6 Ausblick

6.1 Fehlerbehaftete erklärende Variable

a x_i sind oft zufällig. Das ist o.k. $\longrightarrow$ Bedingte Analyse, gegeben die x_i.

Anderes Problem: x mit Messfehlern. (Wahrer Wert nicht zufällig!)

b „Latente" Variable u, v

$$v = \widetilde{\alpha} + \widetilde{\beta} u$$

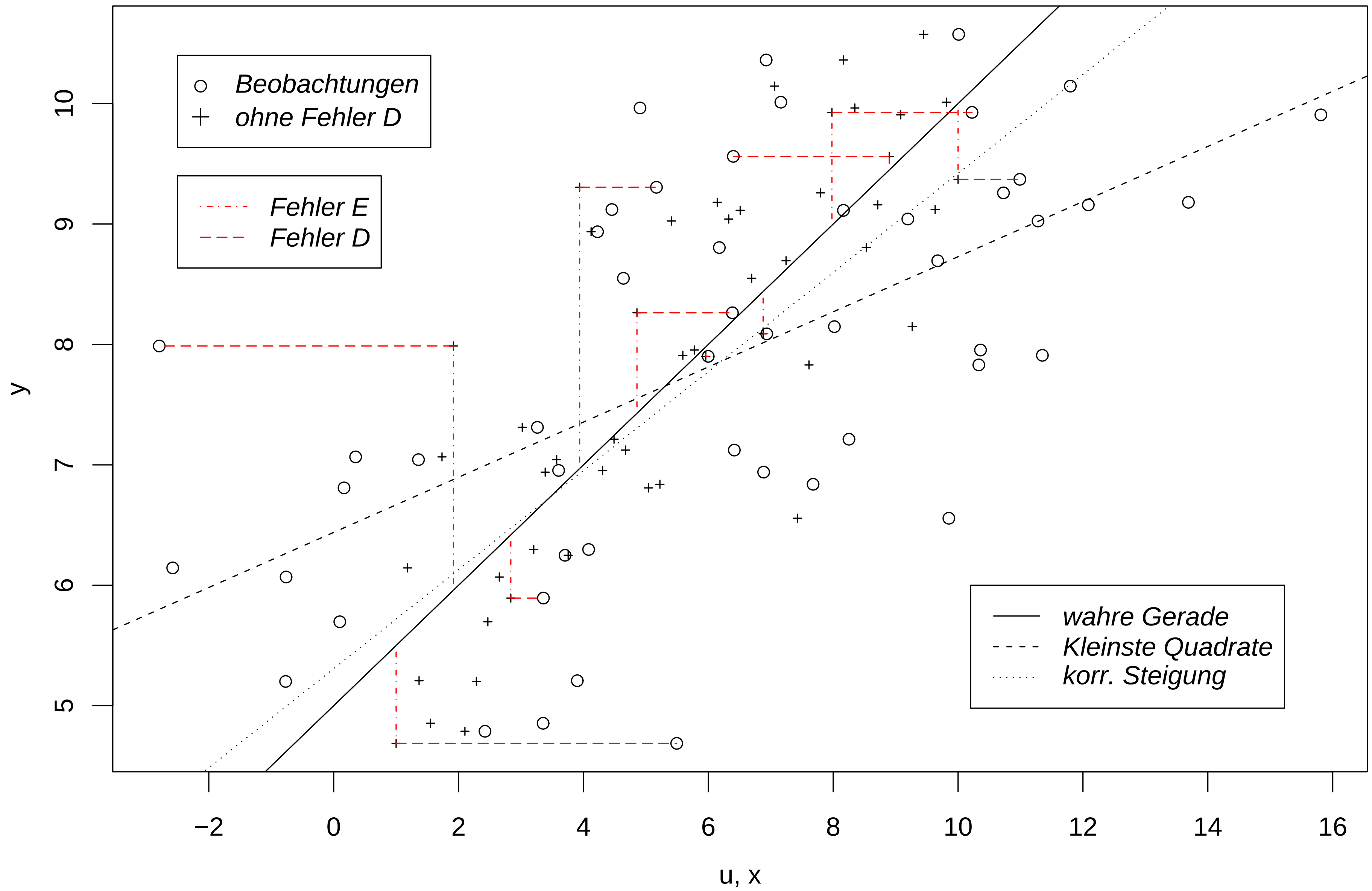
$$X_i = u_i + D_i \;, \qquad Y_i = v_i + E_i = \widetilde{\alpha} + \widetilde{\beta} u_i + E_i$$

