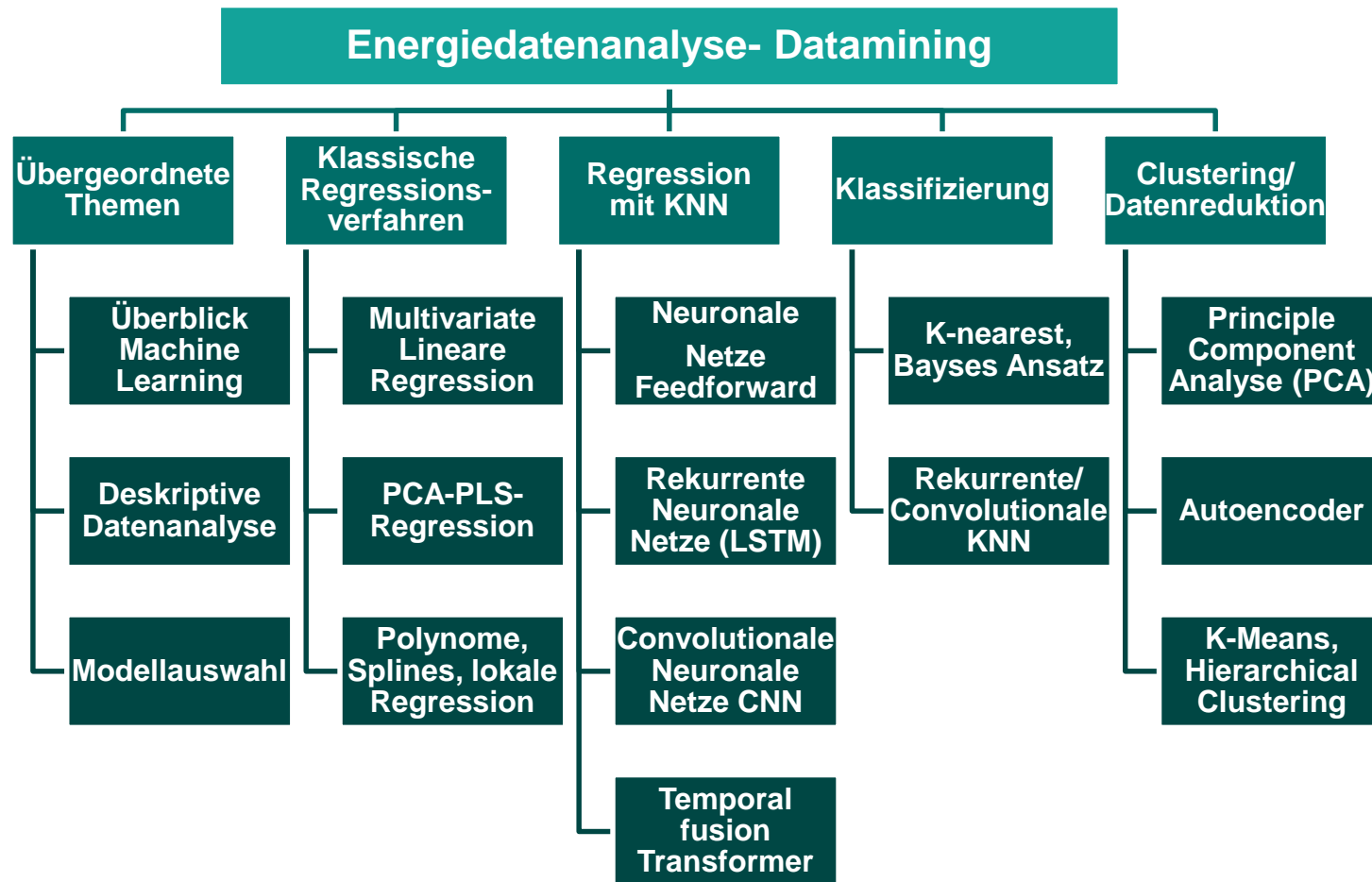


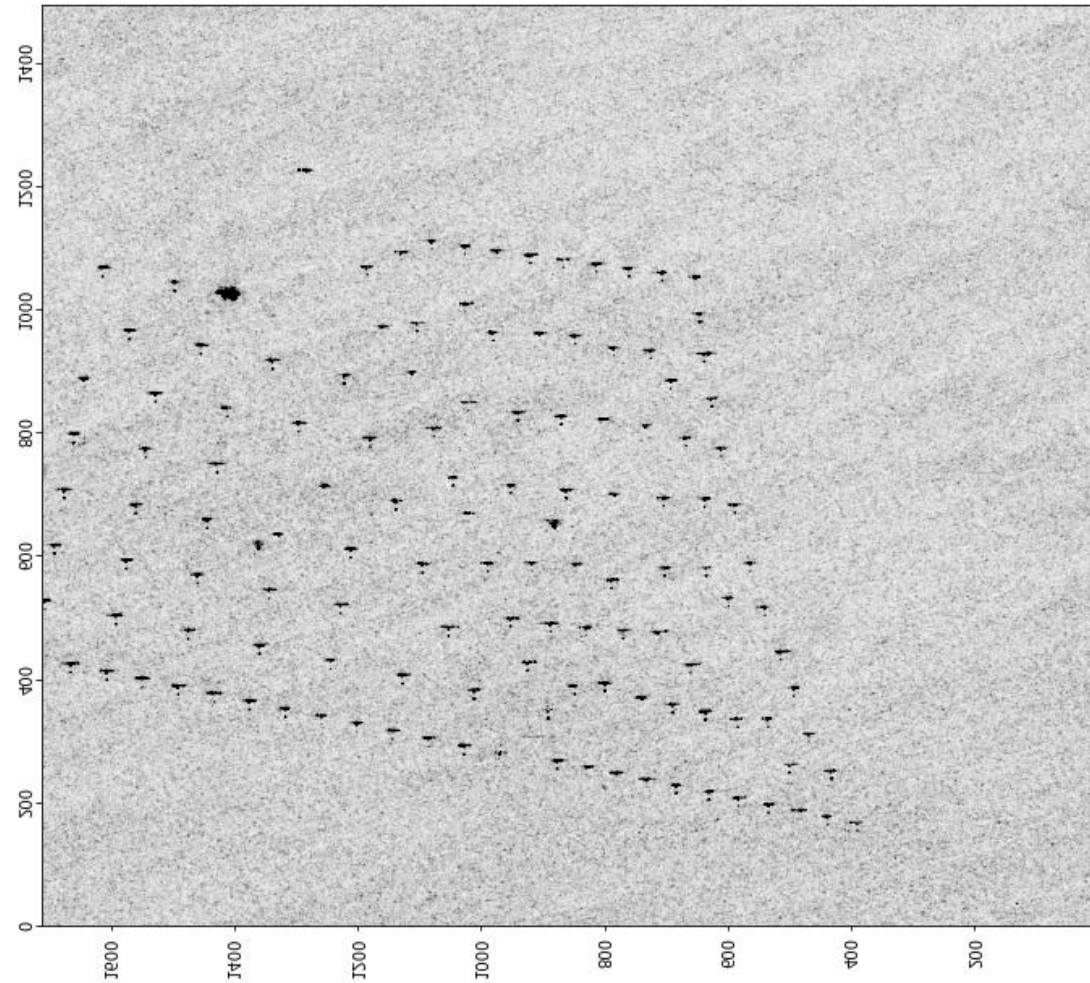
Lineare Regression

Energiedatenanalyse - Datamining

Die Themengebiete der Veranstaltung verknüpfen Modelle des „Machine Learning“ mit energiewirtschaftlichen Fragestellungen



Der Cliffhänger: Was ist das und was hat das mit Regression zu tun?



1. Lineare Regression (Theorie)



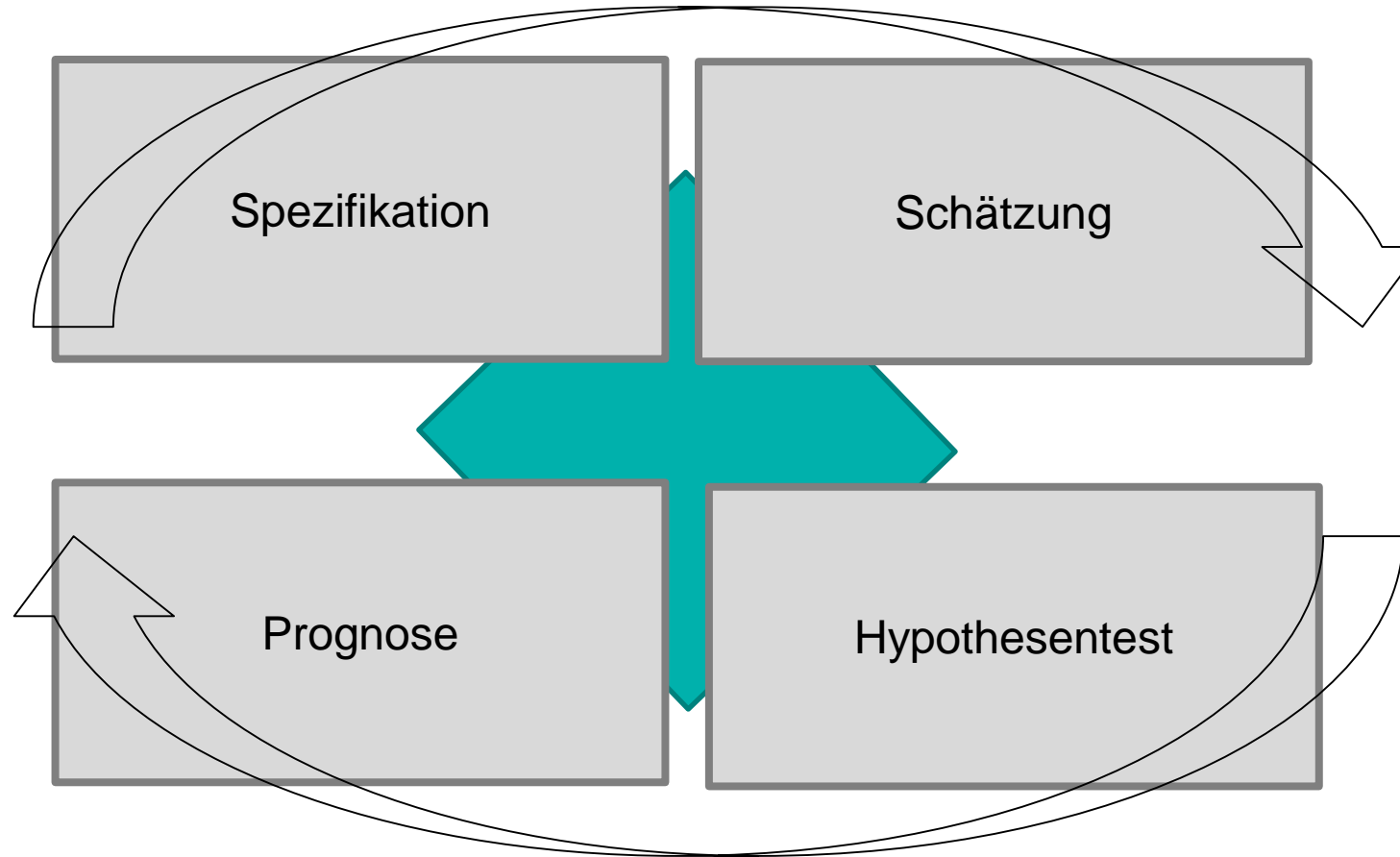
2. Case Study 1: Datenlücken bereinigen

3. Case Study 2: Winkelerkennung von WKAs

4. Case Study 3: Multiple Regression Strompreis auf Brennstoffe und CO2



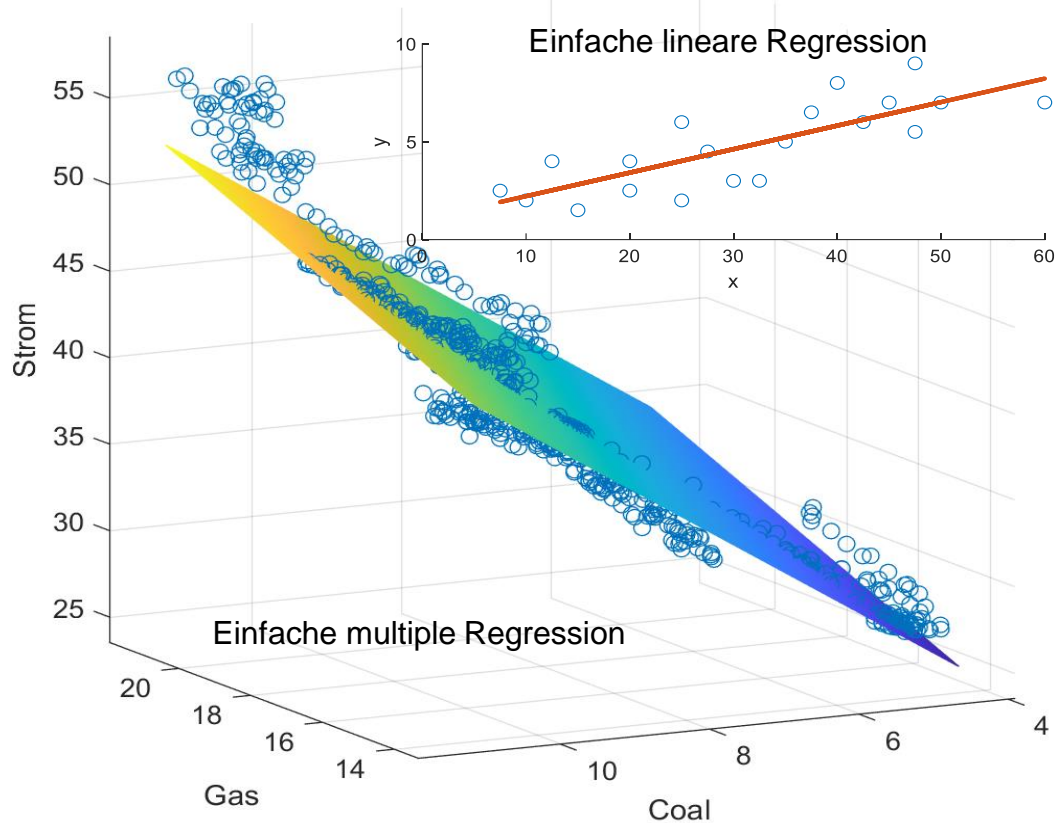
Motivation: Überblick Vorgehensweise lineare Regression



Quelle: In Anlehnung an: Eine Einführung; 6. Auflage; Ludwig von Auer; 2013; Springer Gabler Verlag

Bei Verwendung mehrerer erklärender Variablen ($k \geq 2$) spricht man vom „multiplen“, sonst vom „einfachen“ Regressionsmodell

Einfache und multiple lineare Regression



Bezeichnungen

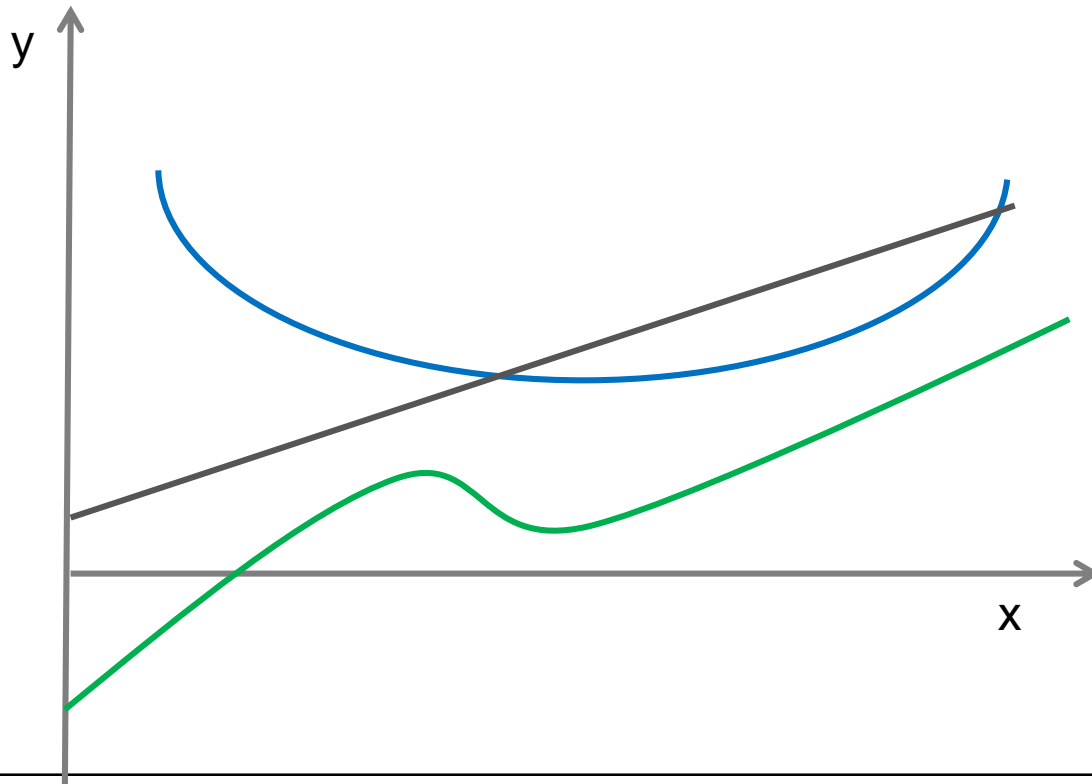
„Sprache“	X	Y
Allgemein	unabhängige Variable	abhängige Variable
	erklärende Variable	erklärte Variable
Modellbildung	exogene Variable	endogene Variable
Maschine Learning	Feature	Output/ Label
Regression	Regressor	Regressand

- A1: Die Regressoren beschreiben vollständig das Verhalten von Y.
- A2: Die Parameter a, b,... sind Konstanten.

Es kommen viele parametrische Funktionen für eine Anpassung einer linearen Regression in Frage

Abbildung des Zusammenhangs

Parametrische Funktionen



Spezifikation und Parameterschätzung

■ Vorgehensweise der Regressionsrechnung

1. Wahl eines geeigneten Funktions-Typs

$y=f(x)$, z.B.:

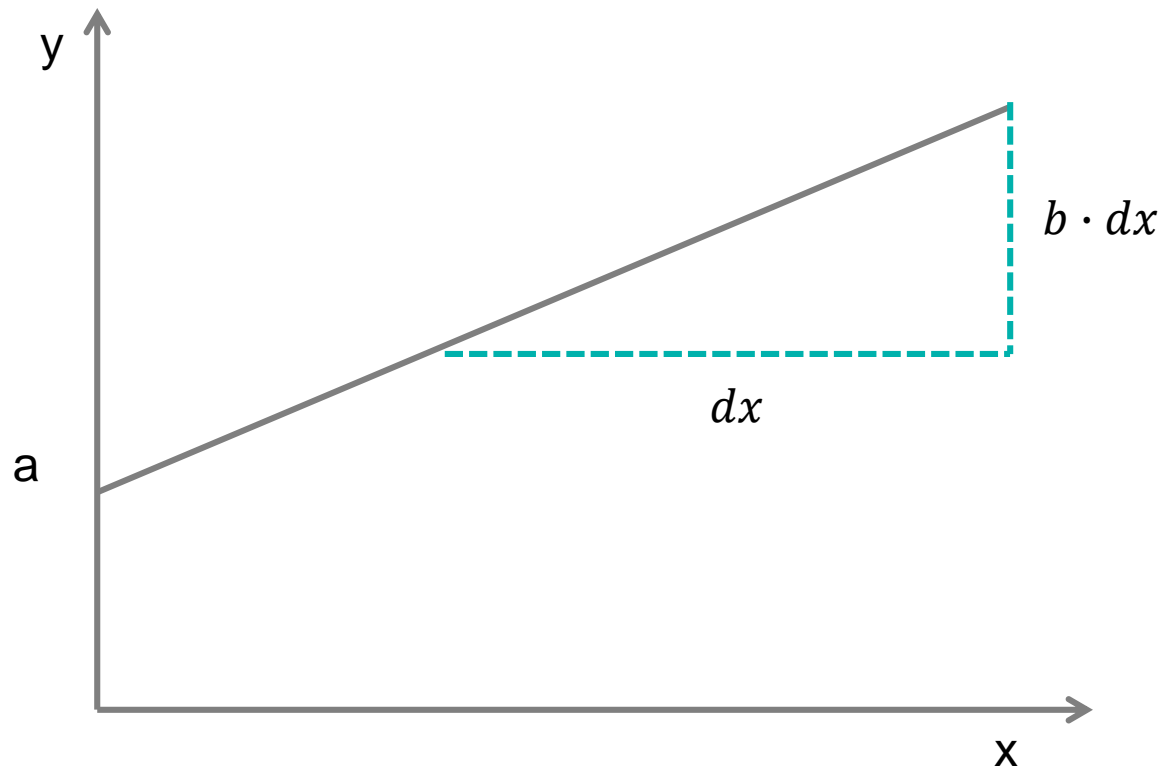
Funktionstyp	Formel
linear	$y = a + bx$
quadratisch	$y = a + bx + cx^2$
exponentiell	$y = e^{a+bx+cx^2}$
trigonometrisch	$y = a + b\sin(x) + b\cos(x)$

2. Anpassung der Funktion $y=f(x)$ an die Punktwolke

■ Berechnung geeigneter Werte für die Parameter

Schritt 1: Wahl eines geeigneten Modells

Abbildung des Zusammenhangs zwischen exogener und endogener Größe



1. Wahl eines geeigneten Funktions-Typ

$$y = f(x)$$

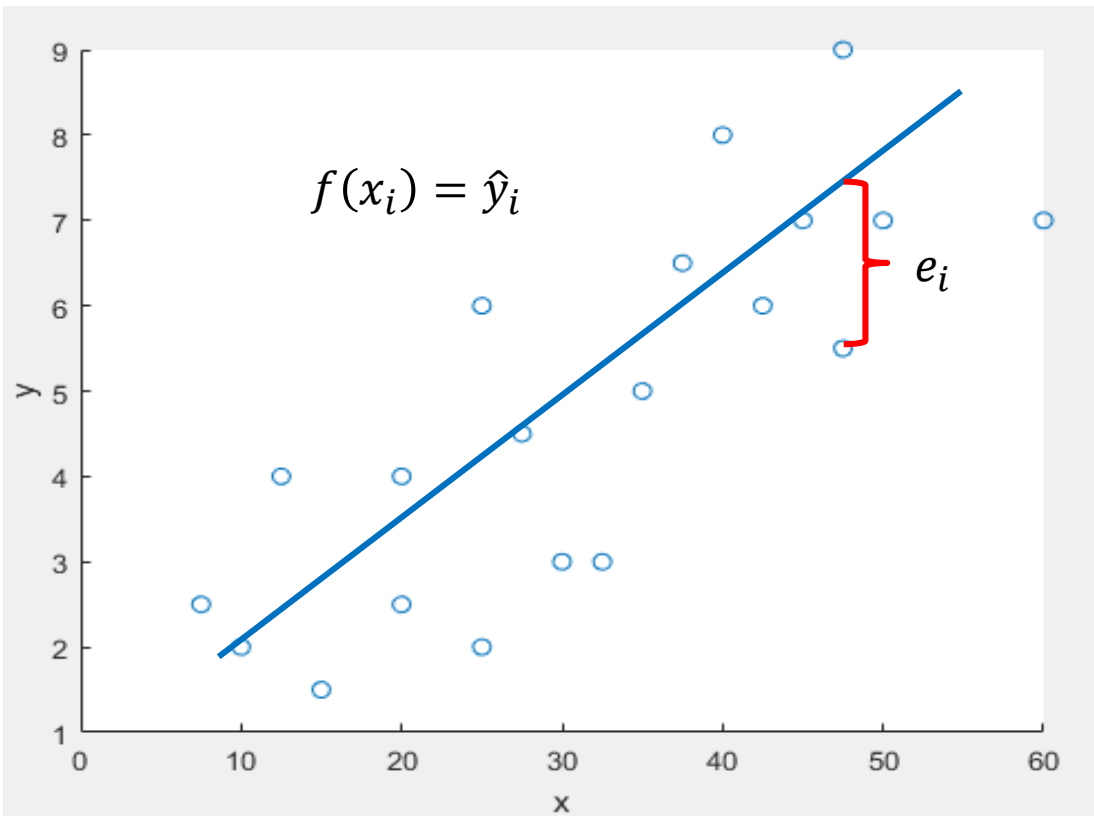
z. B. Geradenfunktion

$$f(x) = a + bx$$

- Beschreibung durch 2 Parameter a und b
- a und b stellen die Regressionsparameter dar

I.d.R. kann das Modell $f(x)=a+bx$ den empirischen Zusammenhang nicht exakt abbilden.

