-1}$.

15-22

Satz 15.32 *Wenn das lineare Gleichungssystem*

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

für $\mathbf{x}$ *konsistent ist, d. h. mindestens eine Lösung besitzt, und* $\mathbf{A}^{-}$ *eine verallgemeinerte Inverse von* $\mathbf{A}$ *ist, dann ist* $\mathbf{x}^{\star} \overset{\text{def}}{=} \mathbf{A}^{-}\mathbf{b}$ *eine Lösung des Gleichungssystems.*

Beweis:
$\mathbf{x}$ sei eine beliebige Lösung, und $\mathbf{A}^{-}$ sei eine verallgemeinerte Inverse von $\mathbf{A}$. Durch Linksmultiplikation von $\mathbf{A}\mathbf{x} = \mathbf{b}$ mit $\mathbf{A}\mathbf{A}^{-}$ erhält man

$$\mathbf{A}\mathbf{A}^{-}\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^{-}\mathbf{b}\,.$$

Für die linke Seite gilt $\mathbf{A}\mathbf{A}^{-}\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x} = \mathbf{b}$, für die rechte Seite gilt $\mathbf{A}\mathbf{A}^{-}\mathbf{b} = \mathbf{A}\mathbf{x}^{\star}$, also gilt

$$\mathbf{b} = \mathbf{A}\mathbf{x}^{\star},$$

so dass $\mathbf{x}^{\star} = \mathbf{A}^{-}\mathbf{b}$ eine Lösung des Gleichungssystems ist.

15-23

Definition 15.33 (Eindeutige verallgemeinerte Inverse)
Für $\mathbf{A} \in \mathbb{R}^{m\times n}$ heißt eine Matrix $\mathbf{A}^{+} \in \mathbb{R}^{n\times m}$ **eindeutige verallgemeinerte Inverse** (*unique generalized inverse*) oder **Moore-Penrose-Inverse** der Matrix $\mathbf{A}$, falls

1. $\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}$,
2. $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+}$,
3. $\mathbf{A}\mathbf{A}^{+}$ ist symmetrisch,
4. $\mathbf{A}^{+}\mathbf{A}$ ist symmetrisch.

15-24

Satz 15.34

1. $\mathbf{A}^+$ *ist eindeutig.*
2. *Wenn* $\mathbf{A}$ *invertierbar ist, gilt* $\mathbf{A}^+ = \mathbf{A}^{-1}$.
3. *Wenn* $\mathbf{A}'\mathbf{A}$ *invertierbar ist, gilt* $\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$.
4. $(\mathbf{A}^+)^+ = \mathbf{A}$.
5. $(\mathbf{A}')^+ = (\mathbf{A}^+)'$.
6. $(\mathbf{A}'\mathbf{A})^+ = \mathbf{A}^+(\mathbf{A}^+)'$.
7. $\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^+\mathbf{A}'$.

15.8 Ergänzungen

Bemerkung 15.a

1. Wenn das Gleichungssystem $\mathbf{Ax} = \mathbf{b}$ konsistent ist, dann ist $\mathbf{x}^{(-)} = \mathbf{A}^-\mathbf{b}$ für jede verallgemeinerte Inverse $\mathbf{A}^-$ von $\mathbf{A}$ eine Lösung und unter diesen Lösungen ist $\mathbf{x}^{(+)} = \mathbf{A}^+\mathbf{b}$ diejenige Lösung mit der kürzesten Länge $(\mathbf{x}'\mathbf{x})^{1/2}$.
2. Wenn das Gleichungssystem $\mathbf{Ax} = \mathbf{b}$ inkonsistent ist, kann eine Näherungslösung im Sinne der Methode der kleinsten Quadrate gesucht werden, d. h. ein Vektor aus der nichtleeren Menge
$$L = \{\mathbf{x} \mid \mathbf{x} \text{ minimiert } (\mathbf{Ax} - \mathbf{b})'(\mathbf{Ax} - \mathbf{b})\}\ .$$
Es gilt
$$L = \left\{\mathbf{x} \mid \mathbf{A}'\mathbf{Ax} = \mathbf{A}'\mathbf{b}\right\},$$
d. h. L besteht aus allen Lösungen der Normalgleichungen. Der Vektor $\mathbf{x}^{(+)} = \mathbf{A}^+\mathbf{b}$ mit $\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^+\mathbf{A}'$ ist derjenige Vektor in L mit der kürzesten Länge $(\mathbf{x}'\mathbf{x})^{1/2}$.