$$D_i \sim \mathcal{N}\langle 0, \sigma_D^2 \rangle \;, \qquad E_i \sim \mathcal{N}\langle 0, \sigma^2 \rangle$$

unabhängig.

Wie gross sind $\widetilde{\alpha}, \widetilde{\beta}$?

c errors-in-variables, functional and structural model d



Beobachtungen
ohne Fehler D
Fehler E
Fehler D
wahre Gerade
Kleinste Quadrate
korr. Steigung
y
u, x
5
6
7
8
9
10
−2
0
2
4
6
8
10
12
14
16

e Schätzung: Falls σ_D^2 bekannt:

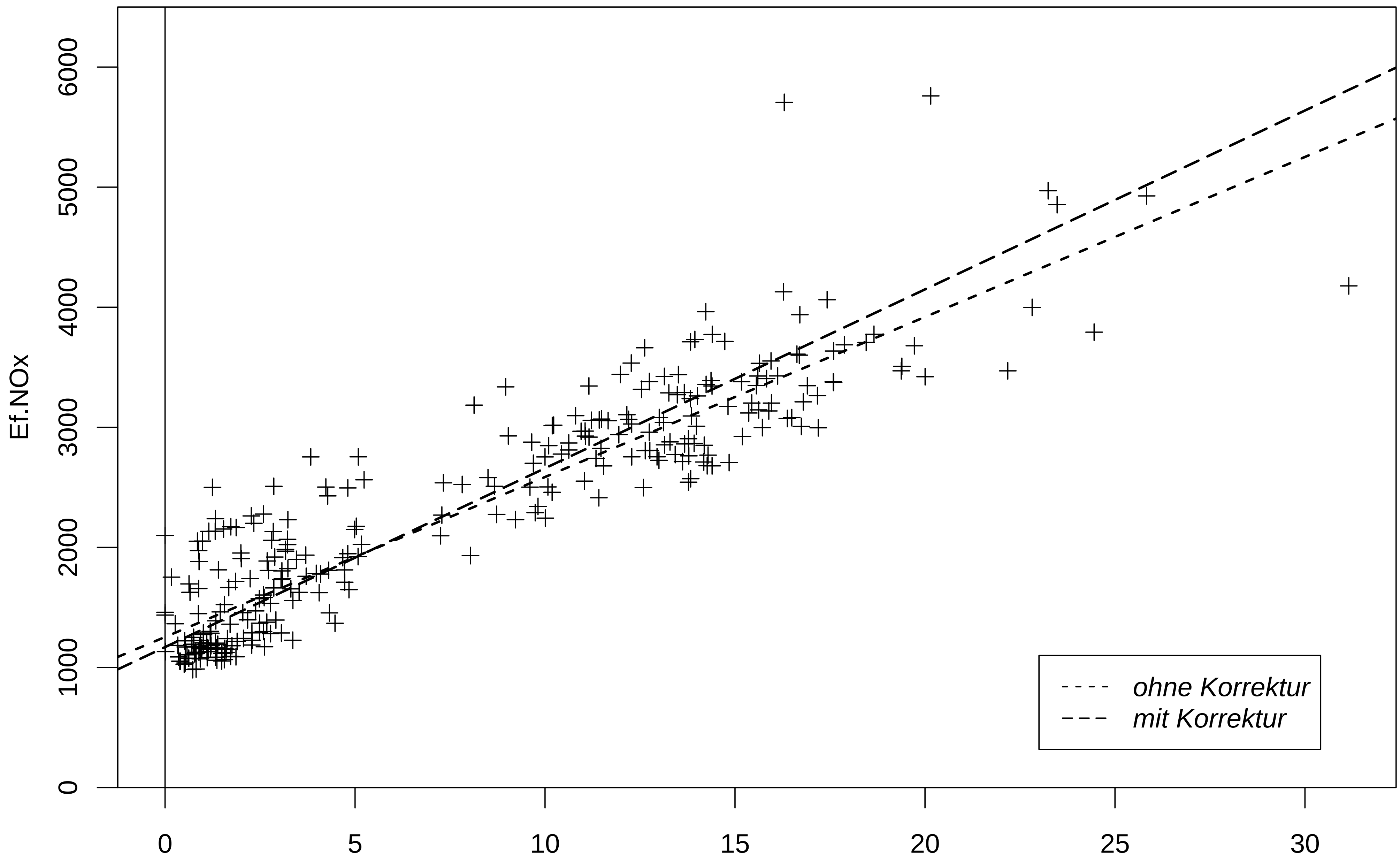
$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2 - \sigma_D^2} = \widehat{\beta}_{LS}/\widehat{\kappa}$$

