universität innsbruck

Stochastische Regressionsanalyse: Intro

Grundlagen der Ökonometrie

Stochastische Regressionsanalyse: Intro

Was bisher geschah:

- Wir uns bisher für existierende Daten interessiert, v.a. Kennzahlen berechnet, z.B. *bedingte Mittelwerte* (OLS). ⇒ *Informationsverdichtung*

Was kommt jetzt?

- Wir interessieren uns nicht länger dafür, was passiert ist (Beobachtungen), sondern *was passieren könnte!*
- Eine einfache Möglichkeit sich dies vorzustellen:
  - wir beobachten nur eine Stichprobe aus einer unbekannten Grundgesamtheit,
  - aber wir interessieren uns für die Grundgesamtheit!
- Wiederum versuchen wir, die unbekannte Grundgesamtheit durch Kennzahlen zu beschreiben, z.B. durch *bedingte Erwartungswerte*.
- Wir benötigen zusätzliche Kennzahlen um die Unsicherheit zu charakterisieren, z.B. *Standardfehler*.

Stochastische Regressionsanalyse: Intro

Vermuteter Zusammenhang in unbekannter Grundgesamtheit:

'Population Regression Function' PRF

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

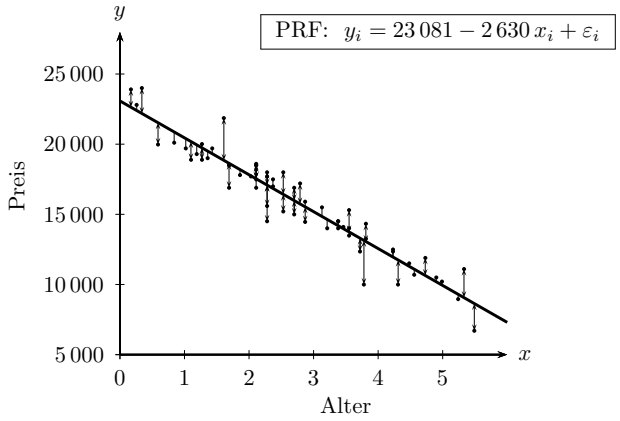
Autobeispiel (Grundgesamtheit, PRF):

$$\text{PRF:} \qquad \text{Preis}_i = 23\,081 - 2\,630\,\text{Alter}_i + \varepsilon_i$$

PRF ist unbeobachtbar! (Forscher kann nur eine *Stichprobe* beobachten!)

Stochastische Regressionsanalyse: Intro

Wenn sie beobachtbar wäre . . .



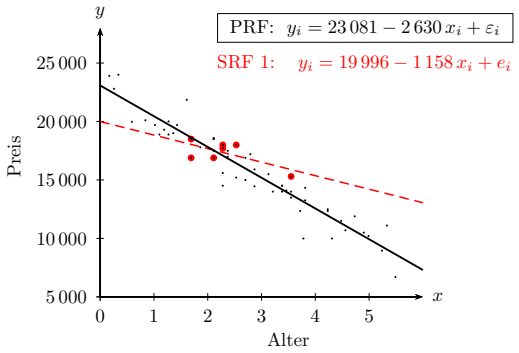
Stochastische Regressionsanalyse: Intro

Ein Forscher kann nur eine Stichprobe mit $n = 7$ beobachten:

Stichprobe zu SRF 1

Obs.	Preis	Alter
3	18000	2.28
7	18000	2.53
14	17700	2.28
21	16900	1.69
31	15300	3.55
56	16900	2.11
59	18500	1.69

'Sample Regression Function' **SRF**



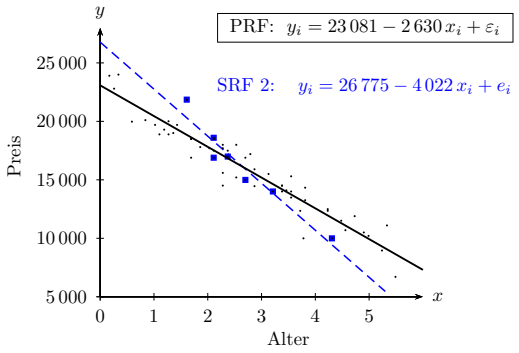
Stochastische Regressionsanalyse: Intro

Ein zweite Forscherin beobachtet andere Stichprobe ($n = 7$):