Bemerkung 15.b

1. Manchmal wird in der Literatur für die eindeutige verallgemeinerte Inverse (Moore-Penrose-Inverse) auch das Symbol $\mathbf{A}^-$ verwendet.
2. Eine verallgemeinerte Inverse mit den ersten beiden Eigenschaften aus Definition 15.33 heißt **reflexive verallgemeinerte Inverse**.
3. Der Begriff **Pseudoinverse** wird in der Literatur uneinheitlich teils für eine verallgemeinerte Inverse, teils für die Moore-Penrose-Inverse und teils für eine reflexive verallgemeinerte Inverse verwendet.

Bemerkung 15.c
In dieser Bemerkung wird die eindeutige verallgemeinerte Inverse sukzessive erst für reelle Zahlen, dann für Diagonalmatrizen, dann für positiv semidefinite Matrizen und schließlich für beliebige rechteckige Matrizen definiert.

1. Für $x \in \mathbb{R}$ ist
$$x^+ \stackrel{\text{def}}{=} \begin{cases} x^{-1} & \text{für } x \neq 0 \\ 0 & \text{für } x = 0 \end{cases}$$
die eindeutige verallgemeinerte Inverse.

2. Für eine Diagonalmatrix $\mathbf{D} \in \mathbb{R}^{n\times n}$ mit den Diagonalelementen $d_1, \ldots, d_n$ ist die Diagonalmatrix $\mathbf{D}^+ \in \mathbb{R}^{n\times n}$ mit den Diagonalelementen $d_1^+, \ldots, d_n^+$ die eindeutige verallgemeinerte Inverse.

3. Für eine positiv semidefinite Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ existiert eine Darstellung

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}',$$

wobei $\mathbf{\Lambda} \in \mathbb{R}^{n\times n}$ eine Diagonalmatrix mit den nichtnegativen Eigenwerten von $\mathbf{A}$ als den Diagonalelementen und $\mathbf{Q} \in \mathbb{R}^{n\times n}$ eine orthogonale Matrix mit $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_n$ ist. Die eindeutige verallgemeinerte Inverse von $\mathbf{A}$ ist durch

$$\mathbf{A}^+ \stackrel{\text{def}}{=} \mathbf{Q}\mathbf{\Lambda}^+\mathbf{Q}'$$

gegeben.

4. Für eine Matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ ist die Matrix $\mathbf{A}'\mathbf{A} \in \mathbb{R}^{n\times n}$ positiv semidefinit und durch

$$\mathbf{A}^+ \stackrel{\text{def}}{=} (\mathbf{A}'\mathbf{A})^+\mathbf{A}'$$

ist die eindeutige verallgemeinerte Inverse von $\mathbf{A}$ gegeben.

Kapitel 16

Anhang: Statistische Grundlagen

16-1

16. Anhang: Statistische Grundlagen

16-2

16.1 Zufallsvariablen und Zufallsvektoren

Bemerkung 16.1

1. **Erwartungswert** einer Zufallsvariablen Z: $\mathbb{E}(Z)$
2. Der **Erwartungswert eines Zufallsvektors** $\mathbf{Z} = (z_1, z_2, \ldots, z_n)'$ mit Werten in $\mathbb{R}^n$ wird komponentenweise gebildet,
$$\mathbb{E}(\mathbf{Z}) \stackrel{\text{def}}{=} (\mathbb{E}(Z_1), \mathbb{E}(Z_2), \ldots, \mathbb{E}(Z_n))' \in \mathbb{R}^n .$$
3. Der **Erwartungswert einer Matrix von Zufallsvariablen** $\mathbf{Z} = [Z_{ij}]_{i=1,\ldots,n;j=1,\ldots,m}$ mit Werten in $\mathbb{R}^{n\times m}$ wird komponentenweise gebildet,
$$\mathbb{E}(\mathbf{Z}) \stackrel{\text{def}}{=} [\mathbb{E}(Z_{ij})]_{i=1,\ldots,n;j=1,\ldots,m}.$$

16-3

Bemerkung 16.2

1. **Varianz** einer Zufallsvariablen Z:

$$\mathbb{V}(Z) \stackrel{\text{def}}{=} \mathbb{E}[(Z - \mathbb{E}(Z))^2] = \mathbb{E}(Z^2) - (\mathbb{E}(Z))^2$$

2. **Kovarianz** von zwei Zufallsvariablen Z_1 und Z_2:

$$\mathbb{C}\text{ov}(Z_1, Z_2) \stackrel{\text{def}}{=} \mathbb{E}[(Z_1 - \mathbb{E}(Z_1))(Z_2 - \mathbb{E}(Z_2))] = \mathbb{E}(Z_1 Z_2) - \mathbb{E}(Z_1)\mathbb{E}(Z_2)$$