$$\widehat{\kappa} = \frac{\widehat{\mathrm{var}}\langle X\rangle - \sigma_D^2}{\widehat{\mathrm{var}}\langle X\rangle}$$

κ: (attenuation coefficient)

f $\widehat{\alpha} = \overline{Y} - \widehat{\beta}\,\overline{X}$

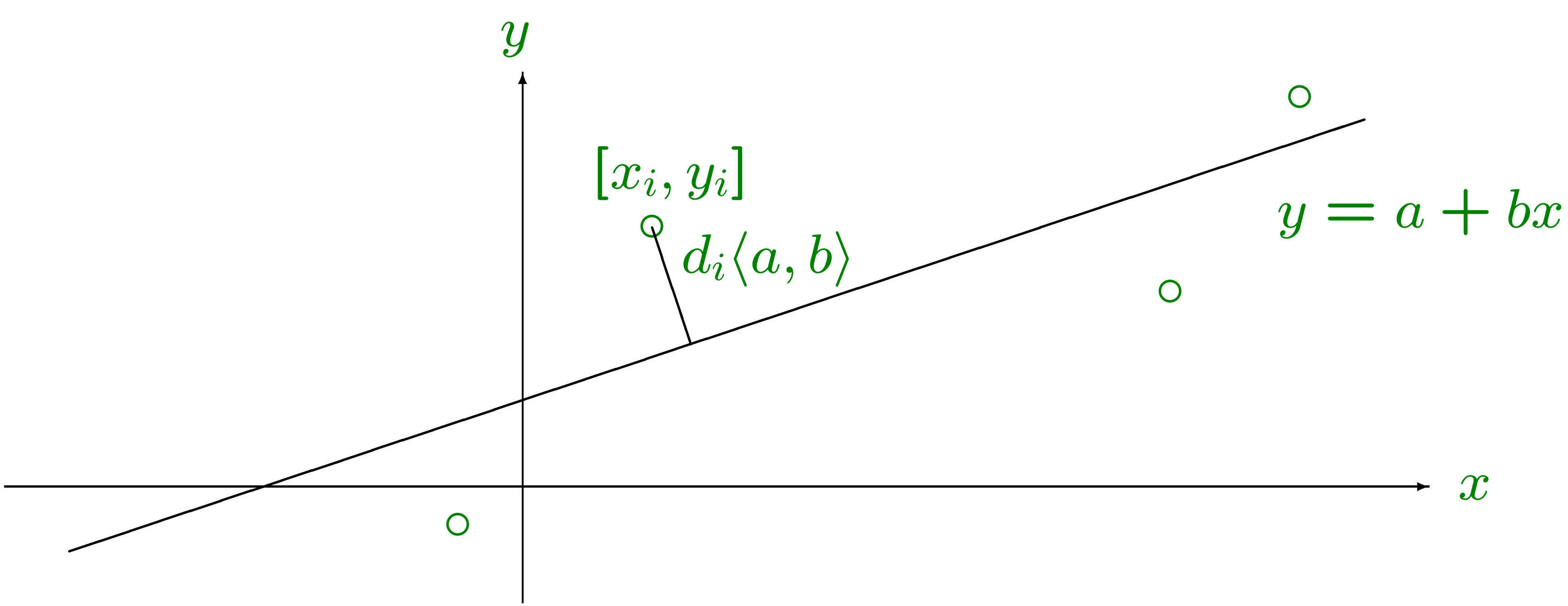
g Im Beispiel der Schadstoffe im Tunnel



i Schätzung: Falls $\gamma = \sigma/\sigma_D$ bekannt:
X „umskalieren", so dass $\gamma = 1$ ist
$\longrightarrow$ orthog. Regression, zurückrechnen.

j Orthogonale Regression: Minimiere $\sum_{i=1}^{n} d_i^2\langle a,b\rangle$.

Hauptkomponenten-Analyse.



k Man muss σ/σ_D oder σ_D kennen.

l Anwendung:

- Vorhersage: gewöhnliche Regression (Achtung!)
- Test für Einfluss ($\beta = 0$): gewöhnliche R.
- Schätzung von $\widetilde{\alpha}, \widetilde{\beta}$: errors-in-variables

6.2 Eichung

a Aus dem Resultat einer (billigen) Mess-Methode
das Resultat einer anderen (teuren) „schätzen".

Bestimmung des Zusammenhangs:
Exakte Werte vorgegeben x_i. Anwendung der billigen Methode $\to Y_i$.

Anwendung: von Y_0 auf x_0 schliessen.