Abbildung der Residuen e



Erweiterung des Modells um eine Störgröße:

- Annahme: Überlagerung nicht-systematischer zufälliger Störeinflüsse ϵ_i
- Entsprechend ist die Gleichung um einen Störterm zu erweitern, welcher eine zufällige Ausprägung in jedem Zeitschritt t aufweist:

$$Y_i = a + bX_i + \epsilon_i$$

Abweichung des gemessenen Wert zur berechneten Geraden

$$e_i = y_i - f(x_i) = y_i - \hat{y}_i$$

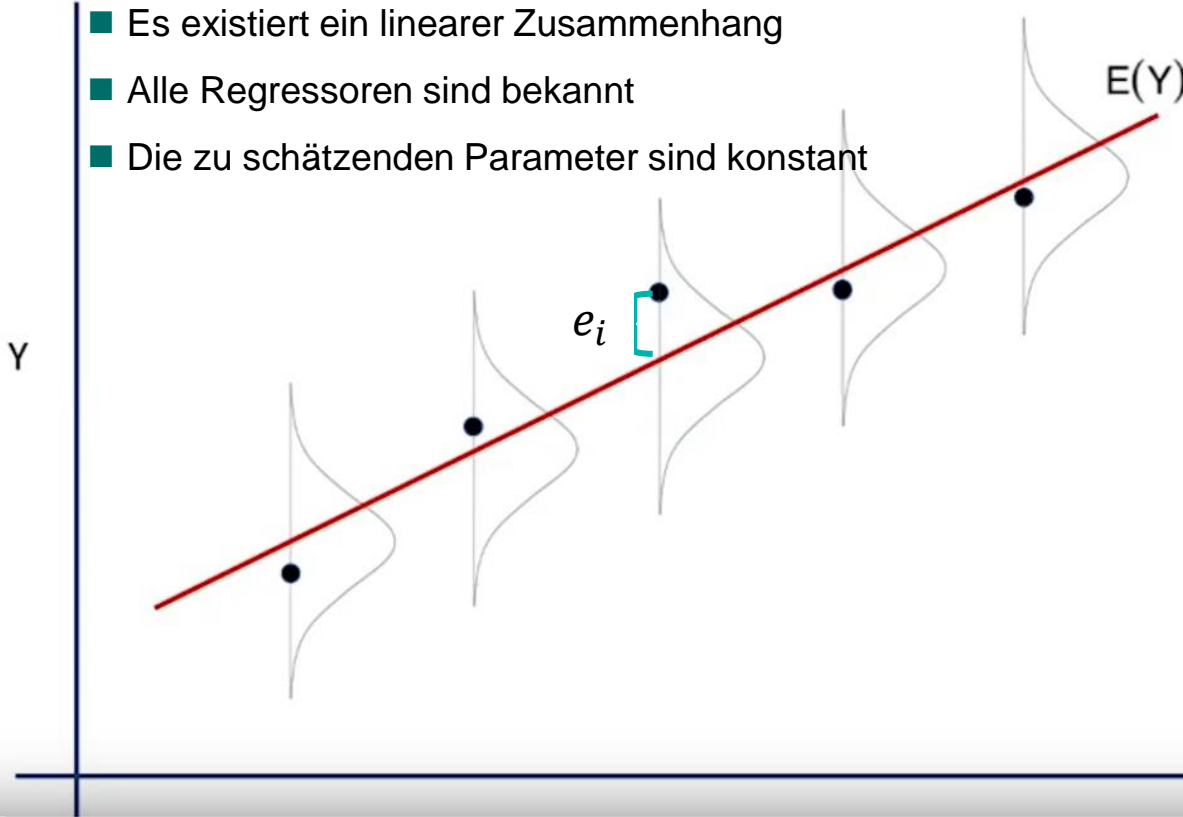
$$= (\text{gemessener } y \text{ Wert}) - (\text{berechneter } \hat{y} \text{ Wert})$$

E misst die Abweichung in der Stichprobe.
 ϵ stellt den grundsätzlichen Modellzusammenhang dar.

Es gibt eine Vielzahl von Annahmen zu den Eigenschaften des resultierenden Modells und der Störgröße

Annahmen bzgl. des Modells

- Es existiert ein linearer Zusammenhang
- Alle Regressoren sind bekannt
- Die zu schätzenden Parameter sind konstant



- Der Erwartungswert beträgt Null:

$$E(\epsilon)=0$$

- Die Störgröße hat eine konstante Varianz (*Homoskedastizität*):

$$\text{Var}(\epsilon)=\sigma^2$$

- Die Störgrößen sind nicht korreliert (fehlende *Autokorrelation*):

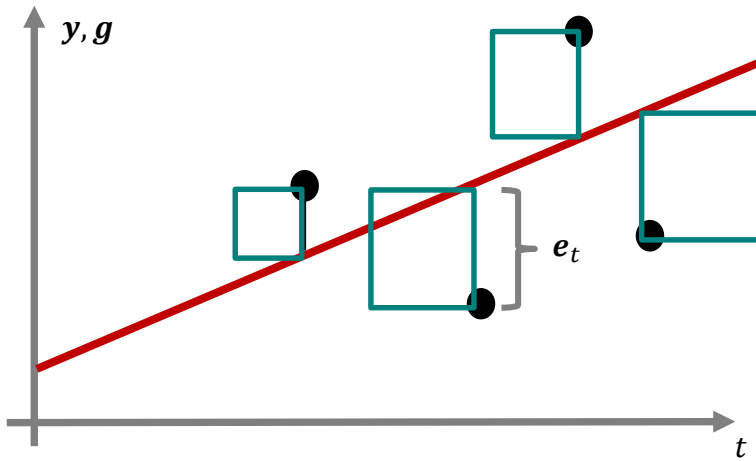
$$\text{Cov}(\epsilon_1, \epsilon_2)=0$$

- Die Störgröße ist unabhängig identisch normalverteilt:

$$\epsilon \sim N(0, \sigma^2)$$

Die Trendbereinigung einer Zeitreihe kann klassisch mit Hilfe von linearen und/oder polynomialen Funktionen vorgenommen werden

KQ-Schätzung



Das Bestimmtheitsmaß R^2 kann zur Beurteilung herangezogen werden:

$$R^2 = 1 - \frac{\sum_{t=1}^n \hat{e}_t^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

KQ: kleinste Quadrate Schätzer; \bar{y} =Mittelwert von y

Modelle zur Trendbereinigung:

- Lineare Regression

$$y_t = a_0 + a_1 \cdot t + e_t$$

- Polynomische Trends

$$y_t = a_0 + \sum_{i=1}^n a_i \cdot t^i + e_t$$

- Schätzung erfolgt mit Methode der kleinsten Quadrate (KQ-Methode)

$$\min_{a_0, \dots, a_1} \sum_{t=1}^n (y_t - a_0 - a_1 \cdot t)^2$$

Ergebnisse:

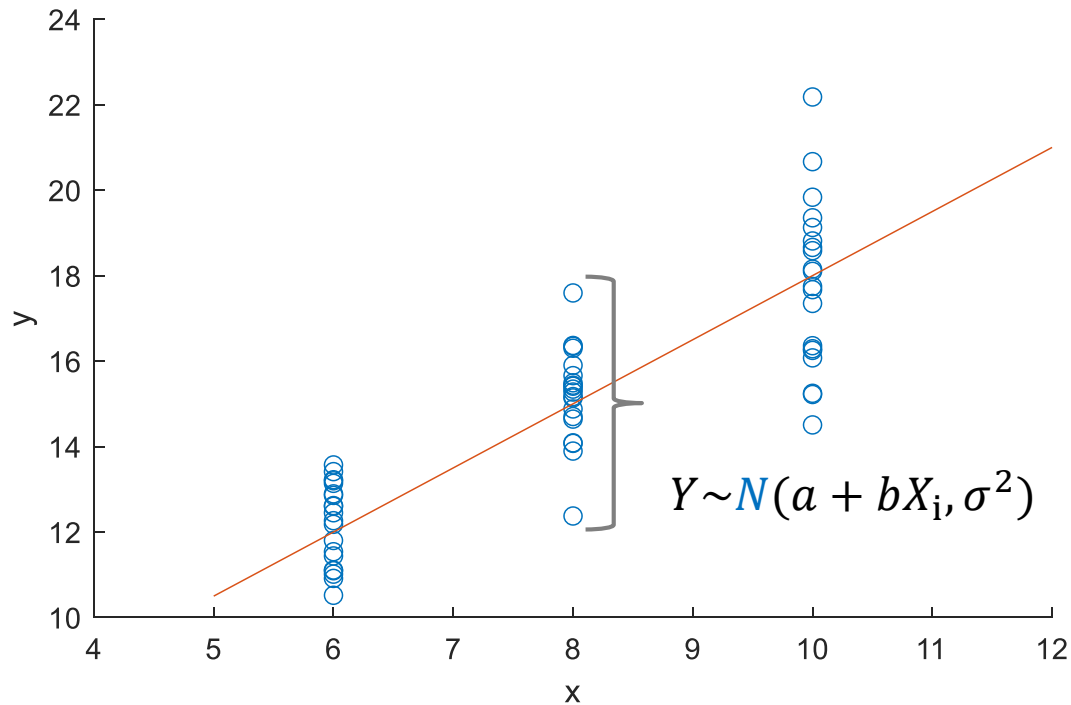
- Resultierender Trend: $\hat{g}_t = a_0 + a_1 \cdot t$

- Residuum: $\hat{e}_t = y_t - \hat{g}_t$

t = Zeit, \hat{y}_t = Schätzwert, $a_0 + a_1$ = Regressionsparameter

Wahrscheinlichkeitsverteilung der Zufallsvariablen Y_i ist den Annahmen nach i.i.d. Normalverteilt

Annahme: $Y_i \sim N(\mu, \sigma)$



- Für den **Erwartungswert** von Y_i gilt:

$$E(Y_t) = E(a + bX_i + \epsilon_i) = a + bX_i$$

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

- und die **Varianz** von Y_i gilt:

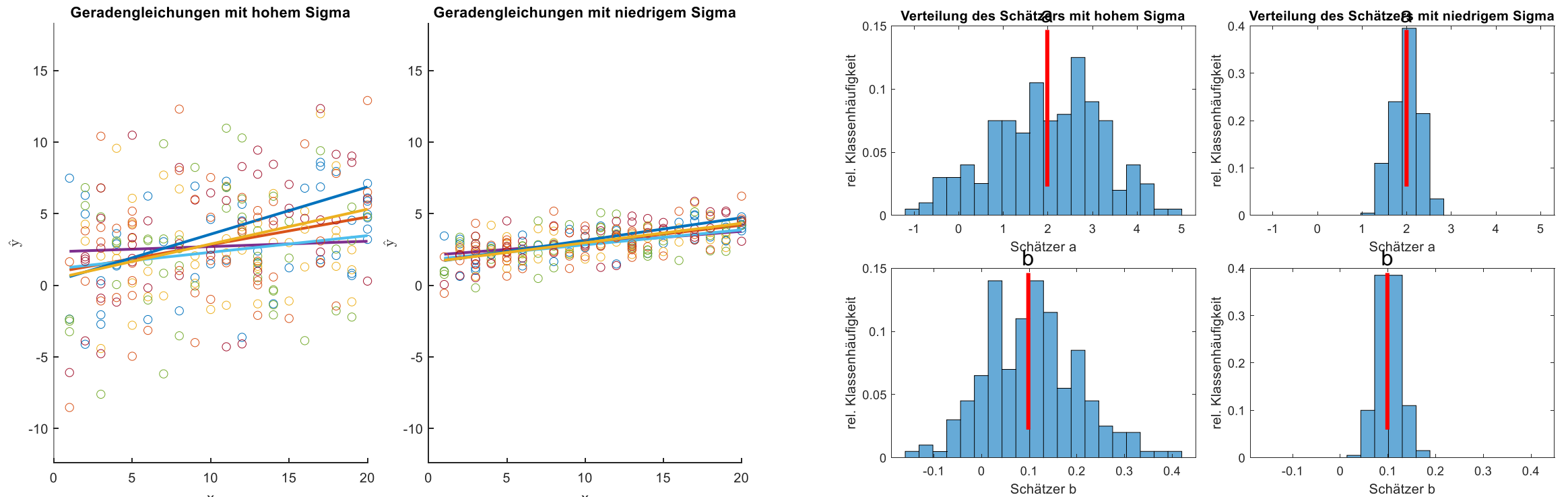
$$\begin{aligned} \text{var}(Y_i) &= E[(Y_i - E(Y_i))^2] \\ &= E[(a + bX_i + \epsilon_i - E(Y_i))^2] \\ &= E[(\epsilon_i)^2] = \sigma^2 \end{aligned}$$

- Schätzer für die Varianz der Residuen

$$[\hat{\sigma}^2] = \frac{1}{n-2} \sum_{i=1}^N e_i^2 = \frac{1}{n-2} \sum_{i=1}^N (y_i - \underbrace{\bar{y} - \hat{b}\bar{x} - \hat{b}x_i}_a)^2$$

Die Varianz der Störgröße hat einen unmittelbaren Einfluss auf die Varianz der Schätzer

- Im vorliegenden Fall wurden einzelne Stichproben (40 Datenpunkte) des Modells $Y \sim N(a + bX_t, \sigma^2)$ vorgenommen. Die Parameter $a = 2$ und $b = 0,1$ wurden nicht verändert. Je Stichprobe wurden die Parameter a und b geschätzt.
- Durch die Störgröße im Modell deren Varianz ebenfalls σ^2 beträgt, resultieren für jede Stichprobe leicht abweichende Parameter.
- Hieraus resultiert die Frage, welche Aussagekraft eine Schätzung hat, die nur eine Stichprobe zugrunde liegen hat

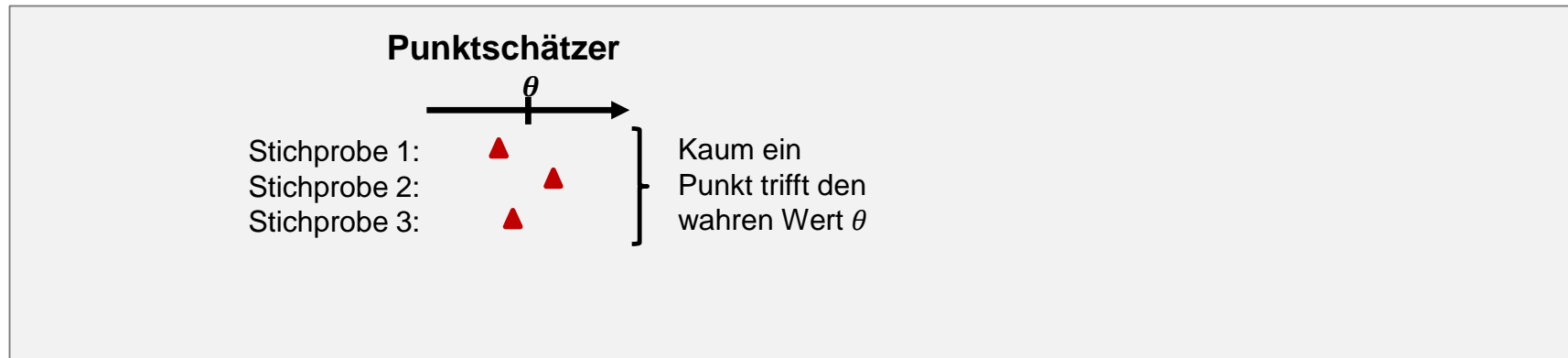


Überblick und Grundbegriffe: Schätzverfahren

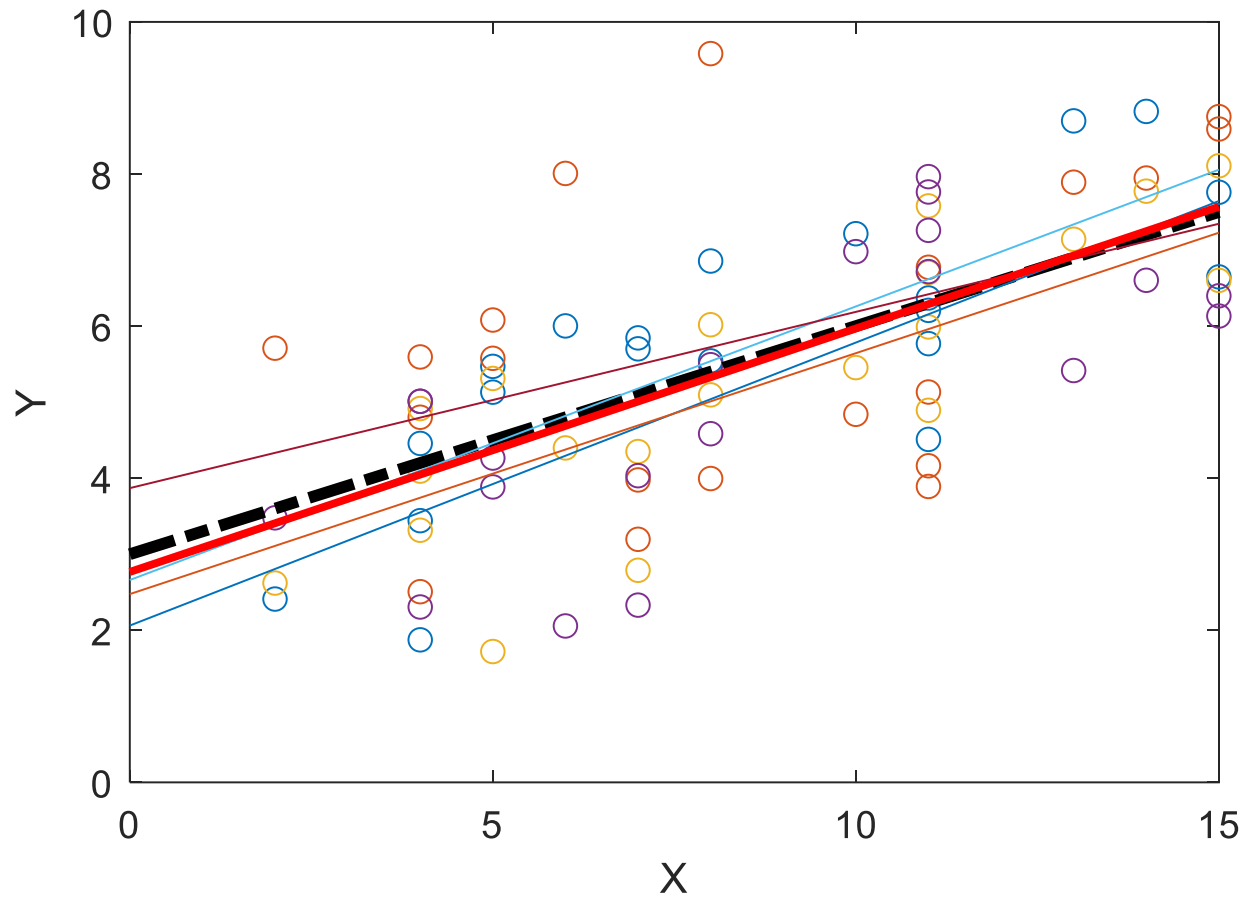
■ Punktschätzer:

Ein Punktschätzer für den zu schätzenden Parameter θ ist eine **Zufallsvariable**, welche mit Hilfe von Stichprobenvariablen Y_1, Y_2, \dots, Y_n berechnet wird.

$\hat{\theta}$ = Punktschätzer für den Parameter θ



In Abhängigkeit der vorliegenden Stichprobe können unterschiedliche Ausgleichsgeraden resultieren



Wovon hängt grundsätzlich die Variabilität von a und b ab?

- Stichprobenumfang n
- Varianz σ^2 der Störgröße
- Spannweite von x

Aus den formulierten Eigenschaften für y lässt sich die Verteilung der KQ-Schätzer herleiten

Lageparameter des Schätzers

Schätzer	\hat{a}	\hat{b}
Erwartungswert	a	b
Varianz	$\sigma^2 \frac{1}{n} \frac{\sum x_i^2}{S_{xx}}$	σ^2 / S_{xx}
Verteilung	$N\left(a, \sigma^2 \frac{1}{n} \frac{\sum x_i^2}{S_{xx}}\right)$	$N(b, \sigma^2 / S_{xx})$

■ Schätzer für die Varianz der Störterme $\hat{\sigma}^2 = \frac{S_{ee}}{n-2}$

Varianz der Störgröße und Regressand

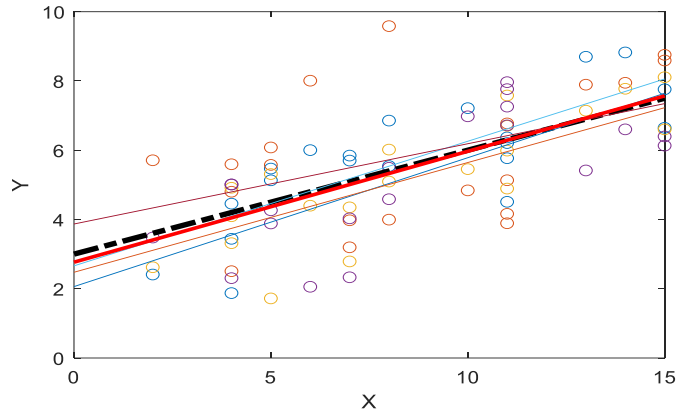
- σ^2 ist die unbekannte “wahre” Varianz der Störgröße e und y
- 1. Die Varianz der geschätzten Parameter \hat{a} , \hat{b} ist somit umso größer je größer die „wahre“ Varianz der Störgröße ist.
- 2. Je größer der Stichprobenumfang, desto kleiner ist die Varianz vom Parameter \hat{a}

S_{xx} ist die Variation der exogenen Variable: $S_{xx} = \sum (x_i - \bar{x})^2$

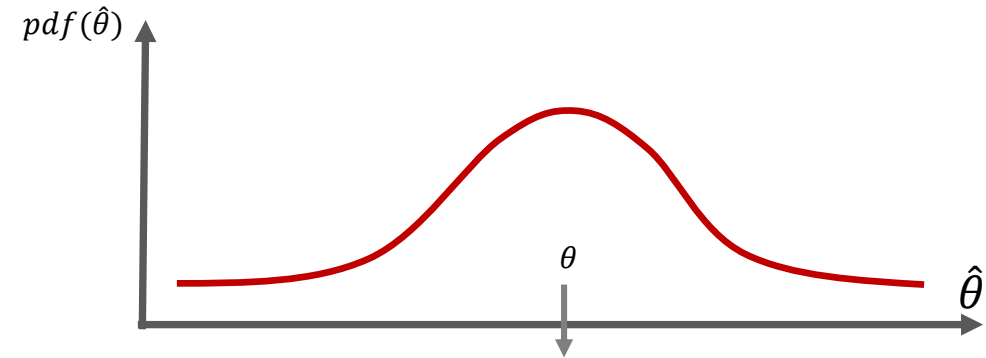
- Die Variation S_{xx} wirkt sich auch minimierend auf die Varianz des Schätzers \hat{b} aus.