Stichprobe zu SRF 2

Obs.	Preis	Alter
1	16990	2.37
16	15000	2.70
25	10000	4.31
29	21850	1.61
35	18600	2.11
56	16900	2.11
60	14000	3.21

'Sample Regression Function 2' **SRF 2**



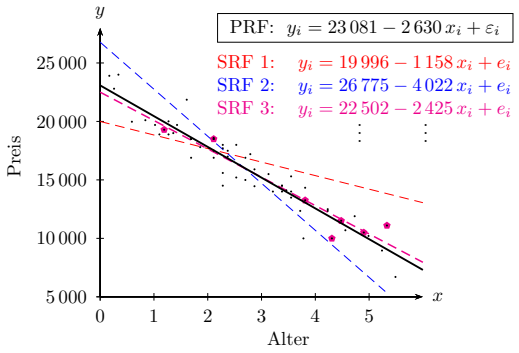
Stochastische Regressionsanalyse: Intro

Ein dritter Forscher zieht nochmals eine andere Stichprobe ($n = 7$):

Stichprobe zu SRF 2

Obs.	Preis	Alter
11	11100	5.33
25	10000	4.31
30	13250	3.81
40	18500	2.11
47	11500	4.48
50	10500	4.90
51	19300	1.19

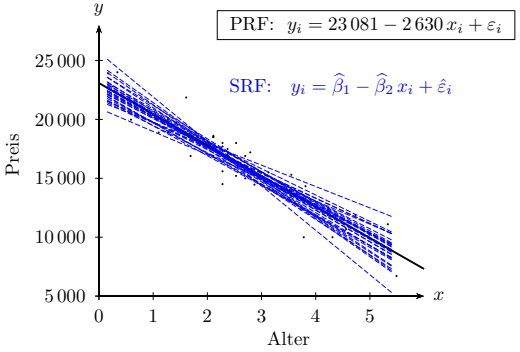
'Sample Regression Function 3' **SRF 3**



Stochastische Regressionsanalyse: Intro

Monte Carlo Simulation: Computer zieht z.B. 1000 Stichproben und berechnet 1000 SRF's:

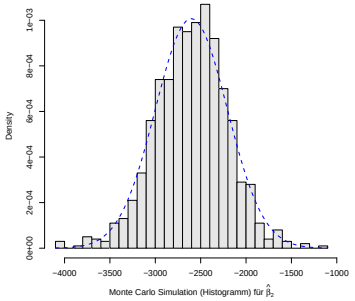
Stichprobe	b_1	b_2
1	22476	-2513
2	23525	-2658
3	22502	-2425
⋮	⋮	⋮
33	19996	-1158
⋮	⋮	⋮
369	26775	-4022
⋮	⋮	⋮
999	23327	-2714
1000	23598	-2875
Mittelwert:	22976	-2603



Stochastische Regressionsanalyse: Intro

Monte Carlo Simulation: Histogramm für Steigungskoeffizienten b_2 der SRF's:

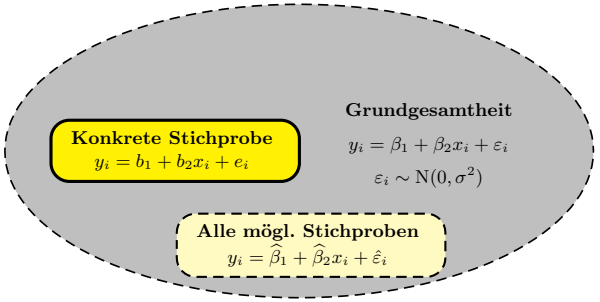
Stichprobe	b_1	b_2
1	22476	-2513
2	23525	-2658
3	22502	-2425
⋮	⋮	⋮
33	19996	-1158
⋮	⋮	⋮
369	26775	-4022
⋮	⋮	⋮
999	23327	-2714
1000	23598	-2875
Mittelwert:	22976	-2603



$\Rightarrow$ **1) Gesetz der Großen Zahl** **2) Zentraler Grenzwertsatz**

Stochastische Regressionsanalyse: Intro

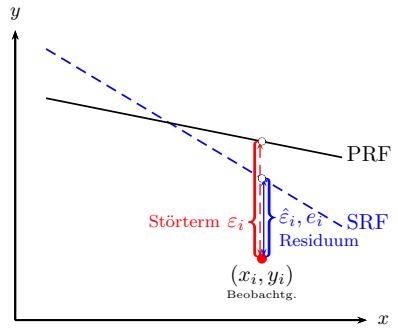
Notation:



β . . . Parameter; $\widehat{\beta}$ . . . Zufallsvariable; b . . . Realisation

Notation

Störterme und Residuen:



Achtung: 'Residuum' wird für Zufallsvariable und Realisation verwendet!