3. Die **Kovarianzmatrix** eines Zufallsvektors $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)'$ ist

$$\mathbb{V}(\mathbf{Z}) \stackrel{\text{def}}{=} [\mathbb{C}\text{ov}(Z_i, Z_j)]_{i,j=1,\ldots,n} \in \mathbb{R}^{n\times n}\,.$$

4. Es gilt

$$\mathbb{V}(\mathbf{Z}) = \mathbb{E}(\mathbf{Z}\mathbf{Z}') - \mathbb{E}(\mathbf{Z})\mathbb{E}(\mathbf{Z}').$$

16-4

Bemerkung 16.3 (Rechenregeln)
$\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)'$ sei ein n-dimensionaler Zufallsvektor mit $\mathbb{E}(\mathbf{Z}) \in \mathbb{R}^n$ und $\mathbb{V}(\mathbf{Z}) \in \mathbb{R}^{n\times n}$. Es sei $\mathbf{a} \in \mathbb{R}^m$ und $\mathbf{A} \in \mathbb{R}^{m\times n}$.

1. $\mathbb{E}(\mathbf{a} + \mathbf{A}\mathbf{Z}) = \mathbf{a} + \mathbf{A}\mathbb{E}(\mathbf{Z})$
2. $\mathbb{V}(\mathbf{a} + \mathbf{A}\mathbf{Z}) = \mathbf{A}\mathbb{V}(\mathbf{Z})\mathbf{A}'$

16-5

16.2 Wahrscheinlichkeitsverteilungen

16.2.1 Multivariate Normalverteilung

Definition 16.4 (Multivariate Normalverteilung)

1. Ein Zufallsvektor $\mathbf{Z} = (Z_1, \ldots, Z_n)'$ ist **multivariat normalverteilt** mit den Parametern $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$ und $\boldsymbol{\Sigma} = [\sigma_{ij}] \in \mathbb{R}^{n\times n}$, wobei $\boldsymbol{\Sigma}$ positiv definit ist, falls $\mathbf{Z}$ die Dichtefunktion

$$f_{\mathbf{Z}}(z_1, \ldots, z_n) = (2\pi \det(\boldsymbol{\Sigma}))^{-n/2} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}, \quad \mathbf{z} \in \mathbb{R}^n$$

besitzt.

2. Notation: $\mathbf{Z} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
3. Falls $\mu = \mathbf{0}$ und $\sigma_{ii} = 1$ für $i = 1, \ldots, n$ heißt **$\mathbf{Z}$ multivariat standardnormalverteilt**.

16-6

Bemerkung 16.5
Für eine multivariate Standardnormalverteilung ist $\mathbf{\Sigma}$ eine Korrelationsmatrix.

Bemerkung 16.6
Für $\mathbf{Z} \sim N_n(\boldsymbol{\mu}, \mathbf{\Sigma})$ gilt:

1. $\mathbb{E}(\mathbf{Z}) = \boldsymbol{\mu}$
2. $\mathbb{V}(\mathbf{Z}) = \mathbf{\Sigma}$
3. $Z_i \sim N(\mu_i, \sigma_{ii})$ für $i = 1, \ldots, n$
4. Falls $\mathbf{\Sigma}$ eine Diagonalmatrix ist, z. B. $\mathbf{\Sigma} = \mathbf{I}_n$, dann ist $\mathbf{Z}$ ist insgesamt stochastisch unabhängig.
5. $\mathbf{\Sigma}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu}) \sim N_n(\mathbf{0}, \mathbf{I}_n)$
6. $(\mathbf{Z} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{Z} - \boldsymbol{\mu}) \sim \chi^2_n$

16-7

Bemerkung 16.7
Es sei $\mathbf{Z} \sim N_n(\boldsymbol{\mu}, \mathbf{\Sigma})$.

1. Für $\mathbf{a} \in \mathbb{R}^n$ gilt $\mathbf{a} + \mathbf{Z} \sim N_n(\mathbf{a} + \boldsymbol{\mu}, \mathbf{\Sigma})$.
2. Es sei $\mathbf{A} \in \mathbb{R}^{m \times n}$ und $\mathbf{A\Sigma A}' \in \mathbb{R}^{m \times m}$ invertierbar, dann gilt $\mathbf{AZ} \sim N_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A}')$.
3. Aus $\mathbf{c} \in \mathbb{R}^n, \mathbf{c} \neq \mathbf{0}$ folgt $\mathbf{c}'\mathbf{Z} \sim N(\mathbf{c}'\boldsymbol{\mu}, \mathbf{c}'\mathbf{\Sigma c})$.