Zusammenfassung der Idee und Ausgaben einer linearen Regression

Der „wahre“ Zusammenhang zwischen X und Y wird durch eine Störgröße überlagert



Die Parameterschätzer anhand der jeweiligen Stichprobe können als Zufallsvariablen aufgefasst werden

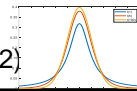


Aus der Schätzung lassen sich Punkt und Intervallschätzer* mit ableiten

Intervallschätzer mit Konfidenz $1 - \alpha$

$$[\hat{b} - t_{1-\alpha/2} * std(\hat{b}); \hat{b} + t_{1-\alpha/2} * std(\hat{b})]$$

$t_{1-\alpha/2}$ = Quantilswert der t – Verteilung $t \sim t(N-2)$



Hierzu wird die Varianz der Residuen geschätzt: $\hat{\sigma}^2 = S_{ee}/(N - 2)$

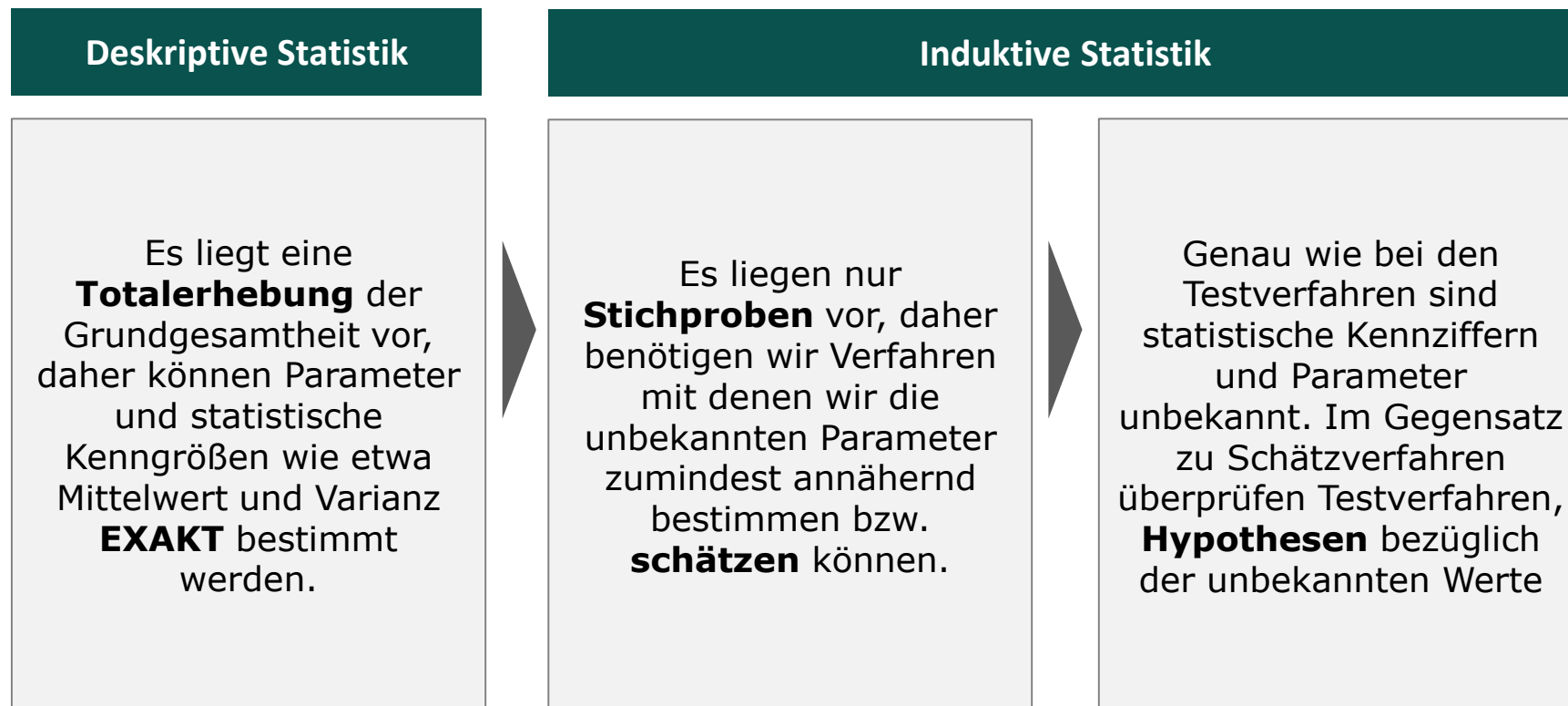
Beispielhaftes Ergebnis

Parameter	Punktschätzer (Est)	Standardabweichung $std(\theta)$ (SE)	Intervall
a	1,02	0,74	$[Est - t_{1-\frac{\alpha}{2}} \cdot SE;$ $Est + t_{1-\frac{\alpha}{2}} \cdot SE];$
b	0,12	0,02	

$$t_{\frac{\alpha}{2}} = \text{tinv}(1 - \frac{\alpha}{2})$$

Anwendungszweck und Eigenschaften der Statistischen Testverfahren

- Hypothesen und Gegenhypothesen werden üblicherweise mit der **Nullhypothese** H_0 und das Gegenteil mit der **Alternative** H_1 bezeichnet



Hypothesentest auf Basis der Gegenhypothese

- Zielsetzung des Hypothesentest:
Eine Vermutung/Behauptung soll gestützt werden.
- Vorgehensweise:
Nachweis, dass die Gegenhypothese widerlegt werden kann.
- Nullhypothese zweiseitiger Test
Dabei ist μ der tatsächliche Erwartungswert der Grundgesamtheit und μ_0 der hypothetische Wert.

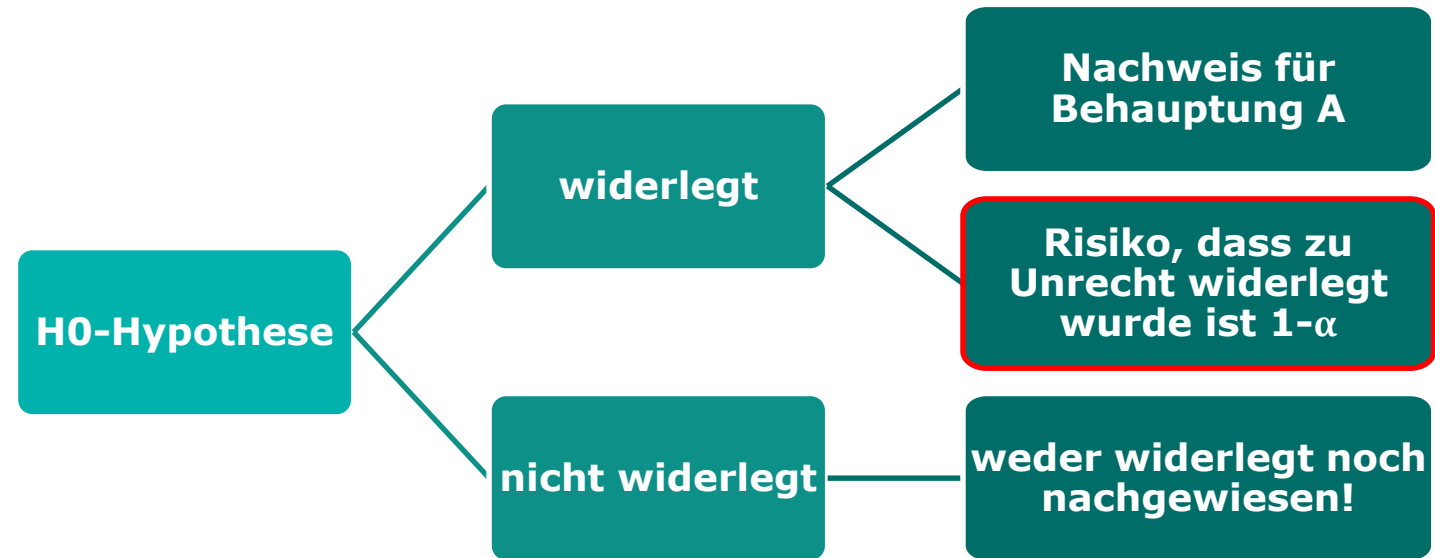
$$H_0: \mu = \mu_0$$

- Nullhypothese einseitiger Test
Hier befindet sich die Alternative jeweils nur auf eine Seite.

$$H_0: \mu \leq \mu_0 \quad \text{oder} \quad H_0: \mu \geq \mu_0$$

Gebrauch von Signifikanztests: Konstruktion der Nullhypothese

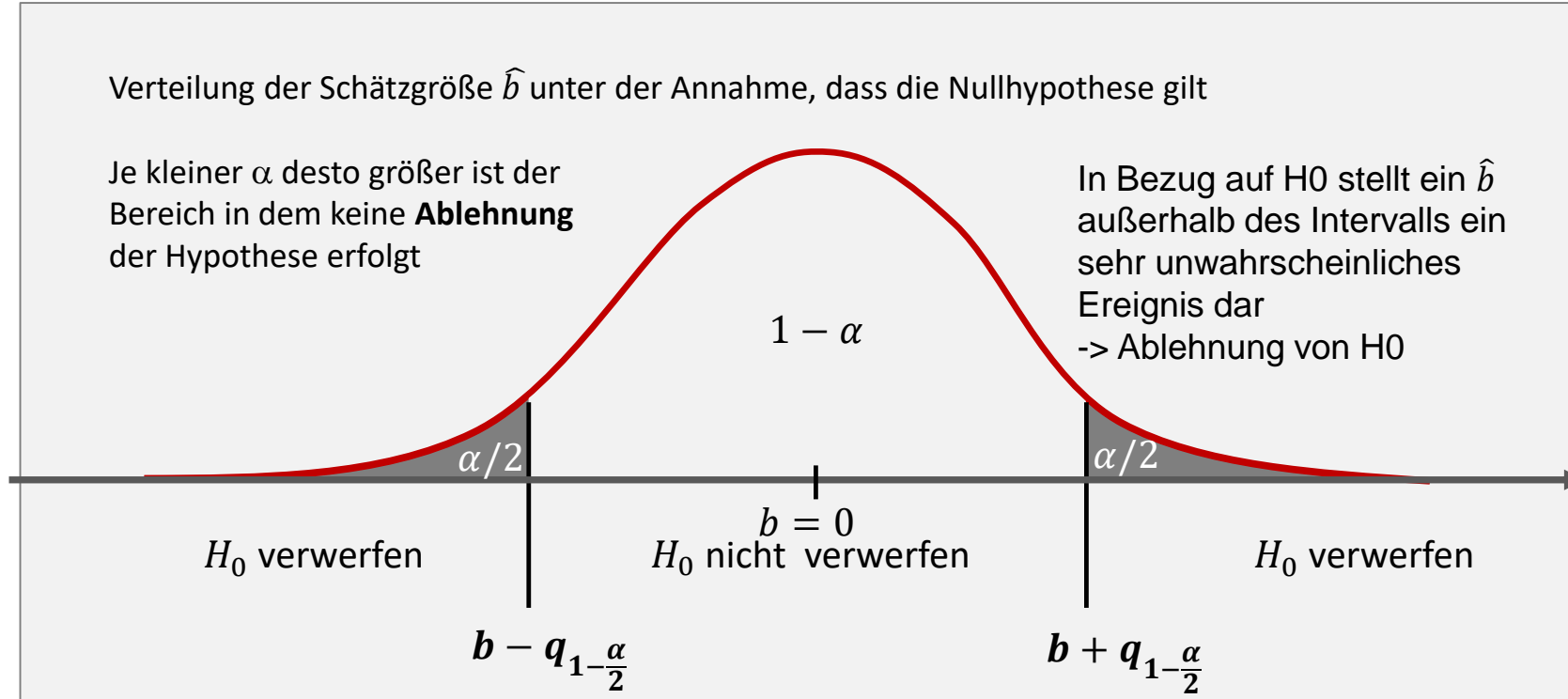
- Signifikanztest ist ein Ausschlussverfahren, welcher einem indirekten Beweis gleichkommt.
- Wollen wir also die Behauptung A als richtig nachweisen, so muss gezeigt werden, dass das Gegenteil von A Falsch ist:
- H_0 = Gegenteil von Behauptung A



- Je kleiner der Wert α festgelegt wird, desto sensibler verhält sich der Test und es wird öfters die Behauptung weder widerlegt noch nachgewiesen.

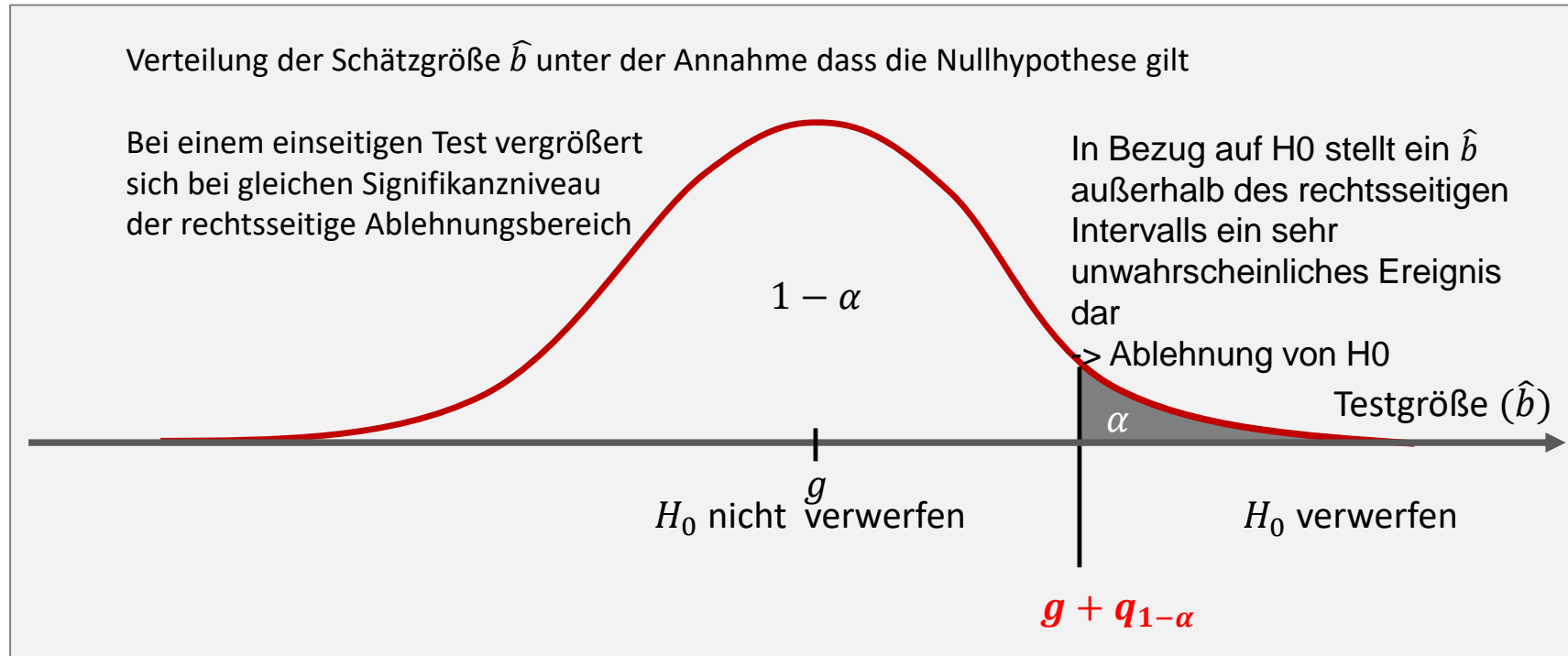
Grundidee des Zweiseitigen Hypothesentest

- Bildung einer Nullhypothese $H_0: b = 0$ mit $\alpha =$ **Signifikanzniveau** i.d.R. **5%** -> **Risiko falscher Widerlegung**
- Bereits abgeleitet wurde die Verteilungsfunktion von \hat{b} : $\hat{b} \sim N(b = 0, \text{var}(\hat{b}))$
- **Fragestellung:** ab welcher Ausprägung des Schätzers \hat{b} muss die Nullhypothese abgelehnt werden?



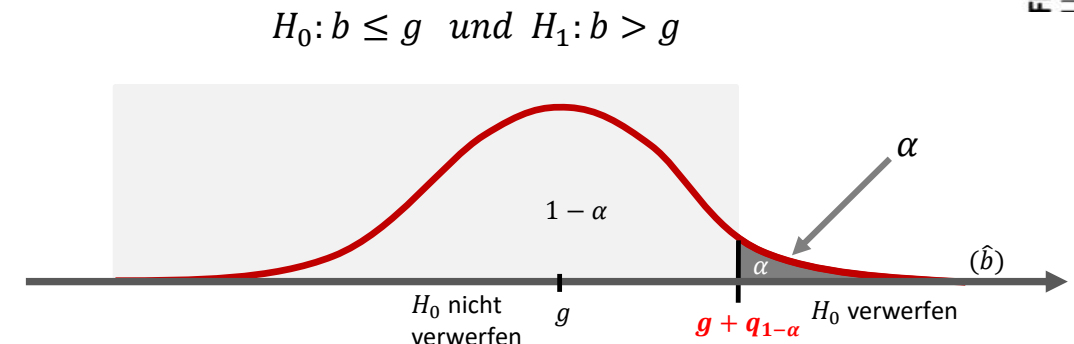
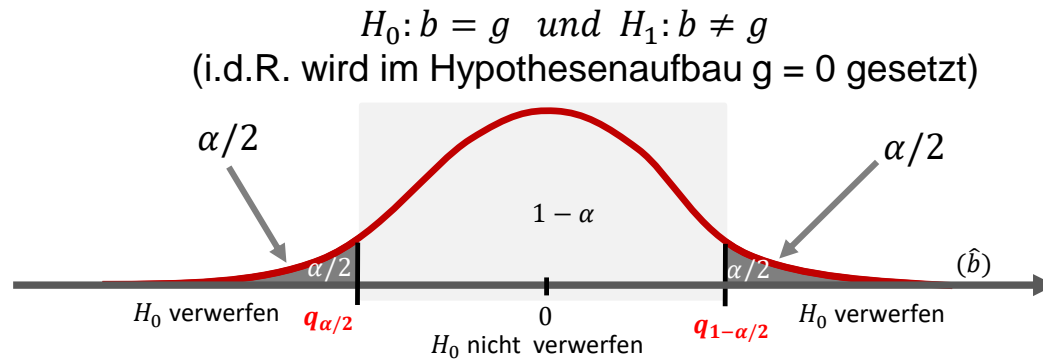
Grundidee des einseitiger Hypothesentest

- Bildung einer Nullhypothese $H_0: b \leq g$ mit $\alpha = \text{Signifikanzniveau}$ i.d.R. 5% -> Risiko erster Art
- **Fragestellung:** ab welcher Ausprägung des Schätzers \hat{b} muss die Nullhypothese abgelehnt werden?



Der Aufbau der Teststatistik hängt von der konkreten Fragestellung ab

- Schritt 1: Formulierung von H_0 und H_1 und Festlegung des Signifikanzniveaus α



- Schritt 2: Formulierung der Testgröße unter H_0

$$t = \frac{\hat{b} - g}{\text{std}(\hat{b})} \quad \text{mit Freiheitsgraden } \vartheta = n - 2$$

- Schritt 3: Ermittlung des kritischen Wertes $t_{1-\alpha/2}$

$$P\{t_{\alpha/2} \leq t \leq t_{1-\alpha/2}\} = 1 - \alpha$$

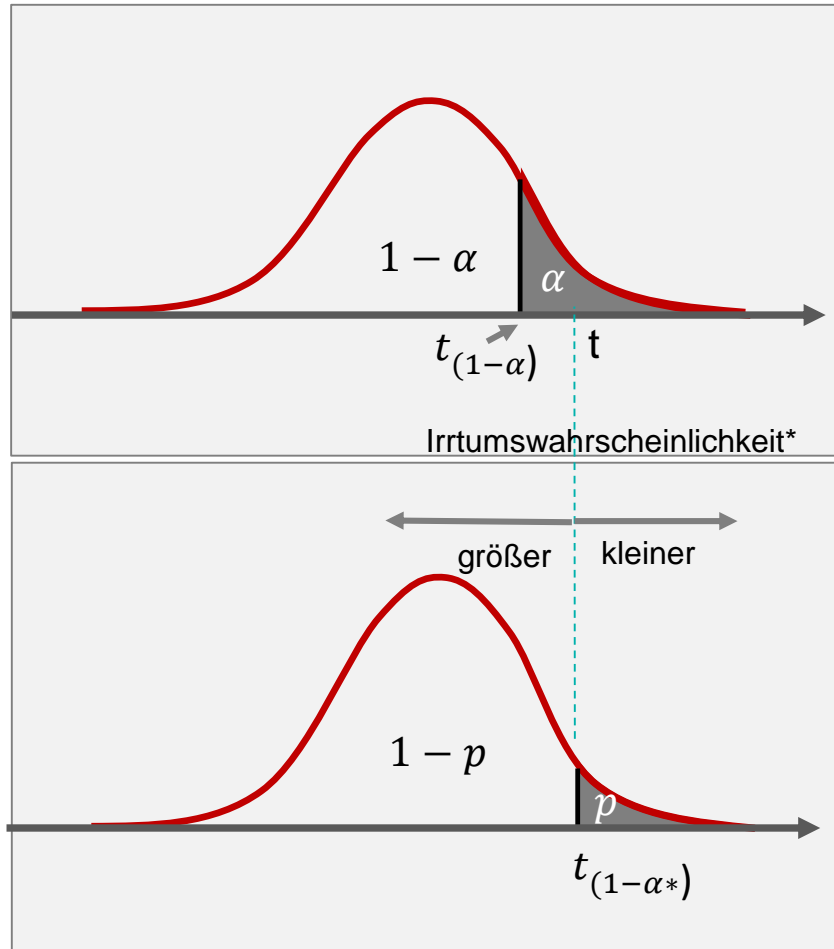
$$P\{t \leq t_{1-\alpha}\} = 1 - \alpha$$

- Schritt 4: Prüfung ob Testgröße kritischen Wert überschreitet/unterschreitet

Wenn $|t| > t_{1-\alpha/2} \rightarrow$ **Ablehnung von H_0**

Wenn $t > t_{1-\alpha} \rightarrow$ **Ablehnung von H_0**

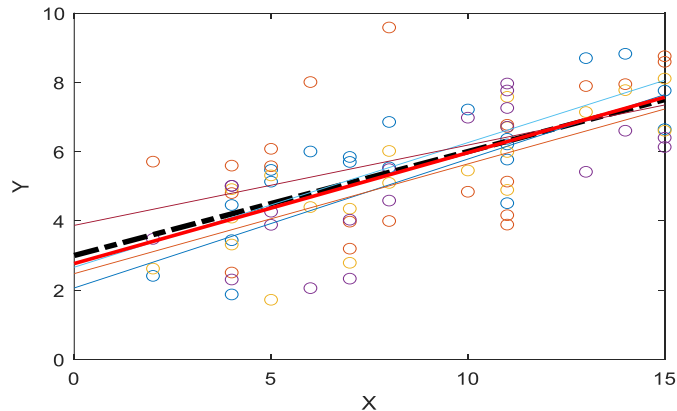
Der P-Wert wird häufig in statistischen Programmen ermittelt und gibt das Signifikanzniveau an, zu dem H_0 gerade noch nicht widerlegt wird



- Häufig wird in der Statistik der sogenannte p-Wert angegeben
- Dieser drückt bei berechneter Testgröße t das Signifikanzniveau aus, welches noch nicht zur Widerlegung von H_0 führt.
- Der Wert p bringt gegenüber dem t -Wert direkter zum Ausdruck, wie deutlich eine H_0 -Hypothese abgelehnt oder akzeptiert worden ist
- Der Wert p kann als Irrtumswahrscheinlichkeit interpretiert werden, H_0 fälschlicherweise abzulehnen.