Notation

- **DGP (Grundgesamtheit):**
  - $\beta_1, \ldots, \beta_k$: unbeobachtbare Parameter der PRF, deterministisch!
  - ε_i: unbeobachtbare Störterme (*errors*), stochastisch!
- **Stichprobe:**
  - *Zufallsvariablen:*
    - $\widehat{\beta}_1, \ldots, \widehat{\beta}_k$: Zufallsvariablen (stochastisch!), Ergebnisse für jede mögliche Stichprobe
    - $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$: Residuen, Ergebnisse für jede mögliche Stichprobe
  - *Realisationen:*
    - $b_1, \ldots, b_k$: d.h. Ergebnis für die *eine* beobachtete Stichprobe, deterministisch!
    - $e_i, \ldots, e_n$): Realisationen der Residuen, deterministisch!

Achtung: In der Literatur wird für die Zufallsvariablen $\widehat{\beta}_h$ und für die deterministischen Realisationen b_h häufig das selbe Symbol $\widehat{\beta}_h$ verwendet!

Stochastische Regressionsanalyse: Intro

OLS: die Mathematik funktioniert gleich

$$\min_{\widehat{\beta}_1,\widehat{\beta}_2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \min_{\widehat{\beta}_1,\widehat{\beta}_2} \sum_{i=1}^{n} (y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i)^2$$

FOC:

$$\frac{\partial \sum_i \hat{\varepsilon}_i^2}{\partial \widehat{\beta}_1} = 2 \sum_i \underbrace{\left(y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i\right)}_{\hat{\varepsilon}_i} (-1) = 0 \quad \Rightarrow \quad \sum_i \hat{\varepsilon}_i = 0$$

$$\frac{\partial \sum_i \hat{\varepsilon}_i^2}{\partial \widehat{\beta}_2} = 2 \sum_i \underbrace{\left(y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i\right)}_{\hat{\varepsilon}_i} (-x_i) = 0 \quad \Rightarrow \quad \sum_i x_i \hat{\varepsilon}_i = 0$$

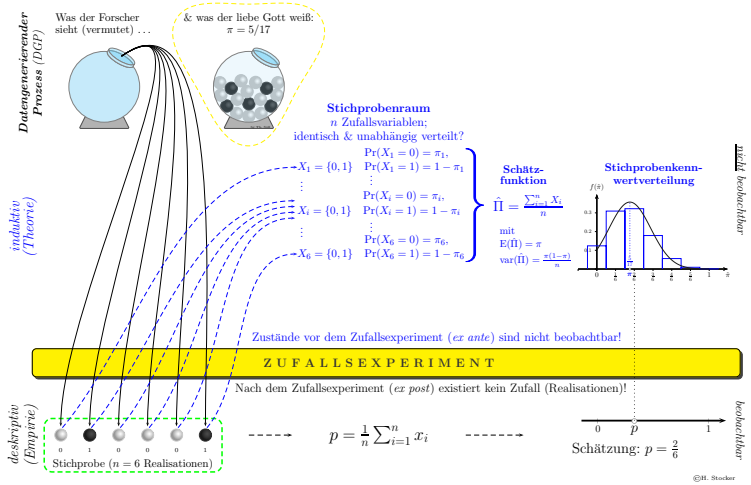
$$\widehat{\beta}_2 = \frac{\widehat{\mathrm{cov}}(x,y)}{\widehat{\mathrm{var}}(x)}, \qquad \widehat{\beta}_1 = \bar{y} - \widehat{\beta}_2 \bar{x}$$