16.2.2 Chiquadratverteilung

16-8

Definition 16.8 (Chiquadratverteilung)
Es sei $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{I})$, dann heißt die Verteilung von $\mathbf{Z}'\mathbf{Z}$ **Chiquadratverteilung** mit Parameter $n \in \mathbb{N}$. Der Parameter n heißt auch **Anzahl der Freiheitsgrade**.
Notation: $X \sim \chi^2_n$, falls X eine Chiquadratverteilung mit Parameter n besitzt.

Bemerkung 16.9
Für $X \sim \chi^2_n$ gilt $\mathbb{E}(X) = n$ und $\mathbb{V}(X) = 2n$.

Dresdner Beiträge zu Quantitativen Verfahren (ISSN 0945-4802)

Ältere Ausgaben (1/94 – 48/08): http://wwqvs.file3.wcms.tu-dresden.de/f-db.htm

49/09 E. Lovász, B. Schipp: The Impact of HIV/AIDS on Economic Growth in Sub-Saharan Africa.
Erschienen in: *South African Journal of Economics*, Vol. 77, Nr. 2, 2009, S. 245-256.

50/09 S. Höse, S. Huschens: Confidence Intervals for Correlations in the Asymptotic Single Risk Factor Model.

51/10 S. Höse, S. Huschens: Confidence Intervals for Quantiles of a Vasicek-distributed Credit Portfolio Loss.

52/10 D. Tillich: Risikomaßzahlen für Kreditportfoliotranchen.
Erschienen in: *AStA Wirtschafts- und Sozialstatistisches Archiv,* Vol. 5, Nr. 1, 2011, S. 59-76. DOI: 10.1007/s11943-011-0095-1

53/10 S. Huschens: Kann es Rückzahlungswahrscheinlichkeiten von 100% geben?
Erschienen in: *bank und markt – Zeitschrift für Retailbanking*, 39. Jg., Heft 3/2010, S. 11.

54/11 S. Höse, S. Huschens: Confidence Intervals for Asset Correlations in the Asymptotic Single Risk Factor Model.
Erschienen in: *Operations Research Proceedings 2010,* Hrsg: B. Hu, K. Morasch, S. Pickl, M. Siegle, Springer, Berlin, 2011, S. 111-116.

55/11 S. Höse, S. Huschens: Stochastic Orders and Non-Gaussian Risk Factor Models.
Erschienen in: *Review of Managerial Science*. Vol. 7, Nr. 2, 2013, S. 99-140. DOI: 10.1007/s11846-011-0071-8

56/11 D. Tillich: Bounds for the Expectation of Bounded Random Variables.
DOI: 10.13140/RG.2.1.2656.7768

57/12 S. Höse, S. Huschens: Credit Portfolio Correlations and Uncertainty.
Erschienen in: *Credit Securitisations and Derivatives – Challenges for the Global Markets,* Hrsg.: D. Rösch, H. Scheule, Wiley: Chichester, 2013, S. 53-70

58/12 S. Fischer: Ratio calculandi periculi - Ein analytischer Ansatz zur Bestimmung der Verlustverteilung eines Kreditportfolios

59/13 D. Tillich, D. Ferger: Estimation of Rating Classes and Default Probabilities in Credit Risk Models with Dependencies.
Erschienen in: *Applied Stochastic Models in Business and Industry,* Vol. 31, Nr. 6, 2015, S. 762-781. DOI: 10.1002/asmb.2089

60/14 C. Lehmann, D. Tillich: Consensus Information and Consensus Rating – A Note on Methodological Problems of Rating Aggregation.
Erschienen in: *Operations Research Proceedings 2014,* Hrsg.: M. Lübbecke et al., Springer, 2016, S. 357-362. DOI: 10.1007/978-3-319-28697-6_50

61/15 C. Lehmann, D. Tillich: Applied Consensus Information and Consensus Rating – A Simulation Study on Rating Aggregation.
Erschienen in: *Journal of Risk Model Validation* Jg. 10, Heft 4, 2016, S. 1-21.

62/16 C. Lehmann: Modellierung der Abhängigkeitsstruktur von Ausfallkörben – Eine Betrachtung für den Spezialfall des Duo-Baskets.

63/16 S. Huschens: Chance (*odd*) versus Wahrscheinlichkeit (*probability*).

64/16 S. Huschens: Stetigkeit in der Statistik.

65/16 S. Huschens: Literaturauswahl zur Statistik.

66/16 D. Tillich, C. Lehmann: Estimation in discontinuous Bernoulli mixture models applicable in credit rating systems with dependent data.

67/16 D. Tillich: Generalized Modeling and Estimation of Rating Classes and Default Probabilities Considering Dependencies in Cross and Longitudinal Section.

68/17 S. Huschens: Risikomaße.

69/17 S. Huschens: Einführung in die Ökonometrie.