Im Rahmen des Hypothesentests wird die Nullhypothese, dass der Parameter Null beträgt geprüft:

Der „wahre“ Zusammenhang zwischen X und Y wird durch eine Störgröße überlagert

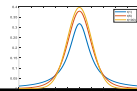


Aus der Schätzung lassen sich Punkt und Intervallschätzer* mit ableiten

Intervallschätzer mit Konfidenz $1 - \alpha$

$$[\hat{b} - t_{1-\alpha/2} * \text{std}(\hat{b}); \hat{b} + t_{1-\alpha/2} * \text{std}(\hat{b})]$$

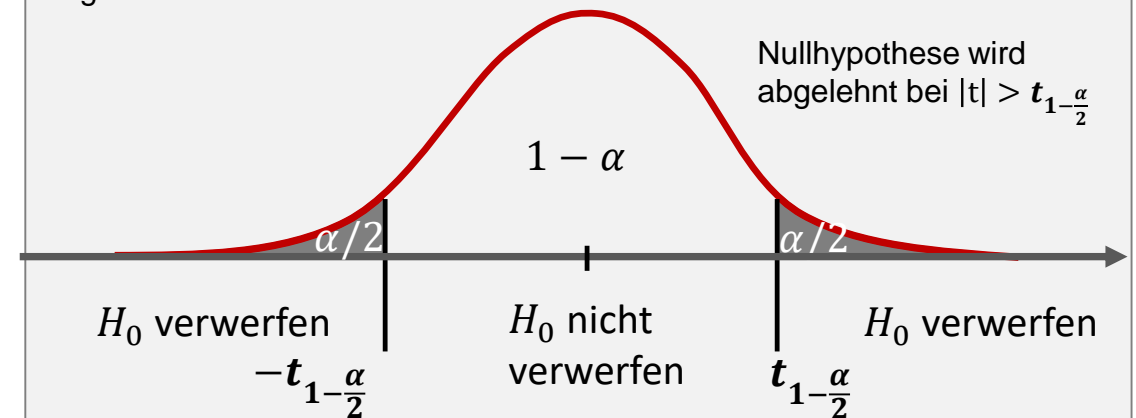
$t_{\alpha/2}$ = Quantilswert der t-Verteilung $t \sim t(N-2)$



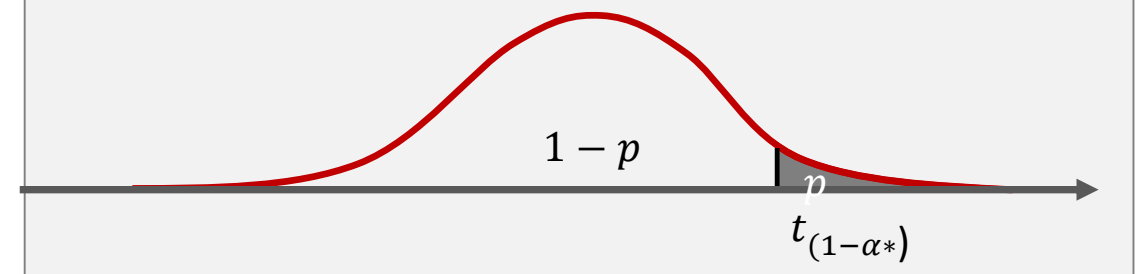
Hierzu wird die Varianz der Residuen geschätzt: $\hat{\sigma}^2 = S_{ee}/(N - 2)$

Hypothesentest P-Wert

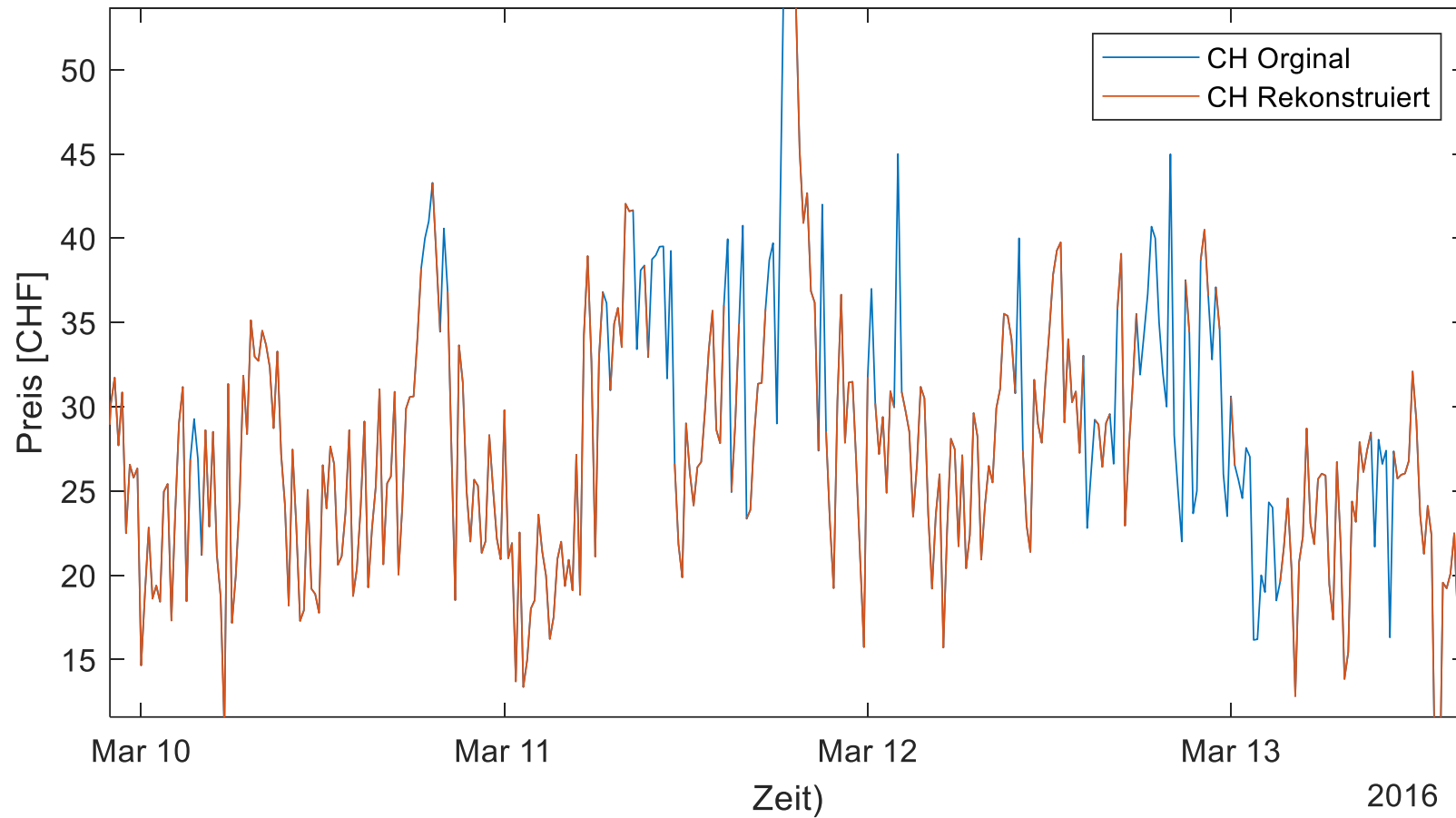
Kann die Hypothese „Parameter ist Null“ mit entsprechender Konfidenz abgelehnt werden?



Bei welcher Signifikanz p würde meine Nullhypothese gerade nicht widerlegt?



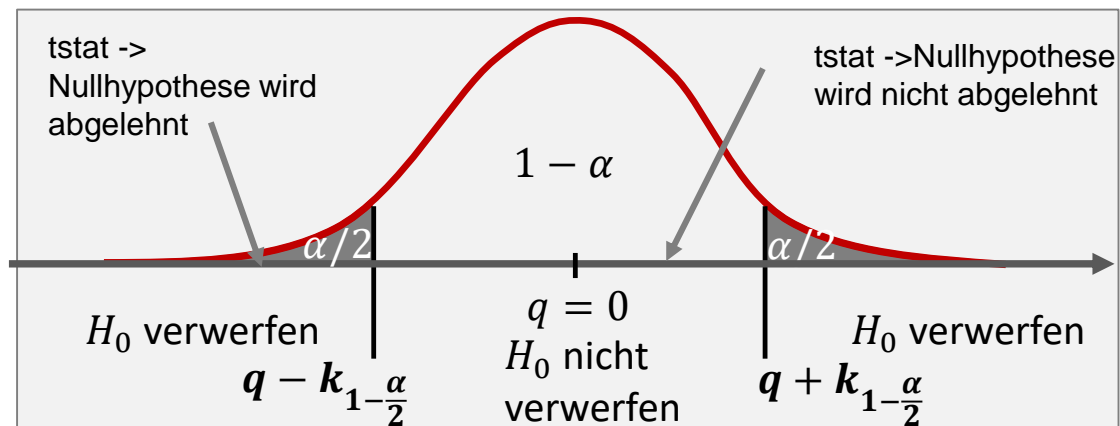
Anwendungsbeispiel 2: Datenimputation Regression am Beispiel der schweizer Strompreise



Im Rahmen des Hypothesentests wird die Nullhypothese, dass der Parameter Null beträgt geprüft:

Hypothesentest

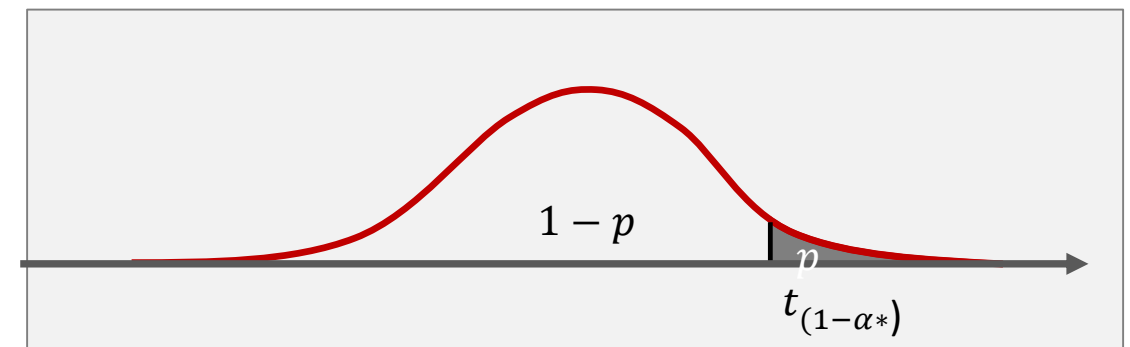
Frage: Kann die Hypothese „Parameter ist Null“ mit entsprechender Konfidenz abgelehnt werden?



	1	2	3	4
	Estimate	SE	tStat	pValue
1 (Intercept)	1.0169	0.7419	1.3708	0.1873
2 x1	0.1201	0.0214	5.6048	2.5568e-05

P-Wert

Frage: bei welcher Signifikanz p würde meine Nullhypothese gerade nicht widerlegt?



Interpretation: ein p in der Nähe von 0 entspricht einer extrem kleinen Irrtumswahrscheinlichkeit beim Ablehnen der Hypothese.

	1	2	3	4
	Estimate	SE	tStat	pValue
1 (Intercept)	1.0169	0.7419	1.3708	0.1873
2 x1	0.1201	0.0214	5.6048	2.5568e-05

1. Einfache lineare Regression

2. Case Study 1: Datenlücken bereinigen

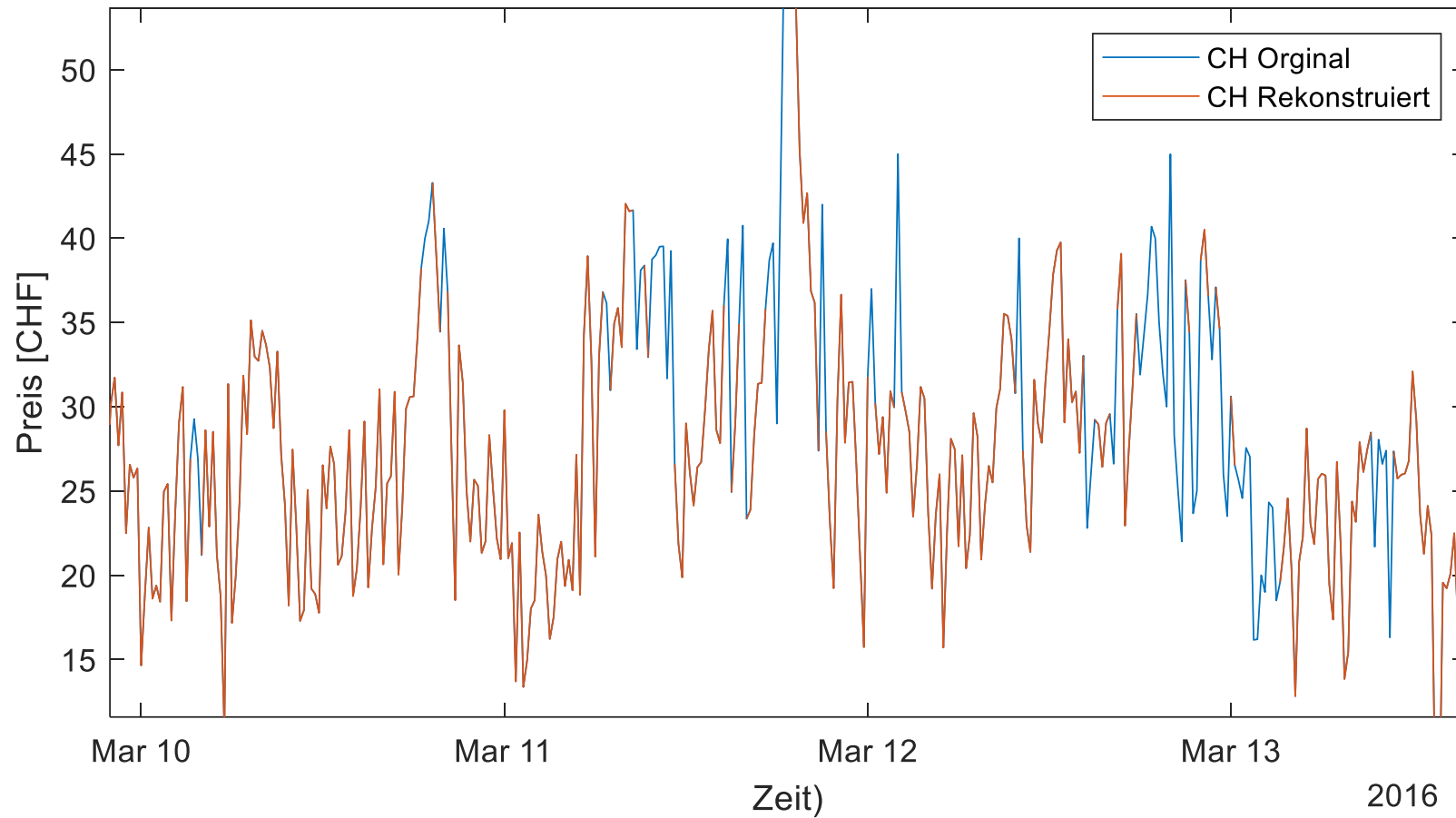
3. Case Study 2: Winkelerkennung von WKAs

4. Multiple lineare Regression

5. Case Study 3: Multiple Regression Strompreis auf Brennstoffe und CO2



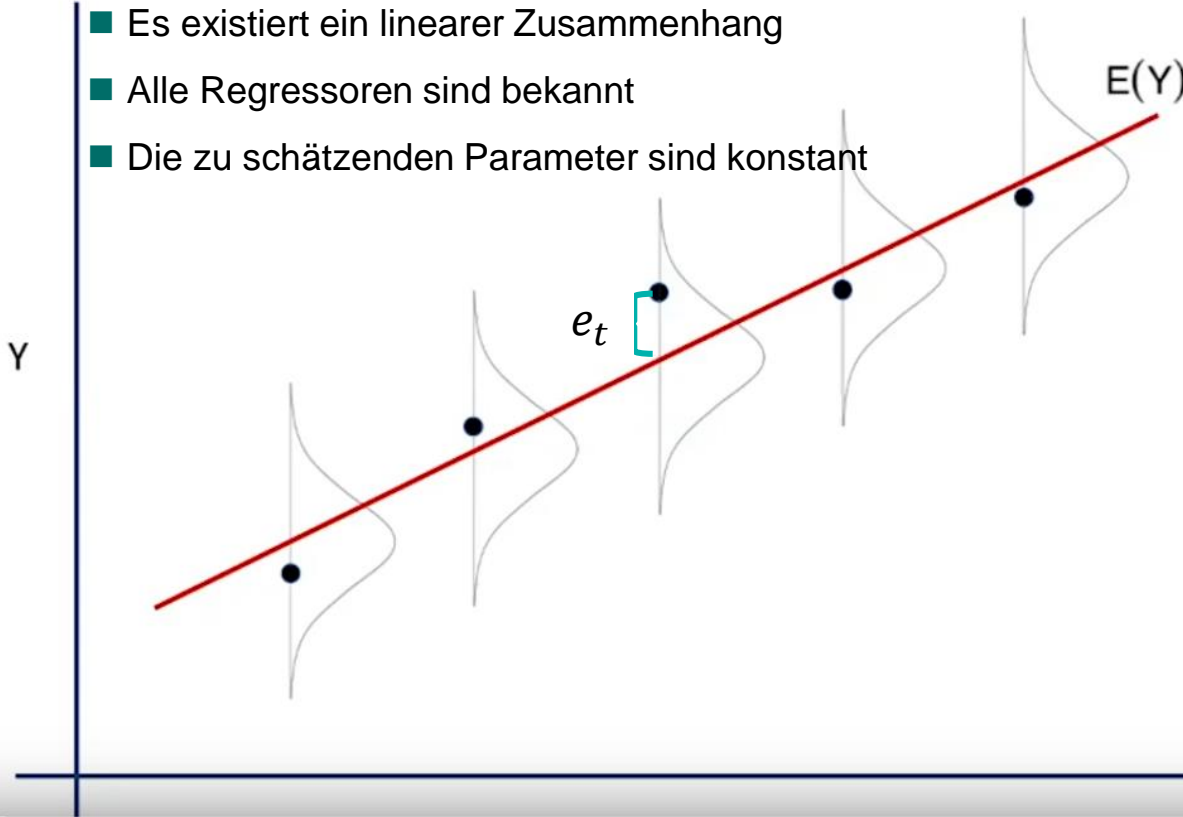
Anwendungsbeispiel 2: Datenimputation Regression am Beispiel der schweizer Strompreise



Es gibt eine Vielzahl von Annahmen zu den Eigenschaften des resultierenden Modells und der Störgröße

Annahmen bzgl. des Modells

- Es existiert ein linearer Zusammenhang
- Alle Regressoren sind bekannt
- Die zu schätzenden Parameter sind konstant



- Der Erwartungswert beträgt Null:

$$E(\epsilon)=0$$

- Die Störgröße hat eine konstante Varianz (*Homoskedastizität*):

$$\text{Var}(\epsilon)=\sigma^2$$

- Die Störgrößen sind nicht korreliert (fehlende *Autokorrelation*):

$$\text{Cov}(\epsilon_1, \epsilon_2)=0$$

- Die Störgröße ist unabhängig identisch normalverteilt:

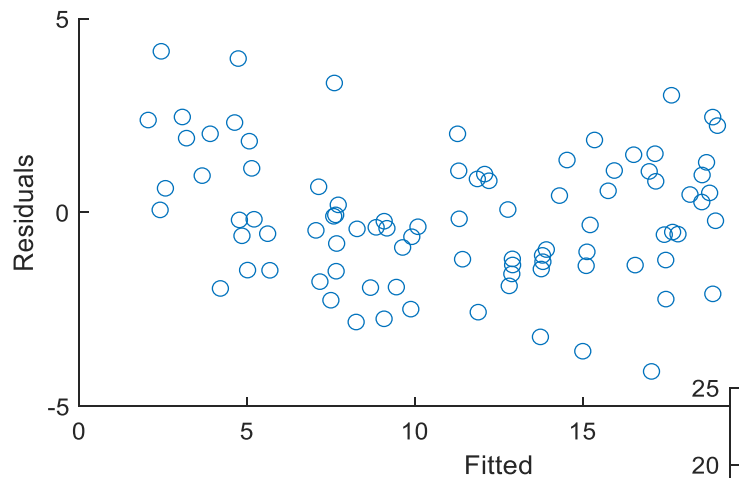
$$\epsilon \sim N(0, \sigma^2)$$

Im Rahmen der Modellbeurteilung werden müssen insb. die Residuen e hinsichtlich ihrer Eigenschaften evaluiert werden

Residuen haben konstante Varianz und um Null verteilt $Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t) = 0$	Unabhängigkeit $Cov(\varepsilon_i, \varepsilon_j) = 0 \text{ für } i \neq j$	Normalverteilt $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$
Graphische Analyse		
Plot Residuen gegenüber X und Y	ACF, PACF	QQ-Plot
Anwendung von Testverfahren		
	Durbin Watson Test	Anderson-Darling-Test Jaque Bera Test

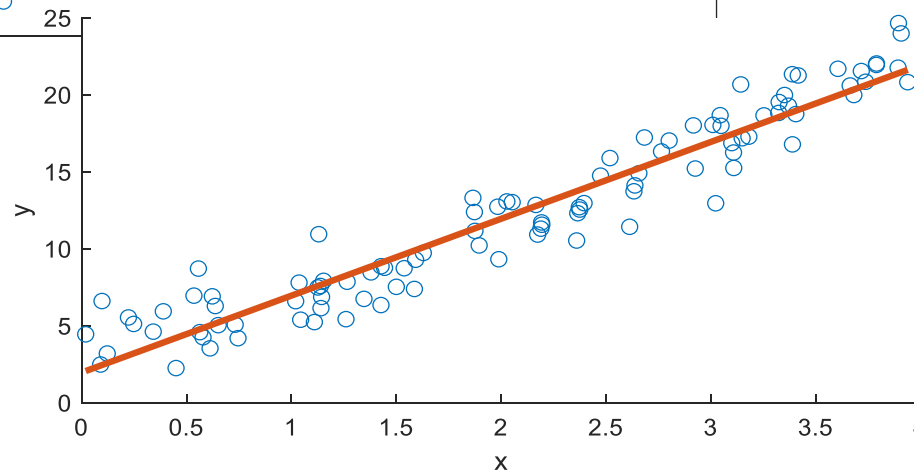
Abbildungen dienen zur visuellen Analyse von Nichtlinearität, nicht konstanter Varianz und Ausreißern

Residual- Fit-Plot

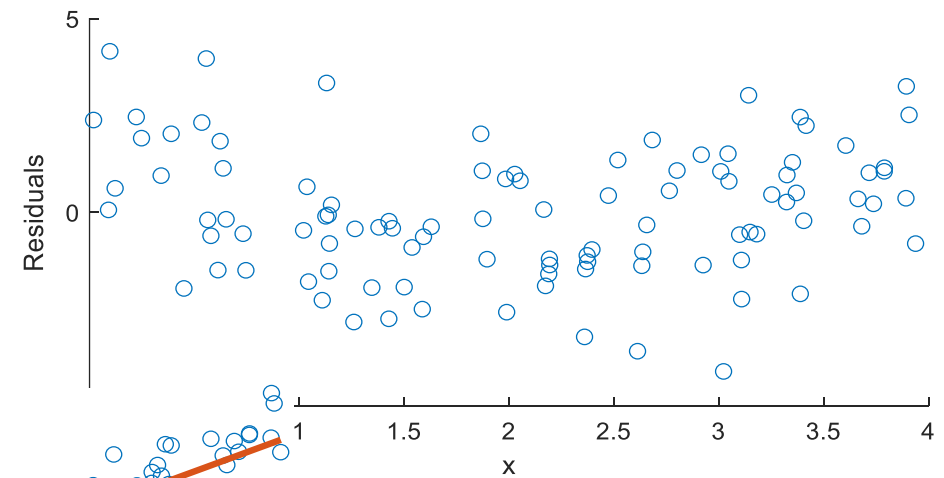


Formulieren die Residuen ein horizontales Band um Null herum?