Stochastische Regressionsanalyse: Intro

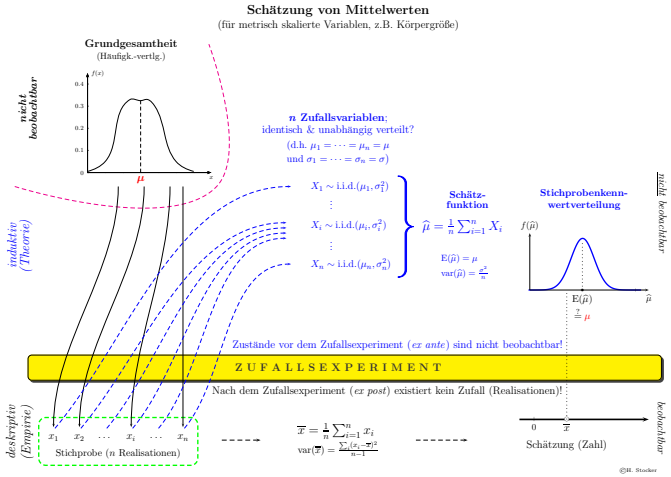
- Das einfache ‘Grundgesamtheit – Stichproben’ Modell ist nützlich für erste Einsichten und um die Notation zu verstehen . . .
  . . . aber es eignet sich weniger gut für tiefere Einsichten.

- Modell von R.A. Fisher ist allgemeiner und zeigt die prinzipiellen Ideen.
- Im Folgenden zeige ich die einfachst mögliche Anwendung, die Schätzung eines Anteils.
- Dieses Modell werden wir später für das Regressionsmodell erweitern . . .

R.A. Fishers Idee: **Schätzung eines Anteils** (Ziehen mit Zurücklegen)



Das funktioniert auch für metrisch skalierte Variablen:



Stochastische Regressionsanalyse: Intro

Dieses Modell werden werden wir später für die lineare Regression erweitern . . .

. . . aber vorher kommt noch die PRF und die bedingte Erwartungswertfunktion (‘*conditional expectation function*’, **CEF**)

PRF: $$y_i = \underbrace{\beta_1 + \beta_2 x_i}_{\textbf{CEF}} + \varepsilon_i$$

Stochastische Regressionsanalyse: PRF & CEF

Grundlagen der Ökonometrie

www.hsto.info/econometrics

PRF & bedingte Erwartungswerte

Übersicht:

- Bisher haben wir uns die Grundgesamtheit als eine nicht vollständig beobachtbare, aber gegebene Anzahl von Beobachtungen vorgestellt.
- Eignet sich weniger gut zur Beschreibung von Prozessen, z.B. Zusammenhang zwischen Inflation und Arbeitslosen.
- Allgemeinere Vorstellung:

  $\Rightarrow$ **Datengenerierender Prozess (DGP):** stochastischer Prozess, der die Daten erzeugt.
- DGP kann durch die *gemeinsame Verteilung* (Dichtefunktion) beschrieben werden.
- Kennzahlen (z.B. bedingte Erwartungswerte) können auch aus gemeinsamer Dichtefunktion berechnet werden!

PRF & bedingte Erwartungswerte

Beispiel: Grundgesamtheit für Autobeispiel mit gerundeten Daten, 100 Beobachtungen :

Obs	Alter (x)	Preis (y)	Obs	Alter (x)	Preis (y)
1	3	15000	51	2	20000
2	3	15000	52	1	20000
3	4	15000	53	5	15000
4	2	20000	54	6	10000
5	4	15000	55	4	15000
⋮	⋮	⋮	⋮	⋮	⋮
48	1	20000	98	2	15000
49	3	20000	99	5	10000
50	4	15000	100	2	20000

$\rightarrow$ Kontingenztabelle (Häufigkeiten auszählen) $\rightarrow$ DGP

PRF & bedingte Erwartungswerte

Grundgesamtheit mit 100 Beobachtungen: absolute Häufigkeiten

		Preis (y)					
		5000	10000	15000	20000	25000	Summe
Alter (x)	1	0	0	0	4	1	5
	2	0	0	2	15	2	19
	3	0	0	10	6	0	16
	4	0	3	25	0	0	28
	5	0	9	5	0	0	14
	6	2	15	1	0	0	18
	Summe	2	27	43	25	3	100