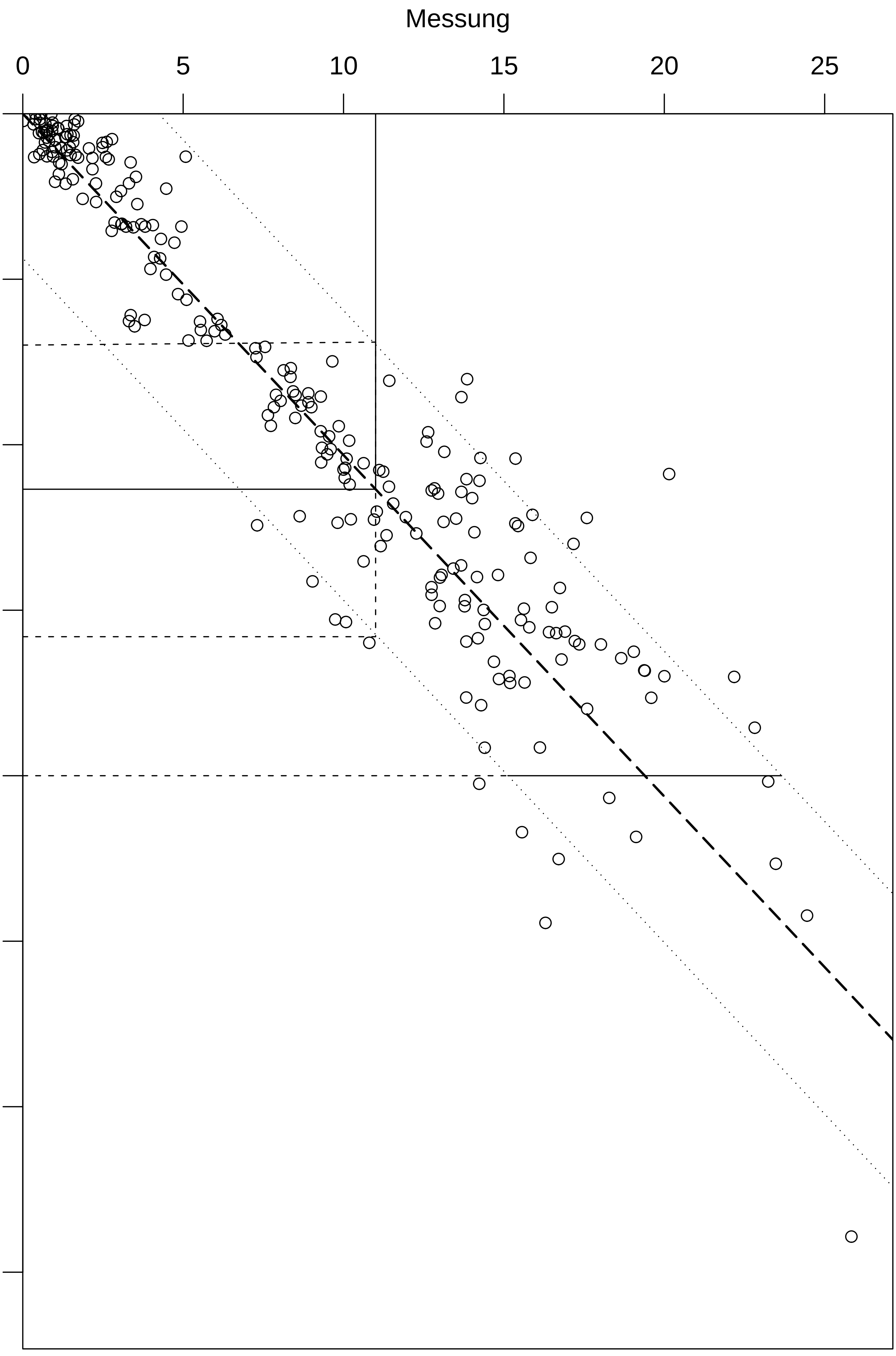
Inverse Regression, Calibration

b Schätzung: $\widehat{\alpha}, \widehat{\beta}$ bestimmen $\longrightarrow \widehat{x} = (Y - \widehat{\alpha})/\widehat{\beta}$

c Wie genau ist dieser Wert? – Vorhersage-Band

$\widehat{\alpha} + \widehat{\beta} x_0 \pm b$ mit $b = q_{0.975}^{t_{n-2}} \, \widehat{\sigma} \sqrt{1 + \frac{1}{n} + (x_0 - \bar{x})^2 / \mathrm{SSQ}^{(X)}}$

Umkehrung: $(y - \widehat{\alpha})/\widehat{\beta} \pm b/\widehat{\beta}$


Messung
0
5
10
15
20
25

d* Einige weitere Stichworte:

- Fehlerbehaftete x-Werte:

  Man verwende eine Schätzung der „wahren Geraden" $\widetilde{\alpha} + \widetilde{\beta}\, x$.

- Überprüfung der Linearität und anderer Modell-Annahmen!

- Periodische Eichung: sollte nicht mit Einzelmessungen erfolgen.

e* Eichproblem und fehlerbehaftete x:

„Naives" Vorgehen beim Eichproblem:

Da man die x-Werte aus Y „vorhersagen" will, vertauscht man die Rollen

aber: Dann ist die falsche Variable mit Messfehlern behaftet.

Errors-in-variables-Modell liefert richtige Gerade mit $\sigma/\sigma_D = 0$.

Aber: Bei Vorhersage war doch Kl.-Quadrate-Schätzung richtig. ???

Nur, wenn für eine neue Beobachtung die bedingte Vert. der Zielgrösse, geg. die Ausgangsgrösse, gleich ist wie bei den „Trainings-Daten"!

Merkpunkte

1. Fehlerbehaftete x-Variable:

   Wenn Koeffzienten geschätzt werden sollen, muss man korrigieren!

   Kleinste Quadrate geben zu flache Gerade.

2. Eich-Problem = inverse Regression

   Aufgepasst mit der Festlegung von x und Y!