Gibt es Ausreißer?



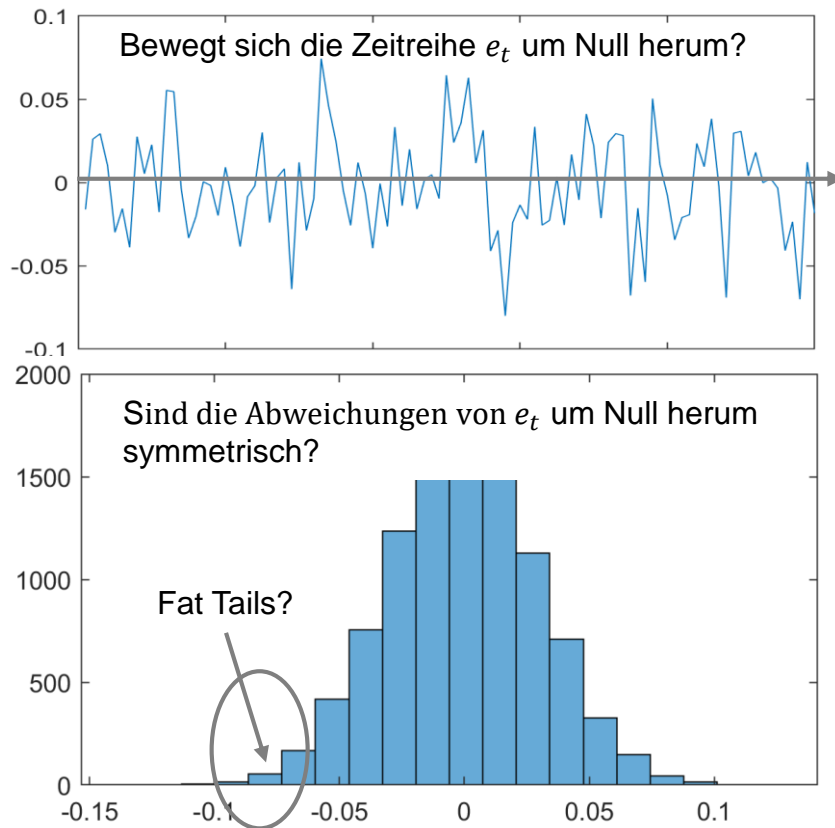
Residual-Regressor-Plot



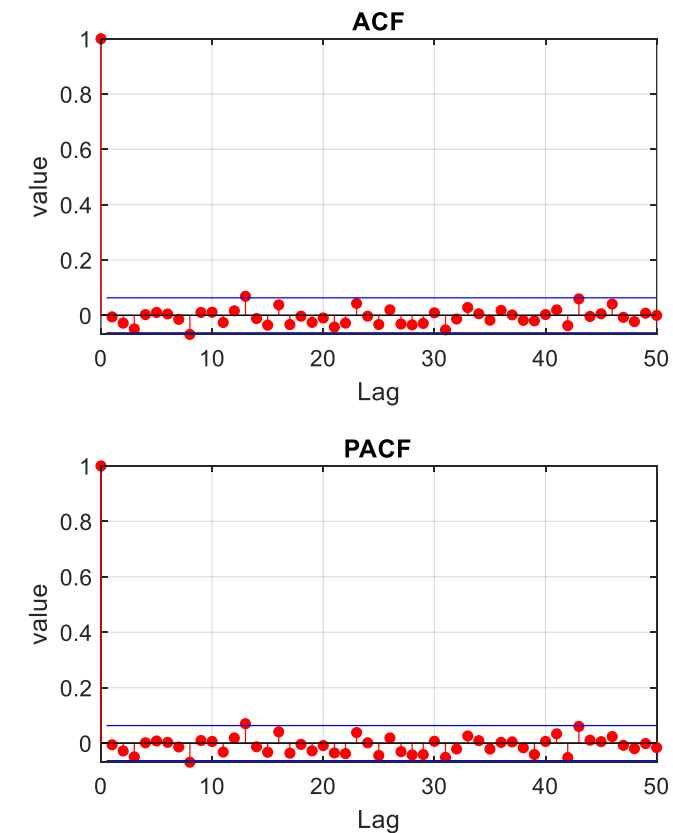
Bei einer einfachen linearen Regression bietet dieser Plot keine Zusatzinformationen

Residuals vs. Order Plot kann visuell insb. bei Zeitreihen Trends und Autokorrelation aufspüren und somit die i.i.d. Annahme überprüfen

Verteilung



Autokorrelation



Die Unabhängigkeit kann mit Hilfe des Durbin-Watson Test überprüft werden

- Berechnung der Prüfgröße:

$$DW = \frac{\sum_{t=2}^N (e_t - e_{t-1})^2}{\sum_{t=1}^N e_t^2} = \frac{\sum_{t=2}^N e_t^2 + e_{t-1}^2 - 2e_t e_{t-1}}{\sum_{t=1}^N e_t^2}$$

- Die Auto-Korrelation beeinflusst maßgeblich den letzten Term des Zählers

$$DW \approx \begin{cases} 0 & \text{falls die Residuen stark positiv korreliert sind} \\ 2 & \text{falls die Residuen unkorreliert sind} \\ 4 & \text{falls die Residuen stark negativ korreliert sind.} \end{cases}$$

Die Unabhängigkeit der Innovationen kann auch mit Hilfe des Ljung Box Pierce Test analysiert werden

- Der Ljung-Box-Pierce-Test zur Überprüfung der Unkorreliertheit der Residuen prüft die ersten k Korrelationen gemeinsam auf Null.
- Nullhypothese: Residuen sind nicht korreliert

- Prüfgröße:

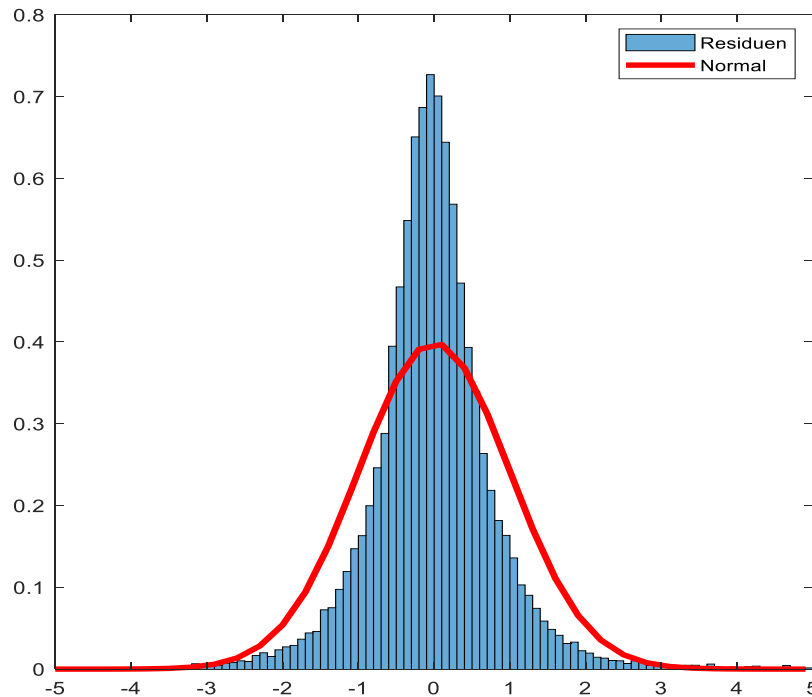
$$Q = n(n + 2) \sum_{\tau=1}^k \frac{\hat{\rho}_{\tau}^2}{n - \tau}$$

geschätzte Autokorrelation

- Q ist approximativ χ^2 -verteilt () mit k-p-Freiheitsgraden
- k sollte größer als die höchste berücksichtigte Modellordnung sein. Faustregel: $k \approx 2\sqrt{n}$
- Die Nullhypothese wird abgelehnt, falls Q größer ist als der kritische Wert, ermittelt aus der χ^2 -Verteilung mit [k-p]-Freiheitsgraden zu einem Signifikanzniveau von α

Verletzungen der Annahmen: Der Jarque Bera nutzt die Eigenschaft, dass jede Normalverteilung eine Schiefe von 0 und Kurtosis von 3 besitzt

Jarque Bera Test



Signifikanzniveau α	0,10	0,05	0,01	0,001
Quantil $\chi^2_{2;1-\alpha}$	4,61	5,99	9,21	13,82

- Der Test basiert auf der Schiefe und Wölbung der Normalverteilung und liefert zwei Testwerte b_1 und b_2

- Parameter für Schiefe:

$$\sqrt{b_1} = \frac{1/n \sum e_t^3}{(1/n \sum e_t^2)^{3/2}} \sim N(0, \sqrt{6n^{-1}})$$

- Parameter für Kurtosis: 3

$$b_2 = \frac{1/n \sum e_t^4}{(1/n \sum e_t^2)^2} \sim N(3, \sqrt{24n^{-1}})$$

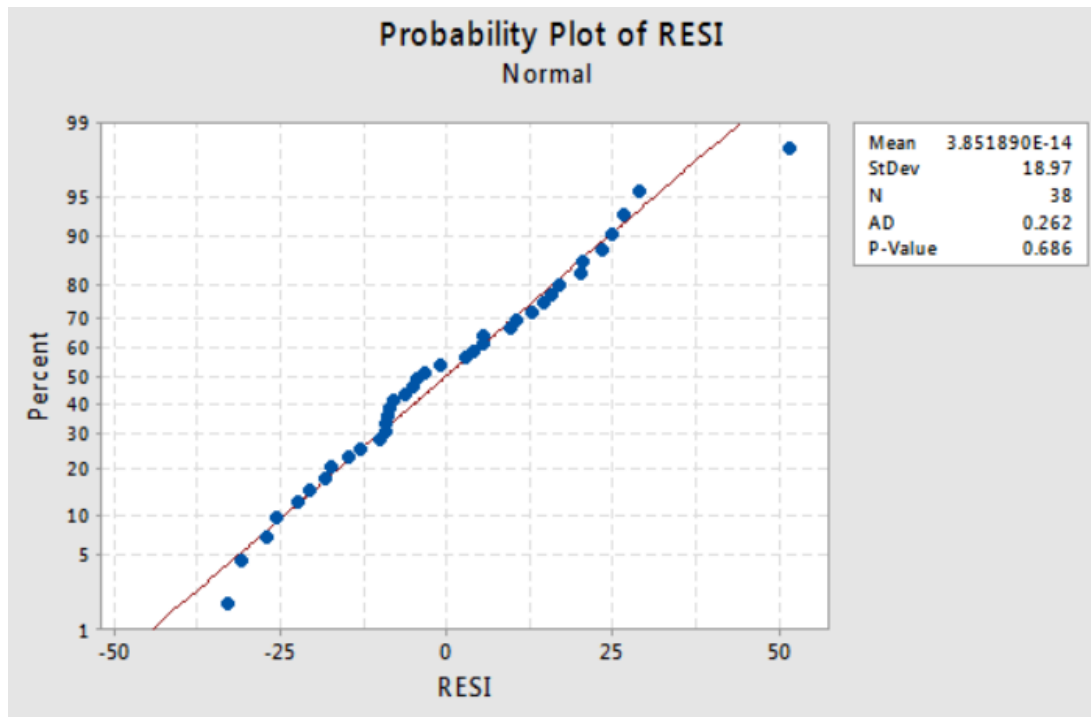
- Diese werden in eine χ^2 - verteilte Teststatistik mit zwei Freiheitsgraden zusammengefasst:

$$JB = \frac{n}{6} b_1 + \frac{n}{24} b_2^2$$

- Ablehnung der H_0 -Hypothese falls JB kritischen Wert übersteigt.

Der Anderson-Darling-Test misst, ob eine Datenstichprobe aus einer Normalverteilung stammt.

Vergleich Abstand der Verteilungen



- Hierbei wird entlang der empirischen Verteilungsfunktion der Residuen der Abstand zur theoretischen Verteilung der H_0 – Hypothese gemessen.

■ Test:

- H_0 : Residuen folgen der Normalverteilung
- H_1 : Residuen sind nicht normalverteilt

- Es resultiert folgende Teststatistik:

$$A_n^2 = -n - \sum_{i=1}^n \frac{2i-1}{n} [\ln(F(X_i)) + \ln(1 - F(X_{n+1-i}))]$$

- X sind die geordneten Residuen

- H_0 wird abgelehnt sobald der p-Wert der Teststatistik kleiner ist als das Signifikanzniveau α (z. B. 0.05)

- Teil I: lineare Regression
 - Spezifikation
 - Punktschätzung
 - Indikatoren für die Qualität der Schätzung
 - Intervallschätzer
 - Hypothesentest
 - **Prognose**

Berechnung der Punktprognose

- Der Aufbau des linearen Modells lautet:

$$\hat{y}_t = \hat{a} + \hat{b}x_t$$

- Hierbei wird vorausgesetzt, dass x_t bekannt ist; beispielsweise mit der Ausprägung x_0

$$\hat{y}_0 = \hat{a} + \hat{b}x_0$$

- Für die Abweichung zwischen prognostizierten und realem Wert y_0 lässt sich folgende Aussage ableiten:

$$\hat{y}_0 - y_0 = (\hat{a} - a) + (\hat{b} - b)x_0 - u_0$$

- Es können zwei Ursachen für Prognosefehler geben
 1. Die Störgröße u kann von 0 abweichen
 2. Die Schätzer der Parameter können vom wahren Wert abweichen

Verlässlichkeit der Punktprognose

- Da \hat{y}_0 von den geschätzten Zufallsvariablen \hat{a} und \hat{b} abhängt, kann die Prognose in Abhängigkeit der unterlegten Stichprobe unterschiedlich ausfallen
- Für eine Angabe der Verlässlichkeit der Prognose \hat{y}_0 könnte dessen Wahrscheinlichkeitsverteilung abgeleitet werden.
- In diesem Fall wird aber auf die Verteilung von $\hat{y}_0 - y_0$ verwiesen

- Da die Punktschätzer erwartungstreu sind, gilt für den **Erwartungswert**:

$$E(\hat{y}_0 - y_0) = E[(\hat{a} - a) + (\hat{b} - b)x_0 - u_0] = E(\hat{a}) - E(a) + E(\hat{b}x_0) - E(bx_0) - E(u_0)$$

$$E(\hat{y}_0 - y_0) = a - a + b_{x0} - b_{x0} - 0 = 0$$

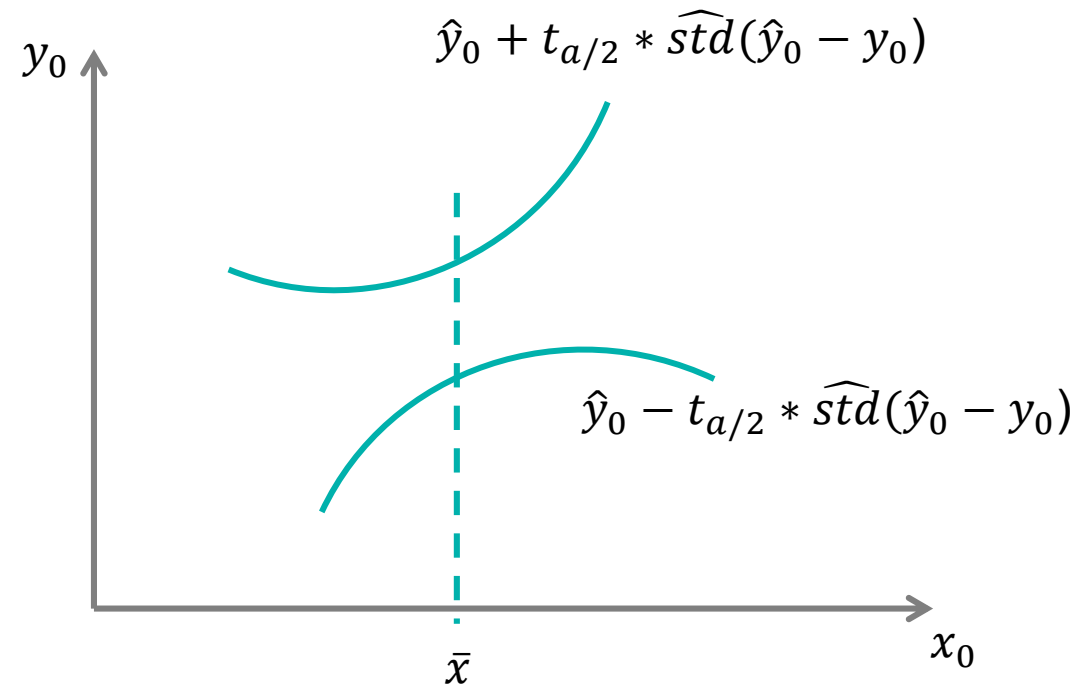
- Für die Varianz des Prognosefehlers gilt:

$$\widehat{var}(\hat{y}_0 - y_0) = \hat{\sigma}_0^2 \left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

- Interpretation:

1. Die Prognosequalität hängt von der Varianz der Störgröße e ab
2. Die Prognosequalität steigt mit Stichprobenumfang
3. Die Prognosequalität ist besser bei einem x nahe dem Mittelwert der Stichprobe von x

Graphische Ableitung des Prognoseintervalls



Ableitung des Prognoseintervalls

- Schritt 1: Schätzung von $\widehat{var}(\widehat{y}_0 - y_0)$ und Festlegung des Signifikanzniveaus

$$\widehat{var}(\widehat{y}_0 - y_0) = \widehat{\sigma}^2 \left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

- Schritt 2: Standardisierung des Prognosefehlers $(\widehat{y}_0 - y_0)$

$$t = \frac{(\widehat{y}_0 - y_0) - E(\widehat{y}_0 - y_0)}{\widehat{std}(\widehat{y}_0 - y_0)} = \frac{(\widehat{y}_0 - y_0)}{\widehat{std}(\widehat{y}_0 - y_0)}$$

- Schritt 3: Ermittlung des $t_{\alpha/2}$ – Wertes
- Schritt 4: Formulierung des Prognoseintervalls

$$PR \left\{ t_{\alpha/2} \leq \frac{(\widehat{y}_0 - y_0)}{\widehat{std}(\widehat{y}_0 - y_0)} \leq t_{\alpha/2} \right\} = 1 - \alpha$$

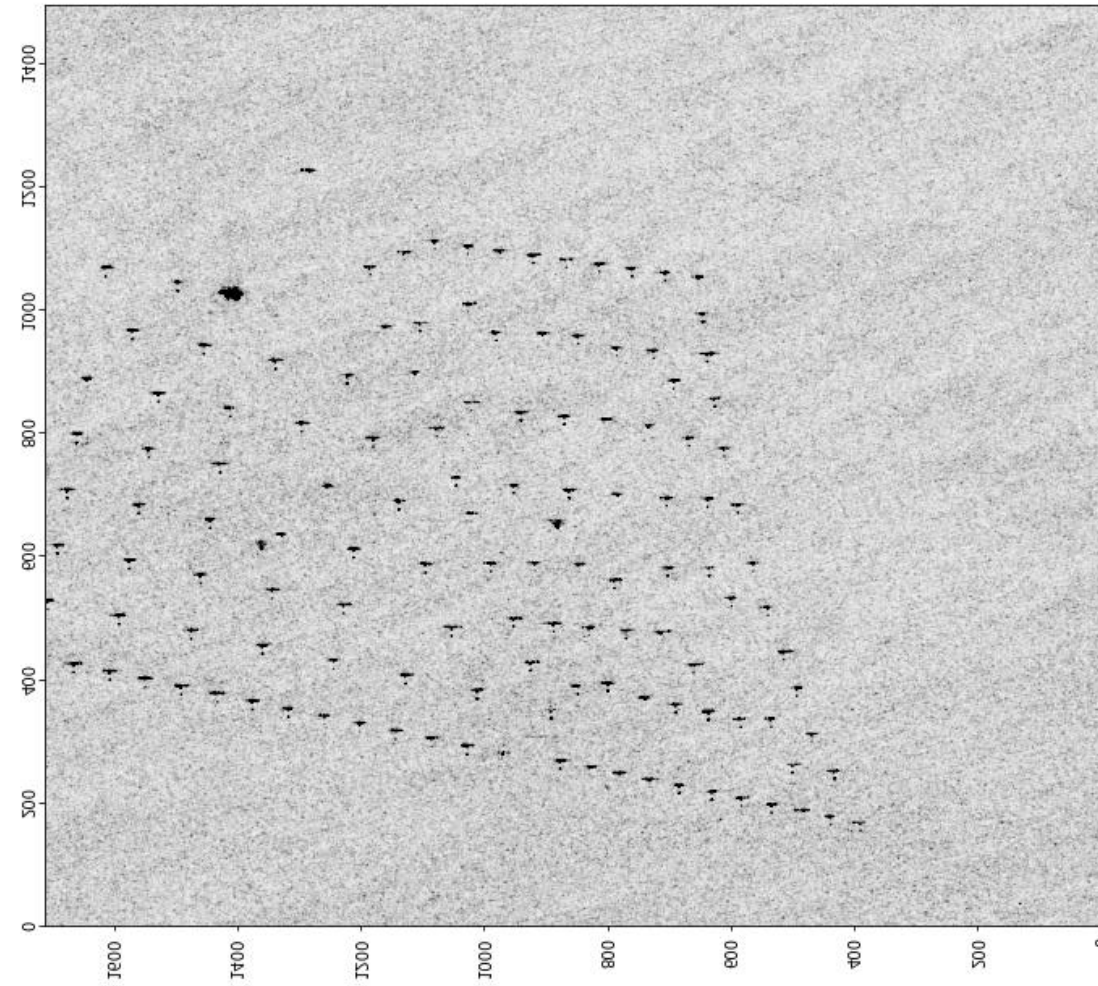
$$PR \left\{ \widehat{y}_0 - t_{\alpha/2} * \widehat{std}(\widehat{y}_0 - y_0) \leq y_0 \leq \widehat{y}_0 + t_{\alpha/2} * \widehat{std}(\widehat{y}_0 - y_0) \right\} = 1 - \alpha$$

Interfall des Schätzer a mit einer Wahrscheinlichkeit von $1-\alpha$: $\left[\widehat{y}_0 - t_{\alpha/2} * \widehat{std}(\widehat{y}_0 - y_0); +t_{\alpha/2} * \widehat{std}(\widehat{y}_0 - y_0) \right]$

1. Einfache lineare Regression
2. Case Study 1: Datenlücken bereinigen
3. Case Study 2: Winkelerkennung von WKAs
4. Multiple lineare Regression
5. Case Study 3: Multiple Regression Strompreis auf Brennstoffe und CO2

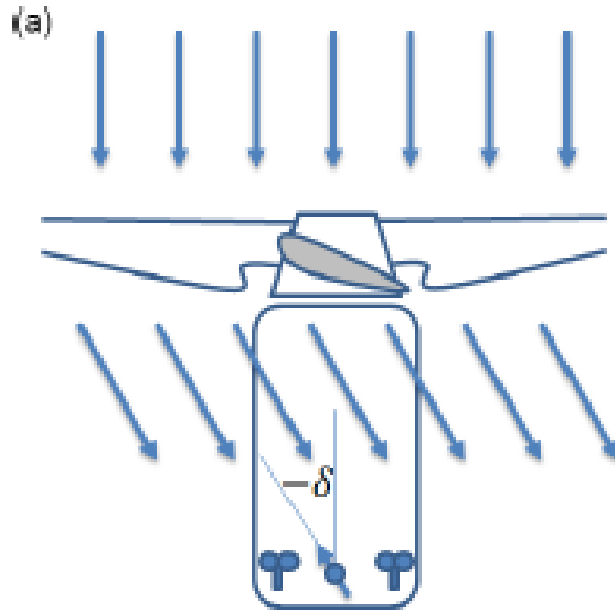


Was ist das?