PRF & bedingte Erwartungswerte

Relative Häufigkeiten $\overset{?}{\leftrightarrow}$ Wahrscheinlichkeiten

		Preis (y)					
		5000	10000	15000	20000	25000	$\Pr(x)$
Alter (x)	1	0	0	0	0.04	0.01	0.05
	2	0	0	0.02	0.15	0.02	0.19
	3	0	0	0.10	0.06	0	0.16
	4	0	0.03	0.25	0	0	0.28
	5	0	0.09	0.05	0	0	0.14
	6	0.02	0.15	0.01	0	0	0.18
$\Pr(y)$		0.02	0.27	0.43	0.25	0.03	1

unbekannte Grundgesamtheit $\rightarrow$ **Wahrscheinlichkeitsfunktion** (Zähldichte)

PRF & bedingte Erwartungswerte

- Die gemeinsame Verteilung des DGP kann durch Kennzahlen beschrieben werden, z.B. bedingte Erwartungswerte $\rightarrow$ OLS
- **Erwartungswert:** mit Wahrscheinlichkeiten gewichtete Summe aller möglichen Ausprägungen des Zufallsexperiments
  (empir. Analogon: arithm. Mittel).

$$\mathrm{E}(y) = \sum_{j=1}^{m} y_j \Pr(y_j) \qquad \text{(diskr. ZV)}$$

- **Bedingter Erwartungswert:** Erwartungswert über eine durch eine *'Bedingung'* definierte Teilmenge (empir. Analogon: bedingter Mittelwert).

$$\mathrm{E}(y|x = \underline{x}) = \sum_{j} y_j \Pr(y_j|x = \underline{x}) \qquad \text{(diskr. ZV)}$$

PRF & bedingte Erwartungswerte

- **Bedingte Erwartungswertfunktion:** (*'Conditional Expectation Function'*, **CEF**) bedingte Erwartungswerte für alle möglichen Ausprägungen der *Bedingung* (empir. Analogon: bedingte Mittelwertfunktion).
- Beispiel: *bedingte Wahrscheinlichkeiten* $\Pr(y|x = \underline{x})$

Preis (y)		5000	10000	15000	20000	25000	Summe
Alter (x)	1	0.00	0.00	0.00	0.80	0.20	1.00
	2	0.00	0.00	0.11	0.79	0.11	1.00
	3	0.00	0.00	0.62	0.38	0.00	1.00
	4	0.00	0.11	0.89	0.00	0.00	1.00
	5	0.00	0.64	0.36	0.00	0.00	1.00
	6	0.11	0.83	0.06	0.00	0.00	1.00

Bedingte Erwartungswerte:

$$\mathrm{E}(y|x = \underline{x}) = \sum_{j} y_j \Pr(y_j|x = \underline{x}) \qquad \text{(diskr. ZV)}$$

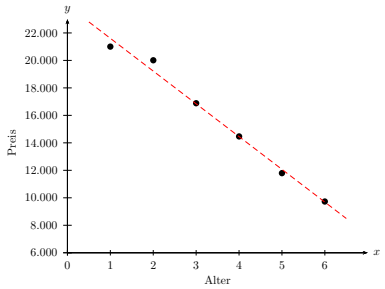
$$\mathrm{E}(\text{Preis}|\text{Alter} = 2) = 0 \times 5000 + 0 \times 10000 + 0.11 \times 15000 + 0.79 \times 20000 + 0.11 \times 25000 = 20000$$

PRF & bedingte Erwartungswerte

Bedingte Erwartungswertfunktion (CEF):

$$\mathrm{E}(y|x=\underline{x}) = \sum_j y_j \Pr(y_j|x=\underline{x}) \qquad \text{(diskr. ZV)}$$

$$\mathrm{E}(\text{Preis}|\text{Alter}=\underline{\text{Alter}}) = \begin{cases} 21000.00 & \text{für Alter = 1} \\ 20000.00 & \text{für Alter = 2} \\ 16875.00 & \text{für Alter = 3} \\ 14464.29 & \text{für Alter = 4} \\ 11785.71 & \text{für Alter = 5} \\ 9722.22 & \text{für Alter = 6} \end{cases}$$



Lineare Approximation an bedingte Erwartungswertfunktion: → OLS
Aber vorher . . .