Ausgangsfrage: Kann anhand von Satellitenaufnahmen eine Fehlstellung von Windkraftanlagen erkannt werden?

Windrichtung und gemessene Windrichtung

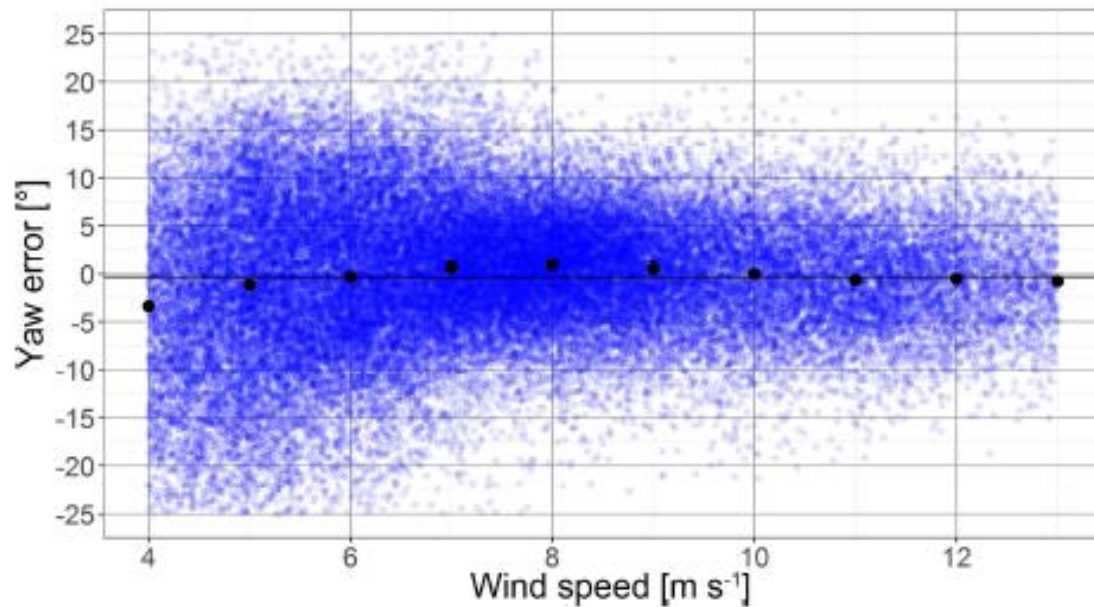


Ursachen der Fehlstellung

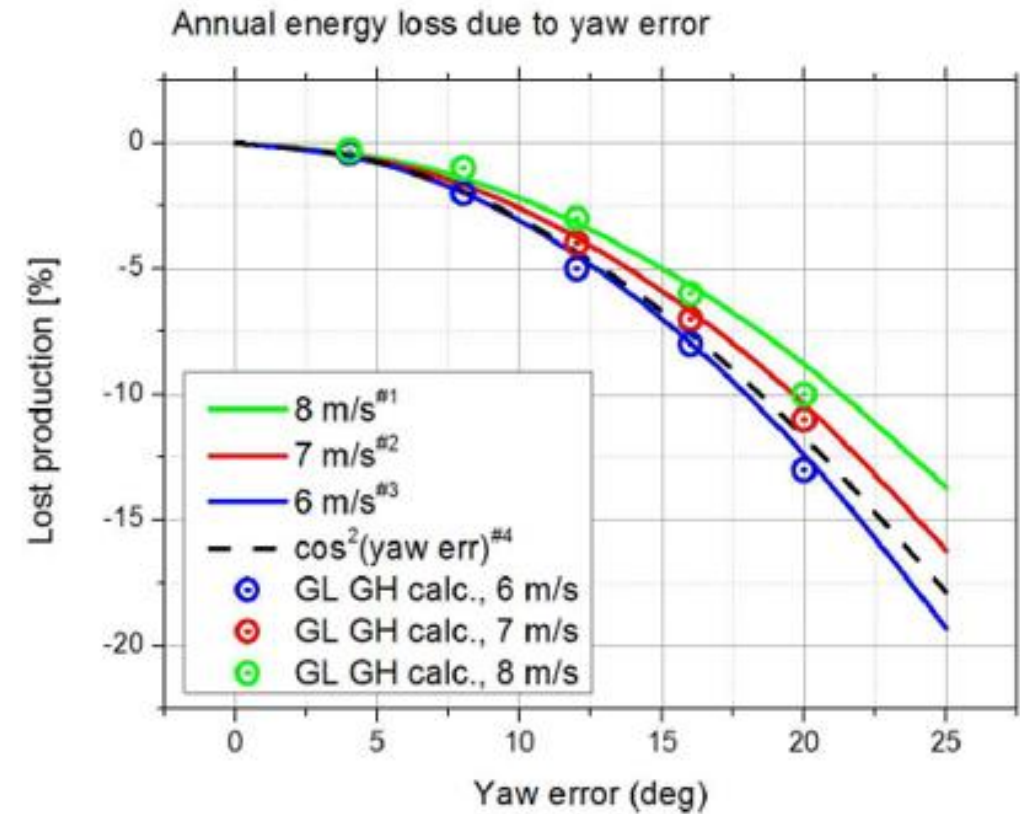
- Abweichung der Windrichtungsmessung:
Gemessene Richtung \neq Reale Richtung
- Abweichungen der Nachführmotoren/True-North-Kalibration:
Soll-Position \neq Ist-Position

Durch eine falsche Ausrichtung der Anlagen können Energieverluste eintreten

Ausrichtungsfehler in Abhängigkeit der Windgeschwindigkeit

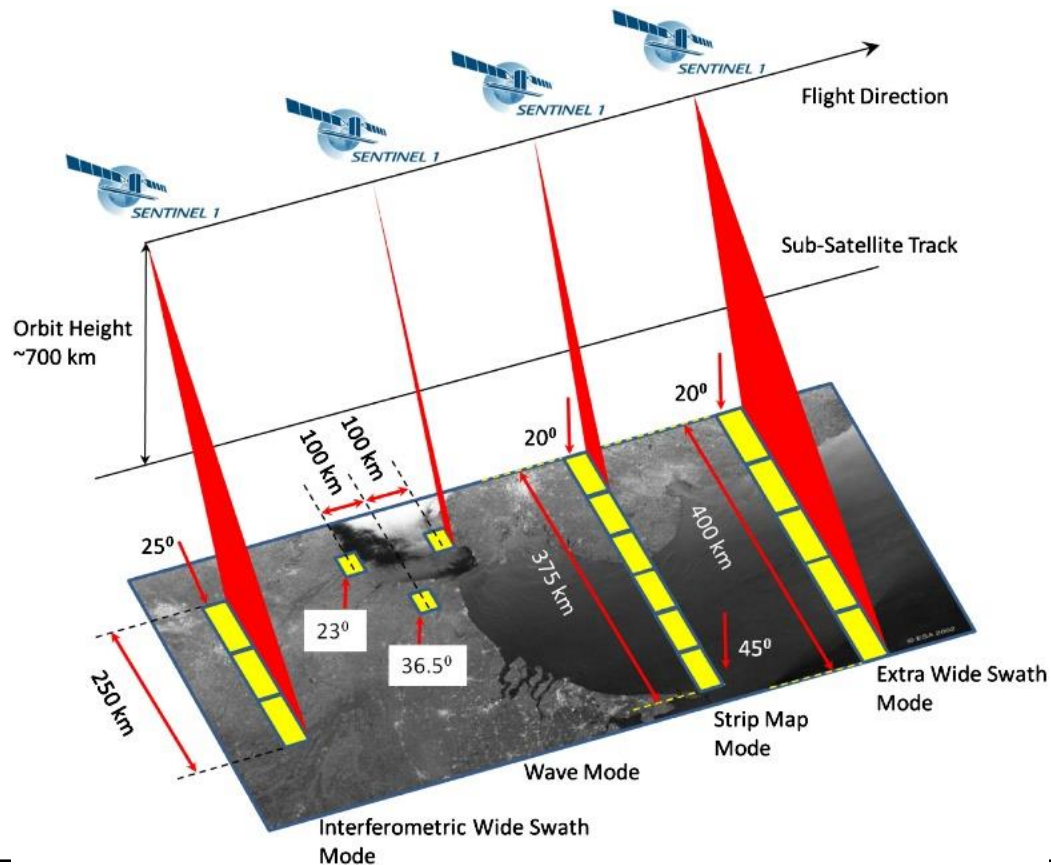


Energieverluste in Abhängigkeit des Ausrichtungsfehlers

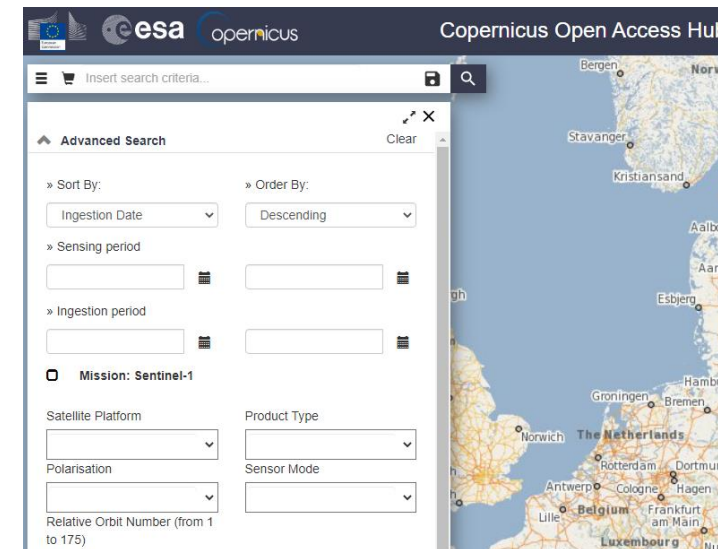


Mit Hilfe von Satellitenaufnahmen soll die tatsächliche Ausrichtung der Anlagen gemessen werden

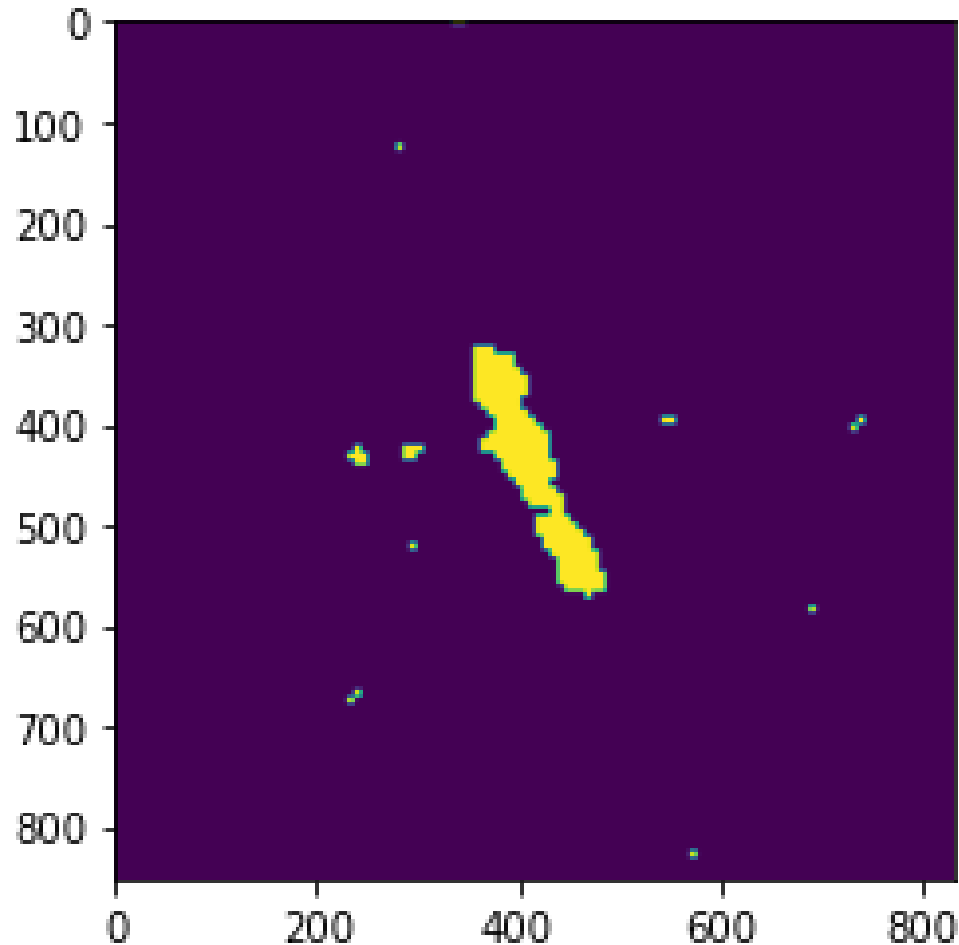
Abtastraster



- Momentan 2 Satelliten mit Radaraufnahmen
- 12 Tage polnaher Orbit
- Freier privater und kommerzieller Zugriff auf Daten



Ausgangspunkt der Analyse ist ein Datensatz, der in einem Raster die Auflösung der Datenpunkte in unterschiedlichen Farbtönen beinhaltet



- Bilder werden als Datenpunkte codiert
- Im dargestellten Bild sind ca. 850x850 Datenpunkte enthalten (Pixel).
- Jedes Pixel besitzt in Abhängigkeit der Farbgebung einen Wert.
- In der Aufbereitung des Bildes wird vereinfacht die Struktur der Windenergieanlage so referenziert, dass
 - Blaue Punkte den Wert Null besitzen
 - Gelbe Punkte (Struktur der WKA) den Wert 1 besitzen

Frage: Wie kann der Winkel der Ausrichtung mit Hilfe einer linearen Regression bestimmt werden?

Die Winkelmessung erfolgt anhand der Steigung einer Geraden durch die Datenpunkte

1. Bestimmung einer Geradengleichung durch die Datenpunkte
2. Normierung der Gerade auf Einheitsvektor
3. „Einzeichnen des Steigungsdreieck“
4. Verhältnis von Y_e und X_e entspricht der Steigung b der Geradengleichung
5. Nutzung des trigonometrischen Tangens zur Bestimmung des Winkels

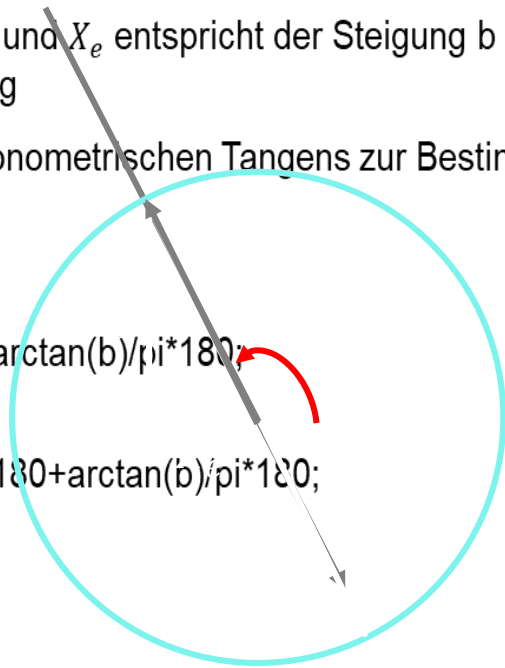
if $b > 0$

Winkel = $\arctan(b)/\pi * 180;$

else

Winkel = $180 + \arctan(b)/\pi * 180;$

end



1. Bestimmung einer Geradengleichung durch die Datenpunkte
2. Normierung der Gerade auf Einheitsvektor
3. „Einzeichnen des Steigungsdreieck“
4. Verhältnis von Y_e und X_e entspricht der Steigung b der Geradengleichung
5. Nutzung des trigonometrischen Tangens zur Bestimmung des Winkels

if $b > 0$

Winkel = $\arctan(b)/\pi * 180;$

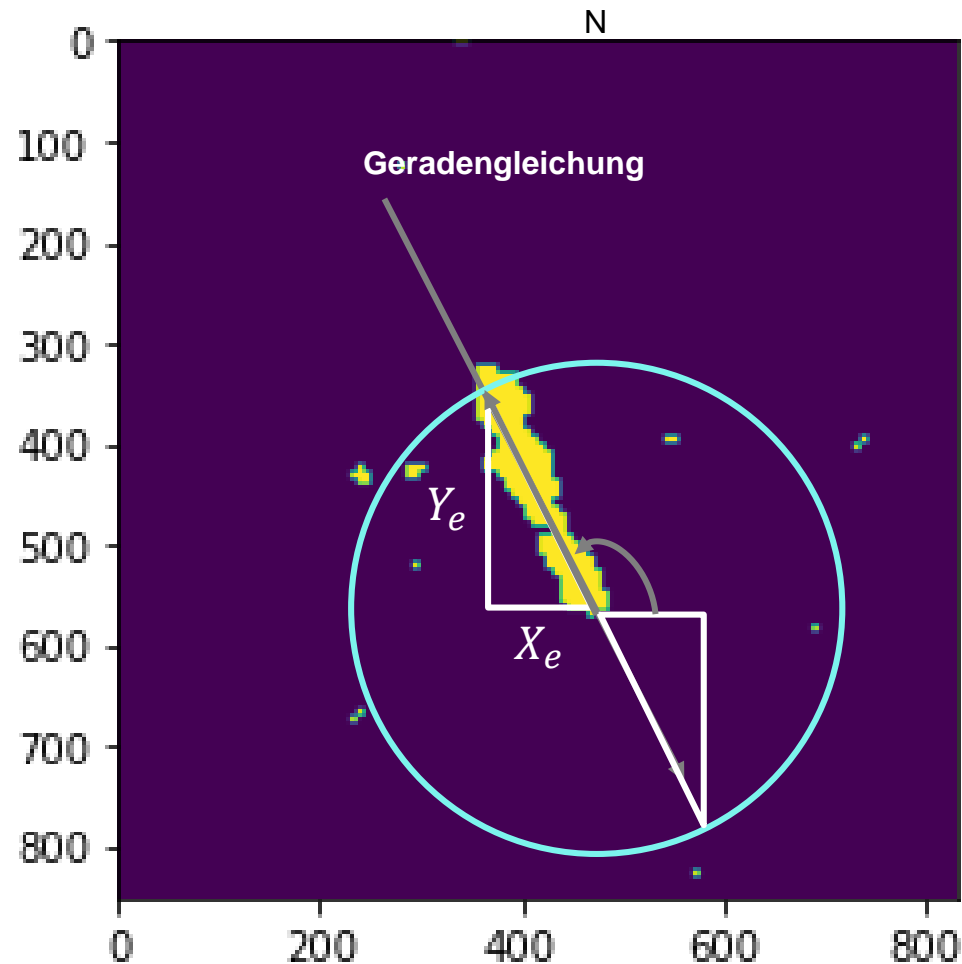
else

Winkel = $180 + \arctan(b)/\pi * 180;$

end

Wie kann dieses Problem in eine lineare Regression überführt werden?

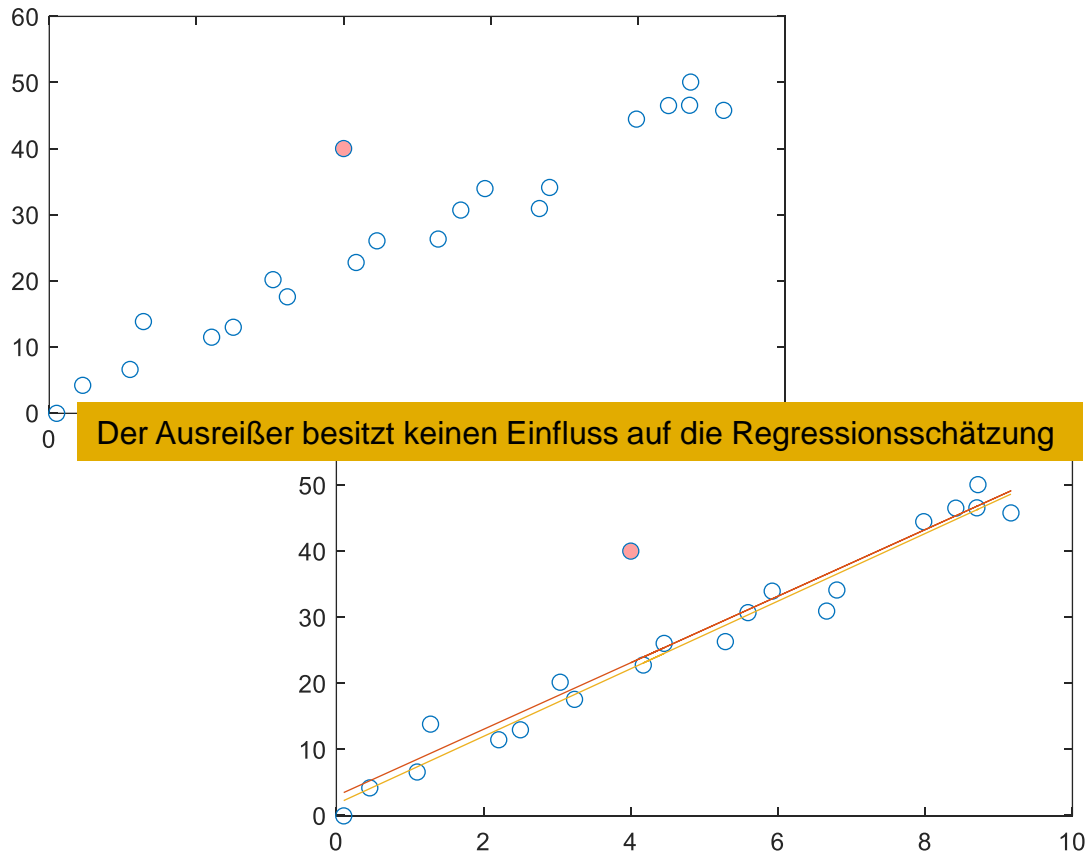
Für die Geradengleichungen muss die Matrix der Pixel in einen X und Y- Vektor zerlegt werden



- Das Bild besitzt 850 x 850 Datenpunkte
- Gelbe Datenpunkte sind mit 1 markiert; blaue mit Null
- Jeder Datenpunkt ist eindeutig durch Y und X Achse bestimmbar
- Flatten der Matrix in einen X und Y-Vektor
 - Berechnung des „neuen“ Y-Vektors
 - Je höher auf der Y-Achse ein „gelbes“ Pixel liegt, desto höher soll die „Y-Ausprägung“ sein
 - $Y_{\text{neu}} = \text{Wert}(Y,X) = \text{Pixel} * (850 - Y(X))$
 - $X_{\text{neu}} = X$

Ein Datenpunkt ist einflussreich, wenn er die Regressionsanalyse übermäßig beeinflusst, wie z. B. die geschätzten Werte \hat{y}_i , die geschätzten Steigungskoeffizienten oder die Ergebnisse des Hypothesentests

Ausreißer: y_i und \hat{y}_i liegen weit auseinander



Ausreißer:

Werte \hat{y}_i und y_i liegen weit auseinander

Leverage:

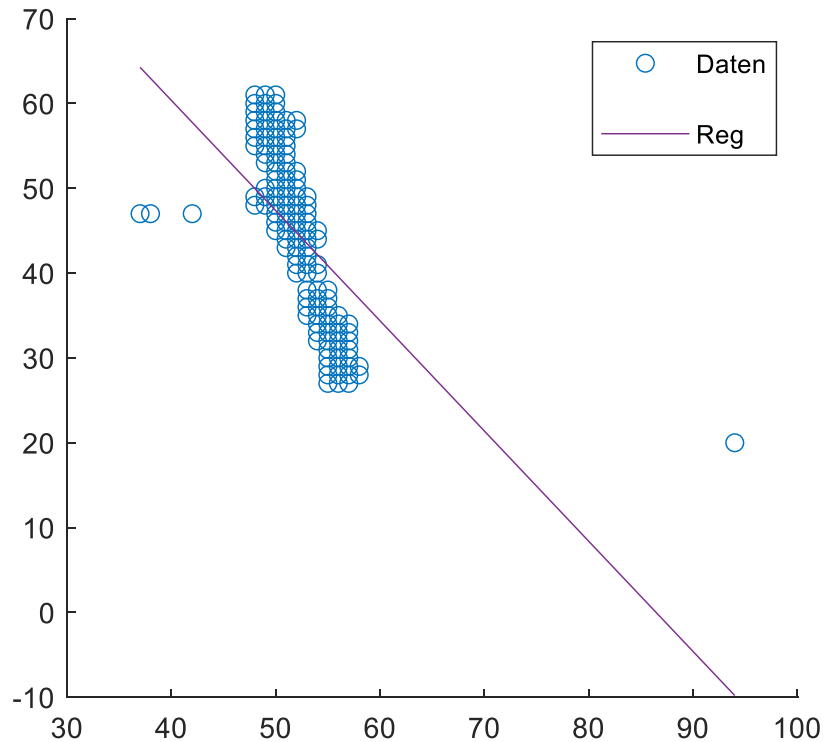
Eine Beobachtung mit ungewöhnlicher Kombination von Regressoren x_{i1}, x_{i2} besitzt ein hohes **leverage**.

Einfluss:

Ein Datenpunkt ist **einflussreich**, wenn er die Regressionsanalyse übermäßig beeinflusst, wie z. B. die geschätzten Werte \hat{y}_i , die geschätzten Steigungskoeffizienten b oder die Ergebnisse des Hypothesentests wie z.B. std oder p-Wert sowie das R^2 der Regression.

Leverage h_{ii} ist ein Distanzmaß zwischen x_i und \bar{x} und quantifiziert den Einfluss von y_i auf \hat{y}_i

Ausreißer



- Man kann zeigen, dass in Matrixschreibweise folgende Beziehung in einer Regression gilt:

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

- mit $H = X(X'X)^{-1}X'$ folgt:

$$\hat{y} = Hy$$

- Auflösung der Matrixschreibweise führt zu:

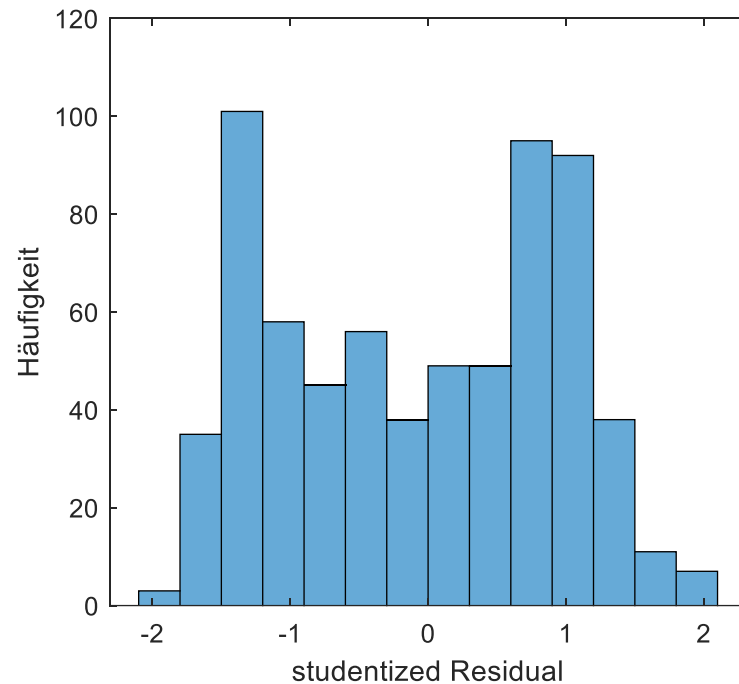
$$\hat{y}_i = h_{i1}y_1 + \dots + h_{ii}y_i + \dots + h_{in}y_n \quad i = 1, \dots, n$$

- Leverage h_{ii} quantifiziert den Einfluss von y_i auf \hat{y}_i .
 - wenn h_{ii} **klein** hat y_i einen geringen Einfluss auf die Schätzung von \hat{y}_i
- Eigenschaften von h_{ii}
 - h_{ii} ist eine Zahl zwischen 0 und 1
 - Die Summe h_{ii} entspricht der Anzahl der Parameter p

Literaturempfehlung: <https://online.stat.psu.edu/stat501/lesson/11/11.2>

Ausreißer können anhand des Residuums bzw. des standardisierten Residuums erkannt werden

Histogramm eines „studentized“ Residuums



■ Standardisiertes Residuum (studentized)

- Berücksichtigung des Leverage eines jeden Datenpunktes

$$r_i = \frac{e_i}{\sqrt{MSE \cdot (1 - h_{ii})}}$$

Individueller Fehler (e_i)
 Mittlerer Fehler (MSE)
 Gewichtung mit Leverage (h_{ii})

h_{ii} ist eine Zahl zwischen 0 und 1

- Daumenregel: ein studentized Wert größer **|3|** kann immer als Ausreißer angesehen werden

Mit Hilfe des Cook-Distanzmaß bzw. der „Difference in fit (DFFITS) können einflussreiche Datenpunkte erkannt werden

Cooks distance (D)

■ Vorgehen:

- Durchführung der Parameterschätzung

■ Metrik D:

$$D_i = \frac{(y_i - \hat{y}_i)^2}{p \cdot MSE} \left(\frac{h_{ii}}{(1 - h_{ii})^2} \right)$$

$p = \text{Anzahl der Parameter}$

Gewichtung mit Leverage

- Ein Datenpunkt ist einflussreich wenn gilt:

$$D > 1$$

Difference in Fit (DFFITS)

■ Vorgehen:

- Anpassung der Regressionsgerade jeweils ohne und mit den betrachteten Datenpunkt
- Vergleich der Ergebnisse mit und ohne den jeweiligen Datenpunkt

- Metrik DFFITS: (i) bedeutet Datenpunkt ist exkludiert:

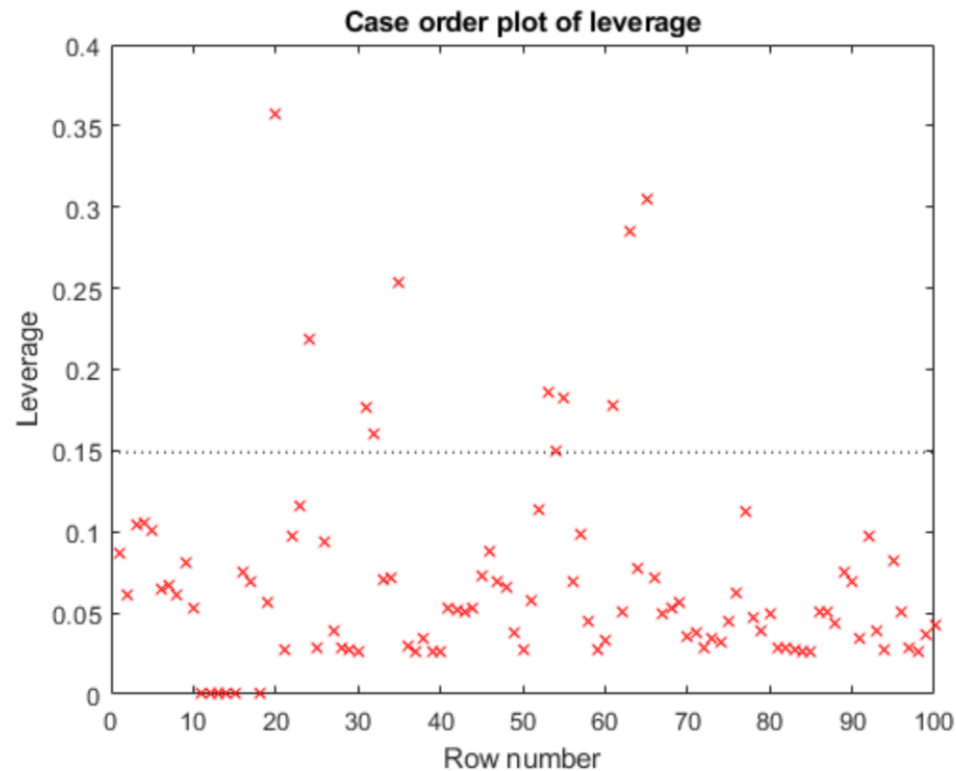
$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{MSE_{(i)} h_{ii}}}$$

- Ein Datenpunkt ist einflussreich wenn gilt:

$$abs(DFFITS) > 2 \sqrt{\frac{p + 1}{n - p - 1}}$$

Umsetzung in matlab: Ausreißeranalyse plotDiagnostics mdl)

Darstellung



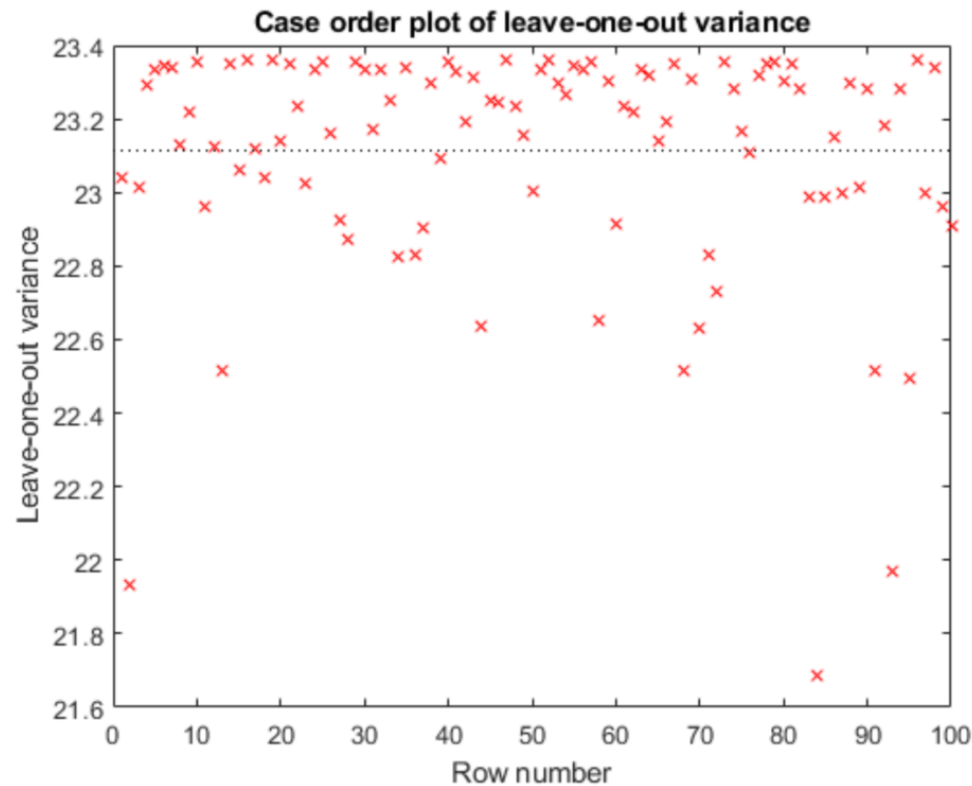
Einstellungsmöglichkeiten

■ plotype — Type of plot'dffits'

Value	Plot Type	Purpose
'contour'	Residual vs. leverage with overlaid contours of Cook's distance	Identify observations with large residual values, high leverage, and large Cook's distance values.
'cookd'	Cook's distance	Identify observations with large Cook's distance values.
'covratio'	Delete-1 ratio of determinant of covariance	Identify observations where the delete-1 statistic value is not in the range of the recommended thresholds.
'dfbetas'	Delete-1 scaled differences in coefficient estimates	Identify observations with large delete-1 statistic values.
'dffits'	Delete-1 scaled differences in fitted values	Identify observations with large delete-1 statistic values in an absolute value.
'leverage'	Leverage	Identify high leverage observations.
's2_i'	Delete-1 variance	Compare the delete-1 variance with the mean squared error

Umsetzung in matlab: Ausreißeranalyse plotDiagnostics mdl)

Darstellung



Erläuterung

- Purpose: The delete-1 variance ($S2_i$) shows how the mean squared error changes when an observation is removed from the data set. You can compare the $S2_i$ values with the value of the mean squared error.
- Definition: $S2_i$ is a set of residual variance estimates obtained by deleting each observation in turn. The $S2_i$ value for observation i is

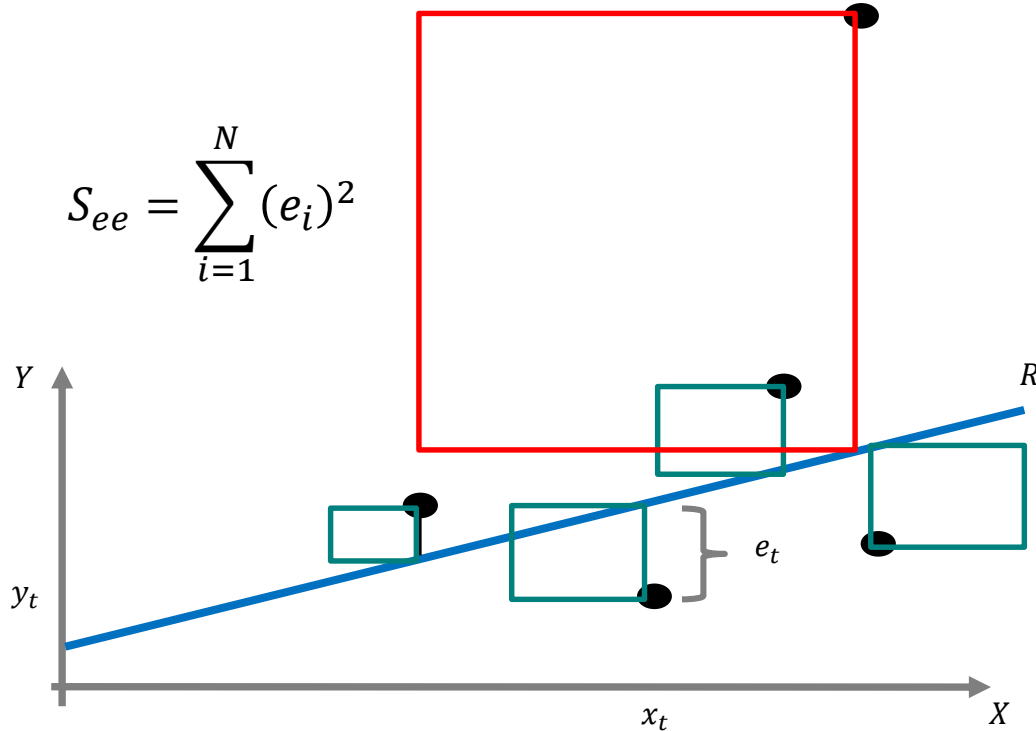
$$S2_i = MSE_{(i)} = \frac{\sum_{j \neq i}^n [y_j - \hat{y}_{j(i)}]^2}{n - p - 1},$$

Einfluss des Ausreißers auf die Zielfunktion kann durch eine Gewichtung verringert werden

Zielfunktion

KQ- Methode für den Punktschätzer eines linearen Regressionsmodells

$$S_{ee} = \sum_{i=1}^N (e_i)^2$$



Mögliche Maßnahmen

- Umstellung der Zielfunktion auf L1-Norm

$$S_{ee} = \sum_{i=1}^N \text{abs}(e_i)$$

- Einfügung von Gewichten

$$S_{ee} = \sum_{i=1}^N (y_i - \mathbf{w}_{ii} b X_i + a)^2 \quad \mathbf{w} = \begin{bmatrix} \frac{1}{\sigma_1^2} & \\ & \frac{1}{\sigma_n^2} \end{bmatrix}$$

- Robuste Regressionstechniken:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N e_i^2 \quad \rightarrow \quad \min_{\theta} \frac{1}{N} \sum_{i=1}^N \rho(r_i)$$

Die „robuste lineare Regression“ behandelt Ausreißer entsprechend einer Gewichtungsfunktion, um deren Einfluss zu begrenzen

Aufbau der robusten Regression

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \epsilon_i^2 \quad \rightarrow \quad \min_{\theta} \frac{1}{N} \sum_{i=1}^N \rho(r_i)$$

Gewichtungsfunktion für iterativen Durchlauf:

$$w(r) = (\underbrace{abs(r) < 1}_{\substack{\text{r größer } |1| \text{ wird} \\ \text{nicht berücksichtigt}}}) * (\underbrace{1 - r^2}_{\substack{\text{r nahe Null} \\ \text{besitzt hohes} \\ \text{Gewicht}}})^2$$

Transformierter Fehlerterm:

$$r = \frac{e}{\tau \cdot \frac{MAD}{0.6745} \cdot \sqrt{1 - h}}$$

Fehler der Regression

Parameter

Erläuterung

- Die gewöhnliche KQ Schätzung ist nur unter den getroffenen Annahmen gültig: insb. $\epsilon \sim N(0, \sigma^2)$
- Robuste Regressionsmethoden bieten eine Alternative zur KQ-Regression

Hinweise:

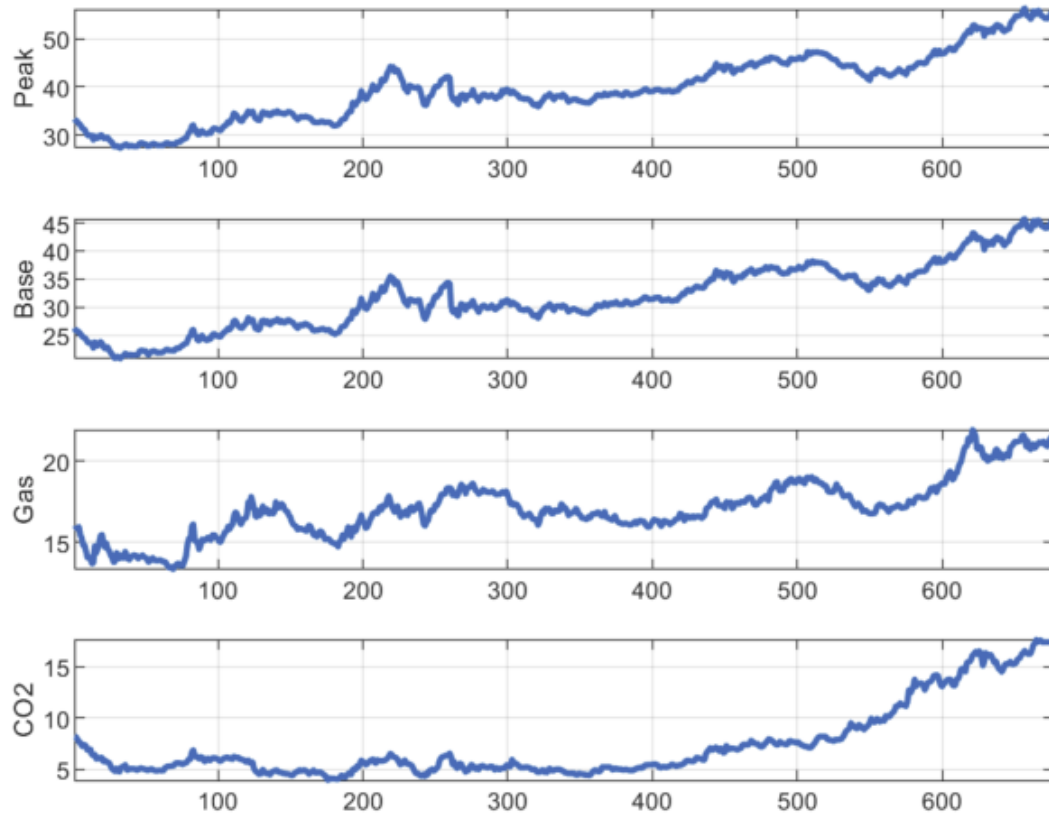
- MAD = absolute Abweichung vom Median der Residuen
- Die Konstante in Höhe von 0.6745 macht den Schätzer unbiased zur Normalverteilung
- h = Leverage Vektor

1. Einfache lineare Regression
2. Case Study 1: Datenlücken bereinigen
3. Case Study 2: Winkelerkennung von WKAs
4. Multiple lineare Regression
5. Case Study 3: Multiple Regression Fernwärmepreise

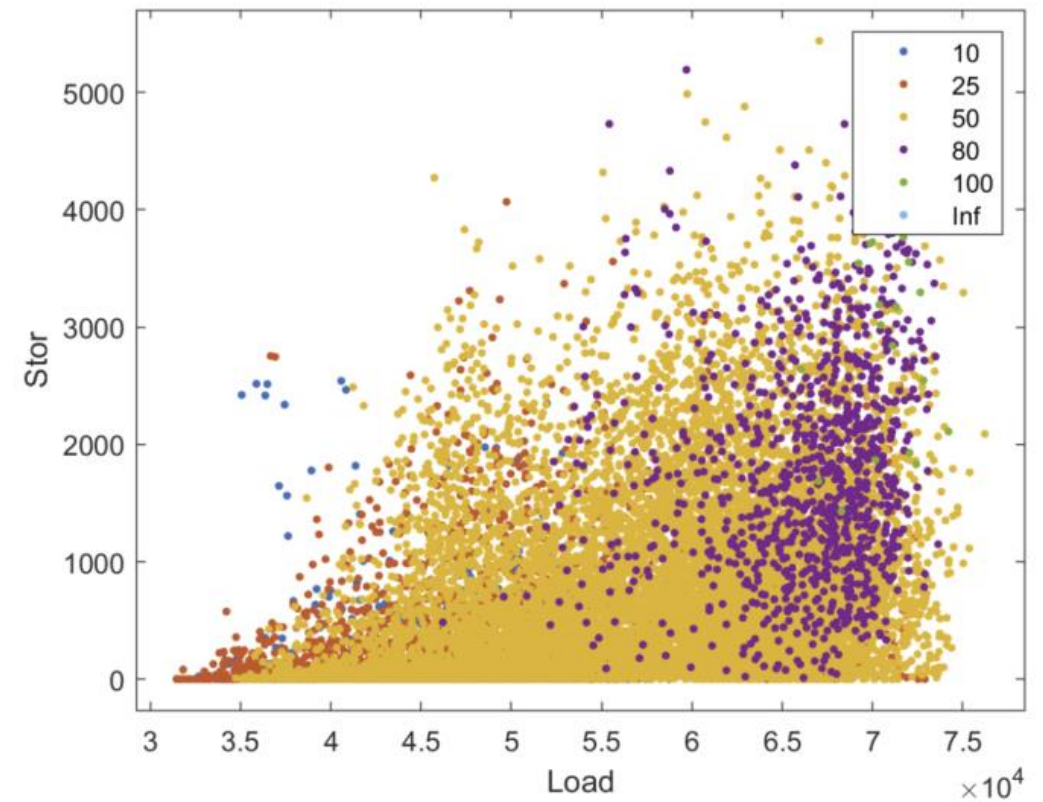


Es gibt eine Reihe von graphischen Visualisierungen, um mehrdimensionale Daten zu analysieren I