PRF & bedingte Erwartungswerte

Eigenschaften der CEF:

- Linearität: wenn x, y, z Zufallsvariablen und a, b beliebige Konstante sind gilt

$$\mathrm{E}(ay + bz|x=\underline{x}) = a\,\mathrm{E}(y|x=\underline{x}) + b\,\mathrm{E}(z|x=\underline{x})$$

- Das Gesetz der iterierten Erwartungen: Für $\mathrm{E}(|y|) < \infty$ gilt

$$\mathrm{E}_{\underline{x}}[\mathrm{E}(y|x=\underline{x})] = \mathrm{E}(y)$$

- Konditionierungstheorem (*'Taking out what is known property'*):

$$\mathrm{E}(g(x)h(y)|x=\underline{x}) = g(\underline{x})\,\mathrm{E}(h(y)|x=\underline{x})$$

Zum Beispiel: $\mathrm{E}(x|x=\underline{x}) = \underline{x}$ oder $\mathrm{E}(xy|x=\underline{x}) = \underline{x}\,\mathrm{E}(y|x=\underline{x})$

PRF & bedingte Erwartungswerte

CEF Störterme:

Zerlegung von y in zwei Teile, CEF & Störterm

$$y = \mathrm{E}(y|x) + \varepsilon$$

daraus folgen CEF Störterme:

$$\varepsilon = y - \mathrm{E}(y|x)$$

Eigenschaften der CEF Störterme folgen aus den Eigenschaften der CEF!

PRF & bedingte Erwartungswerte

Eigenschaften der CEF Störterme:

- Die bedingten Erwartungswerte der CEF Störterme sind Null

$$\mathrm{E}(\varepsilon|x=\underline{x}) = 0$$

(folgt aus Konditionierungstheorem und Gesetz der iterierten Erwartungen, siehe Skript; *'mean independence'*)

- Auch die unbedingten Erwartungswerte der CEF Störterme sind gleich Null:

$$\mathrm{E}(\varepsilon) = \mathrm{E}_{\underline{x}}\left[\mathrm{E}(\varepsilon|x=\underline{x})\right] = 0$$

(folgt aus dem Gesetz der iterierten Erwartungen und dem Konditionierungstheorem)

- Das Produkt der CEF Störterme und dem Regressor x ist Null

$$\mathrm{E}(x\varepsilon) = 0$$

für $\mathrm{E}(|x\varepsilon|) < \infty$ (folgt wieder aus dem Gesetz der iterierten Erwartungen und dem Konditionierungstheorem)

PRF & bedingte Erwartungswerte

- Varianz der CEF Störterme (*'regression variance'*):

$$\text{var}(\varepsilon) := \sigma^2 = \text{E}[(\varepsilon - \text{E}(\varepsilon))^2] = \text{E}(\varepsilon^2)$$

- CEF als bester Prediktor:

$$\text{E}[y - g(x)]^2 \geq \text{E}[y - \text{E}(y|x)]^2 \qquad \forall\, g(x)$$

- **Bedingte Varianz der CEF Störterme:**

$$\text{var}(\varepsilon|x) = \text{E}(\varepsilon^2|x) := \sigma^2(x) \qquad \text{(für } \text{E}(\varepsilon^2) < \infty\text{)}$$

  - **homoskedastisch:** Varianzen der CEF Störterme sind für alle x gleich groß

$$\text{E}(\varepsilon^2|x) = \sigma^2$$

  - **heteroskedastisch:** Varianzen der CEF Störterme hängen von x ab (d.h. sind *nicht* für alle x konstant!)

$$\text{E}(\varepsilon^2|x) = \sigma^2(x)$$

PRF & bedingte Erwartungswerte

Die lineare CEF: (Spezialfall)

$$\text{E}(y|x = \underline{x}) = \beta_1 + \beta_2 \underline{x}$$

In zwei Fällen exakt erfüllt:

1. alle Variablen multivariat normalverteilt (in diesem Ausnahmefall sind die CEF Störterme homoskedastisch)
2. gesättigtes Dummyvariablenmodell (heteroskedastische Störterme)

In anderen Fällen:

1. OLS liefert lineare Approximation an CEF
2. Störterme oft heteroskedastisch

PRF & bedingte Erwartungswerte

Die lineare CEF: (lineare Projektion)