Scatter-Histogramm

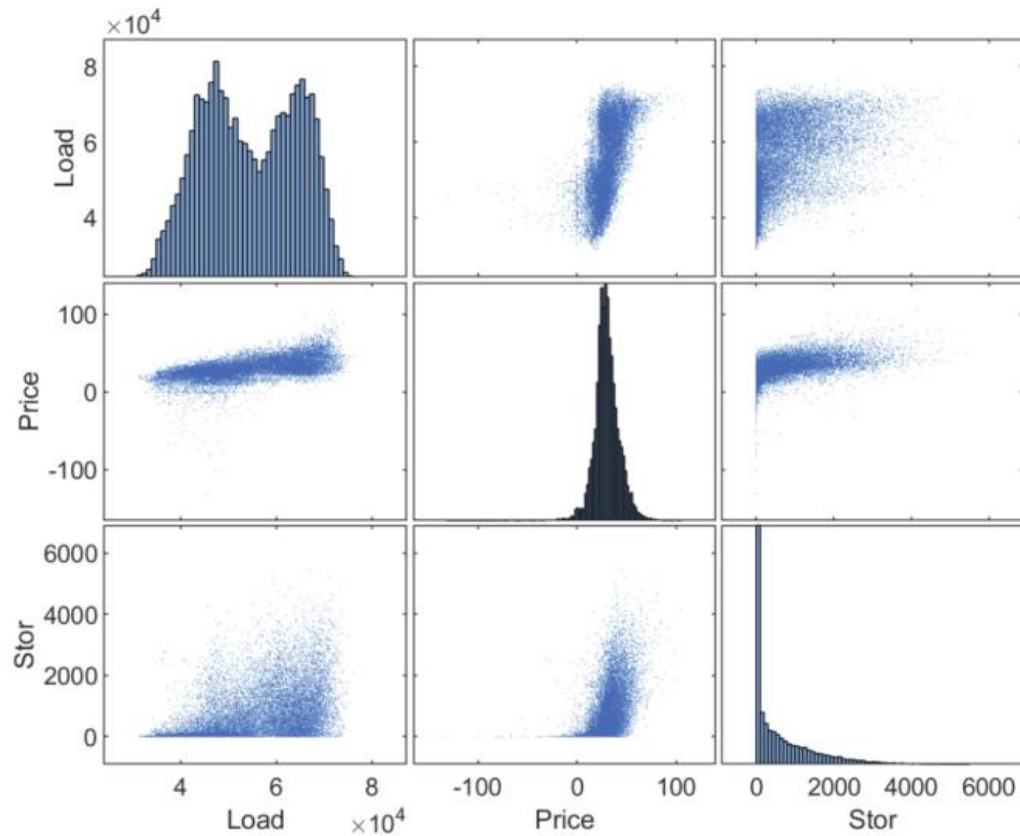


Labeling der Daten

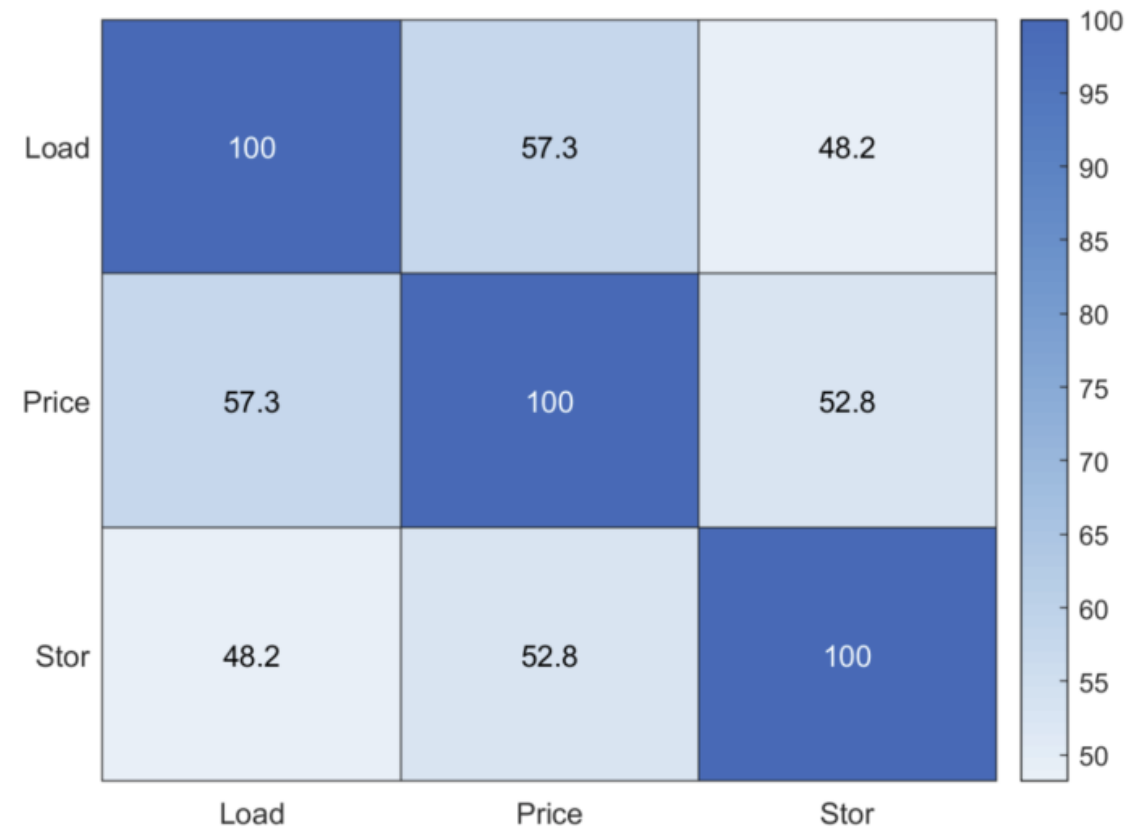


Es gibt eine Reihe von graphischen Visualisierungen, um mehrdimensionale Daten zu analysieren II

Scatter-Histogramm

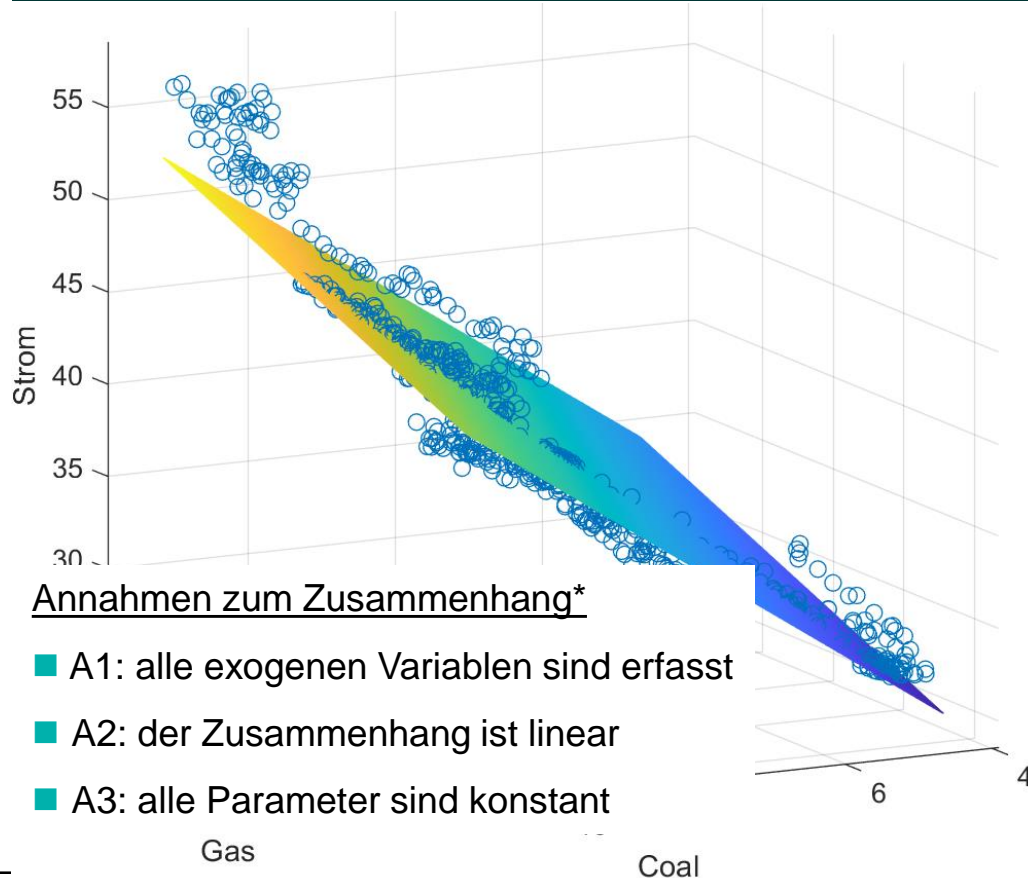


Plotmatrix



y soll folgend auf den funktionalen Zusammenhang von x_1 und x_2 bestimmt werden

Darstellung des Zusammenhangs



Aufbau des Modells

- Aufbau des ökonometrischen Modells:

$$y_t = a + b_1 * x_{1t} + b_2 * x_{2t} + e_t$$

- Resultat des geschätzten Modells

$$\hat{y}_t = \hat{a} + \hat{b}_1 * x_{1t} + \hat{b}_2 * x_{2t}$$

- Formulierung der Residuen der Schätzung

$$e_t = y_t - \hat{y}_t$$

- Daraus folgt folgender Zusammenhang:

$$y_t = \hat{a} + \hat{b}_1 * x_{1t} + \hat{b}_2 * x_{2t} + e_t$$

* Für die Störgrößen gelten die gleichen Annahmen wie im einfachen linearen Fall. Störgröße e hat Erwartungswert von Null; Varianz Störgröße ist konstant; Störgröße weist keine Autokorrelation auf; Störgröße ist normalverteilt

Punktschätzer für \hat{a} , \hat{b}_1 \hat{b}_2

Wiederum kann die KQ-Methode zur Ableitung der Schätzung der Parameter verwendet werden

- Minimierung der Summe der quadrierten Residuen

$$S_{ee} = \sum e_t^2$$

- Aufgelöst nach \hat{e} folgt:

$$e_t = y_t - \hat{a} - \hat{b}_1 \cdot x_{1t} - \hat{b}_2 \cdot x_{2t}$$

- Eingesetzt in die Minimierungsfunktion

$$S_{ee} = \sum (y_t - \hat{a} - \hat{b}_1 \cdot x_{1t} - \hat{b}_2 \cdot x_{2t})^2$$

- Zur Minimierung erfolgt die Ableitung dieser Funktion und deren Nullsetzung

$$\partial S_{ee} / \partial \hat{a} = 0, \quad \partial S_{ee} / \partial \hat{b}_1 = 0, \quad \partial S_{ee} / \partial \hat{b}_2 = 0$$

Punktschätzer der Parameter

- Im Nenner zur Bestimmung von b erscheinen nur exogene Variablen
- Im Zähler sind jeweils auch Kovarianzen zwischen exogener und endogener Variable abgebildet

$$\hat{a} = \bar{y} - \hat{b}_1 \bar{x}_1 - \hat{b}_2 \bar{x}_2$$

$$\hat{b}_1 = \frac{S_{22}S_{1y} - S_{12}S_{2y}}{S_{11}S_{22} - S_{12}^2}$$

$$\hat{b}_2 = \frac{S_{11}S_{2y} - S_{12}S_{1y}}{S_{11}S_{22} - S_{12}^2}$$

Definition des Bestimmtheitsmaß

- Die Variation der zu erklärenden Variable ist definiert:

$$S_{yy} = \sum (y_t - \bar{y})^2$$

- Die erklärende Variation bildet sich wie folgt:

$$S_{\hat{y}\hat{y}} = \sum (\hat{y}_t - \bar{y})^2, S_{yy} = S_{\hat{y}\hat{y}} + S_{ee}$$

- Daraus folgt für das Bestimmtheitsmaß:

$$R^2 = \frac{S_{yy} - S_{ee}}{S_{yy}} = \frac{S_{\hat{y}\hat{y}}}{S_{yy}} \text{ mit } S_{\hat{y}\hat{y}} = \sum_{k=1}^p \hat{b}_k * S_{ky}$$

- Das Bestimmtheitsmaß wird beeinflusst durch die Kovarianzen, gewichtet mit dem geschätzten Parameter b.

- Definition der Korrelation:

$$R_{12}^2 = \frac{S_{12}^2}{S_{11}S_{22}}$$

Weitergehende Fragestellung: Unverzerrtheit und Effizienz der KQ-Methode

- Bildung des Erwartungswertes -> Schätzer ist Erwartungstreu

$$E(\widehat{b}_1) = b_1; E(\widehat{b}_2) = b_2; E(\widehat{a}) = a$$

- Bildung der Varianzen:

$$\text{var}(\widehat{b}_1) = \frac{\sigma^2}{S_{11}(1 - R_{12}^2)}$$

$$\text{var}(\widehat{b}_2) = \frac{\sigma^2}{S_{22}(1 - R_{12}^2)}$$

$$\text{var}(\widehat{a}) = \frac{\sigma^2}{N} + \overline{x_1^2} \text{var}(\widehat{b}_1) + \overline{x_2^2} \text{var}(\widehat{b}_2) + 2 \overline{x_1 x_2} \text{var}(\widehat{b}_1) \text{var}(\widehat{b}_2)$$

- Kovarianz:

$$\text{cov}(\widehat{b}_2, \widehat{b}_1) = \frac{-\sigma^2 R_{12}^2}{S_{12}(1 - R_{12}^2)} \text{ mit } R_{12}^2 = \frac{S_{12}^2}{S_{11}S_{22}}$$

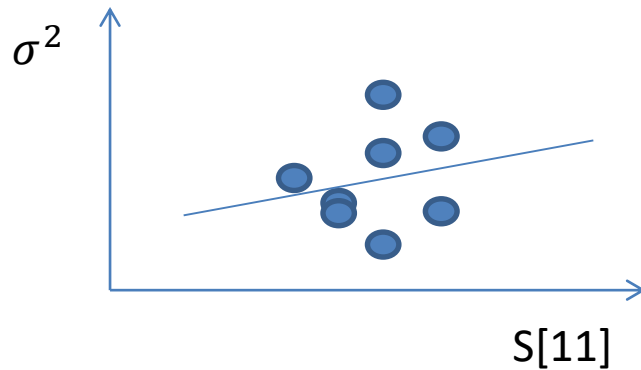
Interpretation der Formeln

$$\text{var}(\widehat{b_1}) = \frac{\sigma^2}{S_{11}(1 - R_{12}^2)} = \frac{\sigma^2}{S_{[11]}}$$

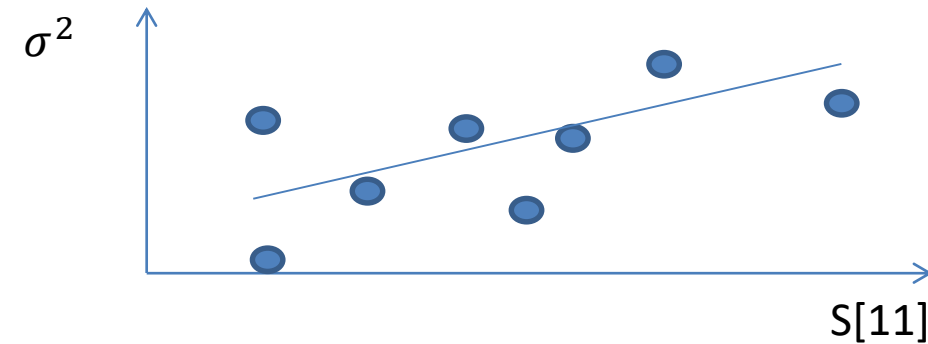
$S_{[11]} = \text{autonome Variation}$

- Fall A: Varianz des Schätzers ist hoch
- Fall B: Varianz des Schätzers ist niedrig

Fall A: σ^2 ist im Verhältnis zu $S_{[11]}$ groß



Fall B: σ^2 ist im Verhältnis zu $S_{[11]}$ klein



Schätzformeln für die Varianzen und Kovarianzen der Punktschätzer

- Gegenüber der vorangegangenen Folie wird nun davon ausgegangen, dass die Varianz der Störgröße unbekannt und damit geschätzt werden muss

$$\widehat{\sigma^2} = \frac{S_{ee}}{N - p}$$

- Mit Hilfe des Schätzers für die Varianz der Störterme können wiederum die Varianz-Kovarianz der Punktschätzer bestimmt werden

Wahrscheinlichkeitsverteilung der Schätzer

- Die Schätzer sind normalverteilt:

$$\hat{b}_1 \sim N\left(b_1, \frac{\sigma^2}{s_{11}(1-R_{12}^2)}\right)$$

- Intervallschätzer

- Schritt 1 Schätzung von σ und $\text{std}(\hat{b}_1)$

$$\widehat{\text{std}}(\hat{b}_1) = \sqrt{\frac{\hat{\sigma}^2}{s_{11}(1-R_{12}^2)}}$$

Standardisierung von \hat{b}_1 lautet:

$$t = \frac{\hat{b}_1 - E(\hat{b}_1)}{\widehat{\text{std}}(\hat{b}_1)}$$

$$t \sim t(N-3), \text{ da } p=3 \text{ ist}$$

- Ermittlung eines $t_{\alpha/2}$ – Wertes und Bildung des Intervalls und Ersetzen von t

$$\Pr\left[t_{\alpha/2} \leq t \leq t_{\alpha/2}\right] = 1 - \alpha$$

$$\left[\hat{b}_1 - t_{\alpha/2} * \widehat{\text{std}}(\hat{b}_1); \hat{b}_1 + t_{\alpha/2} * \widehat{\text{std}}(\hat{b}_1)\right]$$

Verallgemeinerung in Matrixschreibweise (Hinweis: <https://online.stat.psu.edu/stat501/lesson/5/5.4>)

- Aufstellung des Regressionsmodells:

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_y = \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}}_x \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}}_b + \underbrace{\begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}}_e$$

$$y = Xb + e$$

Ableitung der Punktschätzer*:

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

Berechnung R^2

$$R^2 = \frac{[y'X(X'X)^{-1}X'y - N\bar{y}^2]}{y'y - N\bar{y}^2}$$

Ableitung der Varianz Kovarianz Matrix:

$$V(\hat{b}) = \sigma^2(X'X)^{-1}, \hat{\sigma}^2 = \frac{S_{ee}}{N-p} \quad p=\text{Anzahl der Parameter}$$

Ableitung der Intervallschätzer:

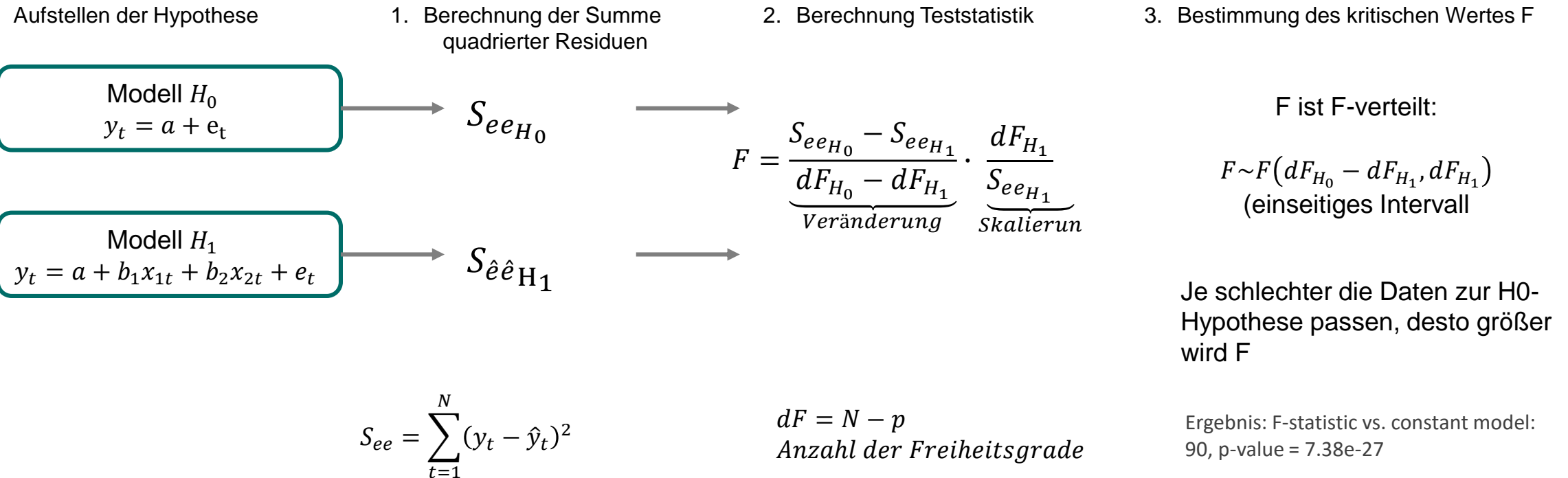
$$\left[s'\hat{b} - t_{\alpha/2} * \hat{\sigma}\sqrt{s'(X'X)^{-1}s}; s'\hat{b} + t_{\alpha/2} * \hat{\sigma}\sqrt{s'(X'X)^{-1}s} \right]$$

* 1) $X'y = X'Xb \quad | \cdot X'$;

2) $(X'X)^{-1}X'y = (X'X)^{-1}X'Xb \quad | \cdot (X'X)^{-1} \quad | (X'X)^{-1}X'X = I$

Der F-Test wird für den Fall angewendet, wenn mehrere Linearkombinationen gleichzeitig getestet werden sollen

- Vorgehensweise: Gegenüberstellung der SSE zweier Modelle
- H0-Hypothese: : *Parameter die nicht im einfacheren Modell enthalten sind, sind Null*



1. Einfache lineare Regression
2. Case Study 1: Datenlücken bereinigen
3. Case Study 2: Winkelerkennung von WKAs
4. Multiple lineare Regression
5. Case Study 3: Multiple Regression Strompreis auf Brennstoffe und CO₂



Aufgabe: Analyse der Terminmarktpreise

- Teil 1: Erstellen Sie ein neues jupyter-notebook
 - Importieren Sie die Tabelle „Terminpreise
 - Erstellen Sie geeignete Plots zur Analyse der Zusammenhänge
 - Interpretieren Sie die Plots und leiten geeignete Modellhypothesen für den Strompreis ab

- Teil 2:
 - Erstellen Sie eigene Modellzusammenstellungen auf Basis der Modellhypothesen
 - Analyse und Interpretation der Ergebnisse
 - Parameter, t-Werte, p-Werte, Standardabweichung (Se)
 - Fit: RMSE
 - Ausgabe der Darstellung im Zeitablauf