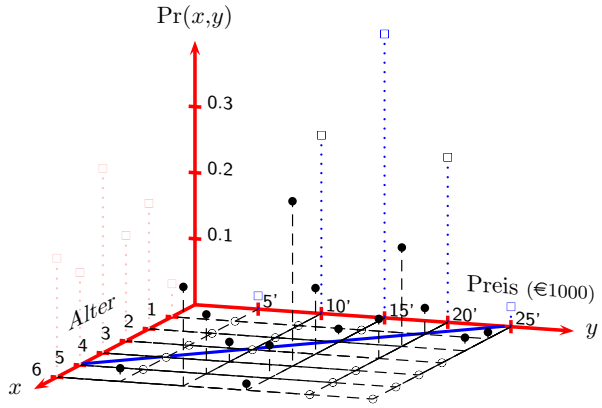
Im statistischen Intermezzo wird gezeigt, dass für die Koeffizienten einer linearen CEF gilt

$$\beta_2 = \frac{\text{E}[x - \text{E}(x)][y - \text{E}(y)]}{\text{E}[x - \text{E}(x)]^2} = \frac{\text{cov}(y, x)}{\text{var}(x)}$$

$$\beta_1 = \text{E}(y) - \beta_2\, \text{E}(x)$$

Diese Parameter beschreiben den **DGP**!
(systematischer Teil der PRF)

PRF & bedingte Erwartungswerte



Deterministische vs. stochastische Regressoren

- 'Bedingen auf' (*'conditioning'*) bedeutet ein 'Festhalten per Annahme', z.B. $\mathrm{E}(y|x = \underline{x})$ (d.h. kein Eingriff in die Natur)
- Dagegen: *Experiment:* für ein tatsächlich 'festgehaltenes' (*kontrolliertes*) x werden mehrere y gezogen (z.B. Feldexperimente von R.A. Fisher).
  $\Rightarrow$ *deterministische Regressoren*
- Beobachtungsdaten: $\rightarrow$ tatsächliches 'Festhalten' von x kaum möglich!
- Vorstellung: es werden jeweils x, y Paare gezogen, d.h., auch die x sind stochastisch!
  $\Rightarrow$ *stochastische Regressoren*
- Mathematik ist etwas aufwändiger (z.B. sind dann auch bedingte Erwartungswerte stochastisch) und es ist ein etwas erweitertes Annahmenset erforderlich, aber Schlußfolgerungen bleiben weitestgehend intakt!

Stochastische Regressionsanalyse für stetige ZV

Grundlagen der Ökonometrie

www.hsto.info/econometrics

Stetige Zufallsvariablen (ZV)

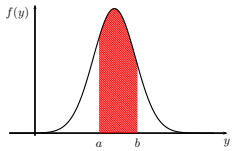
Wiederholung:

- Univariate Dichtefunktion: (ZV y)

$$f(y) \qquad \text{mit} \int_{-\infty}^{+\infty} f(y)\, dy = 1$$

- Dichte und Wahrscheinlichkeit:

$$\Pr(a < y < b) = \int_a^b f(y) dy \qquad \text{mit } \Pr(y = \underline{y}) = 0!$$

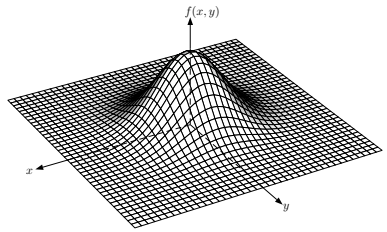


Stetige Zufallsvariablen

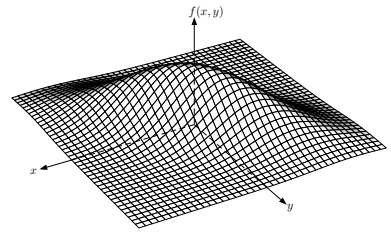
Wiederholung:

- Bivariate Dichtefunktion: (Zufallsvariablen y, x)

$$f(y, x) \qquad \text{mit} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(y, x)\, dy\, dx = 1$$



$\mathrm{corr}(x, y) = 0$



$\mathrm{corr}(x, y) = -0.7$

Stetige Zufallsvariablen

Wiederholung: Stetige ZV

- Randdichte: sei $f(y,x)$ eine gemeinsame Dichte der Zufallsvariablen y und x, dann ist die Randdichte

$$f_y(y) = \int_{\mathbb{R}} f(y,x)\,dx$$

- (unbedingter) Erwartungswert:

$$\mathrm{E}(y) = \int_{-\infty}^{+\infty} y\, f_y(y)\,dy$$

Stetige Zufallsvariablen

Wiederholung:

- Bedingte Dichte:

$$f_{y|x}(y|x) = \frac{f(y,x)}{f_x(x)}$$

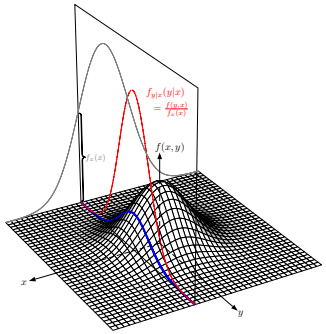
- Bedingter Erwartungswert:

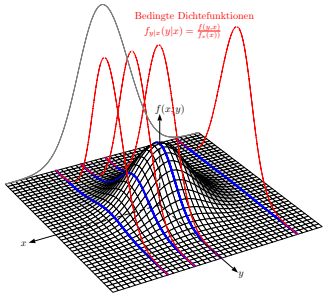
$$\mathrm{E}(y|x) = \int_{-\infty}^{+\infty} y\, f_{y|x}(y|x)\,dy$$

- Stochastische Unabhängigkeit:

$$f(y,x) = f_y(y) f_x(x) \qquad \text{bzw.} \qquad f_{y|x}(y|x) = \frac{f(y,x)}{f_x(x)} = \frac{f_y(y) f_x(x)}{f_x(x)} = f_y(y)$$

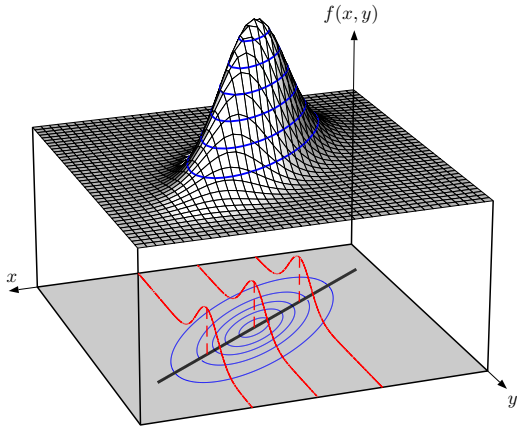
Stetige Zufallsvariablen

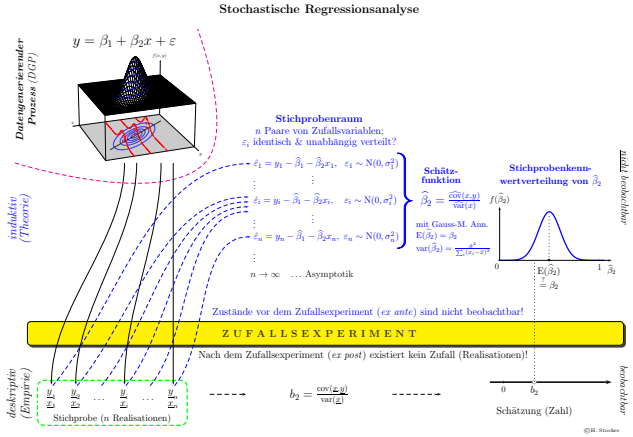




$$\mathrm{E}(y|x) = \int_{\mathbb{R}} y\, f_{y|x}(y|x)\,dy$$

Datengenerierender Prozess und CEF mit Störtermen





DGP $\rightarrow$ PRF $\rightarrow$ Stichprobenkennwertverteilung

- DGP und dessen gemeinsame Verteilung unbeobachtbar; meist komplizierte Verteilung mit oft heteroskedastischen Störtermen.
- Stichprobenkennwertverteilung:
  - Lage unbekannt!
  - Wir wissen: empirische Schätzung ist eine Realisation aus dieser Verteilung!
- Die Eigenschaften der Stichprobenkennwertverteilung sowie deren Parameter (Erwartungswert & Varianz) stehen im Zentrum des nächsten Kapitels!
- Stichprobenkennwertverteilungen werden für Hypothesentests benötigt!

Thanx . . .