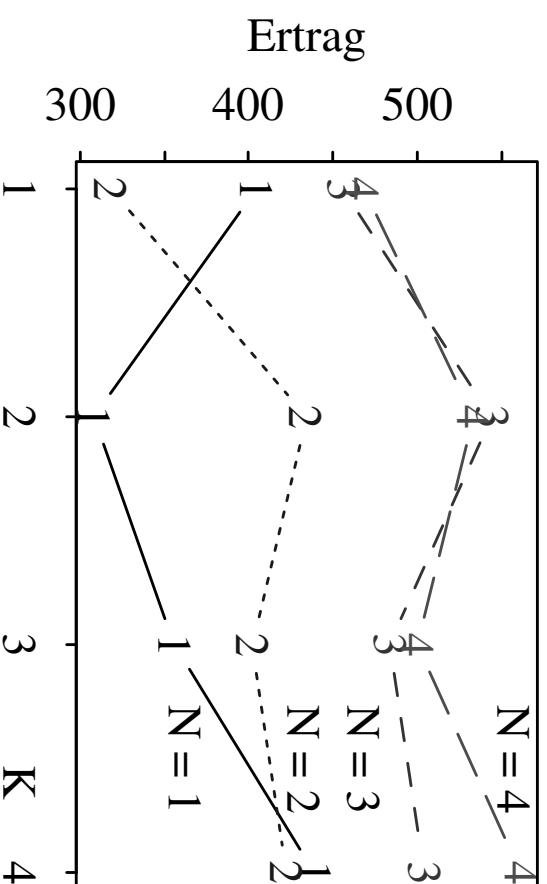


## 16 Ausblick

### 16.1 Varianzanalyse, Allgemeines Lineares Modell

#### a Varianzanalyse mit festen Effekten



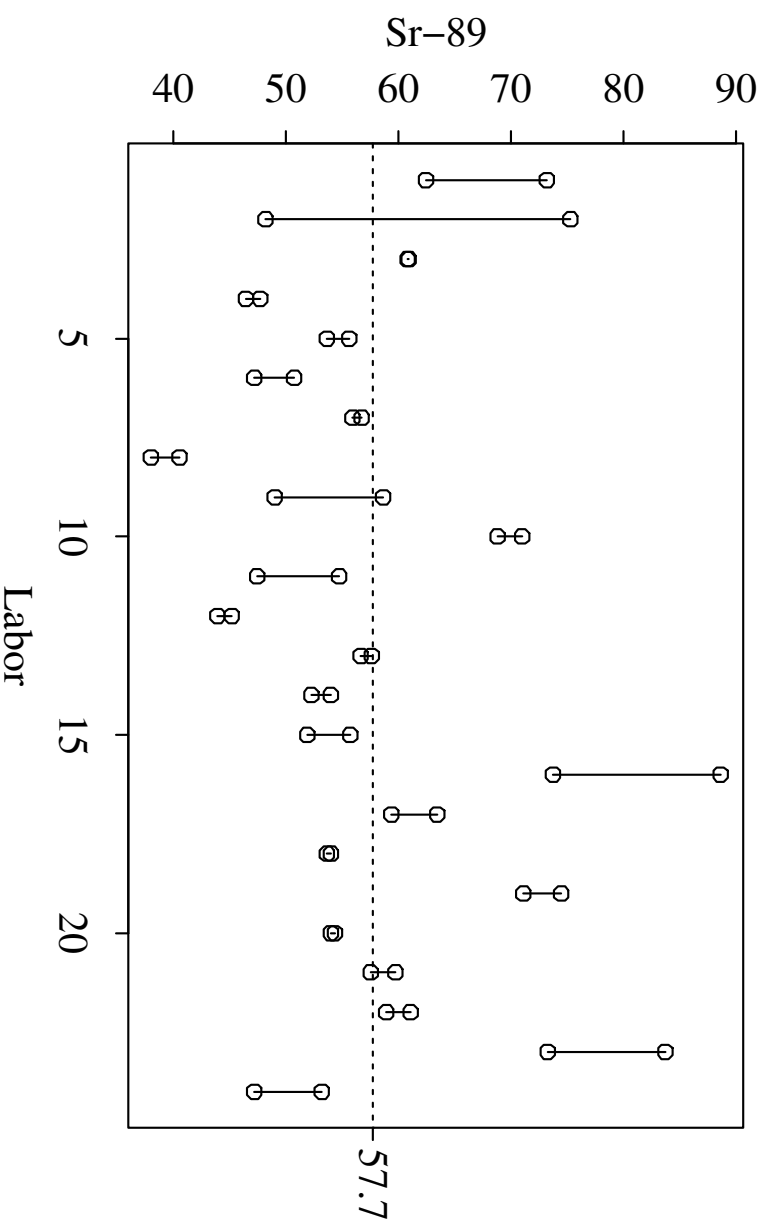
Unterkapitel der Regression ... mit nützlichen Vereinfachungen

#### b **Versuchsplanung.** Geplante Versuche ermöglichen Ursachenforschung

Möglichst wenige Versuche für möglichst viel Information

Optimale Produktionsbedingungen finden → response surfaces

- c **Zufällige Effekte** Ringversuch



- d **Repeated Measures, Split Plot Designs** Verlaufskurven, Spaltanlagen  
 e **Zufällige Effekte allgemein**  
 f **Programme SAS**
- R: lme noch unvollständig und anspruchsvoll

## 16.2 Korrelierte Fehler, Zeitreihen-Regression

a **Korrelierte Fehler** Sequence Plot, Tests

b **AR-Modell für Fehler**

$$Y_i = \bar{x}_i^T \bar{\beta} + E_i, \quad E_i = \alpha E_{i-1} + U_i, \quad U_i \text{ unabhängig.}$$

c **Generalized Least Squares**

Korrelationen zwischen  $E_i$ 's bekannt  $\rightarrow$  Verrillg. KI.Qu.

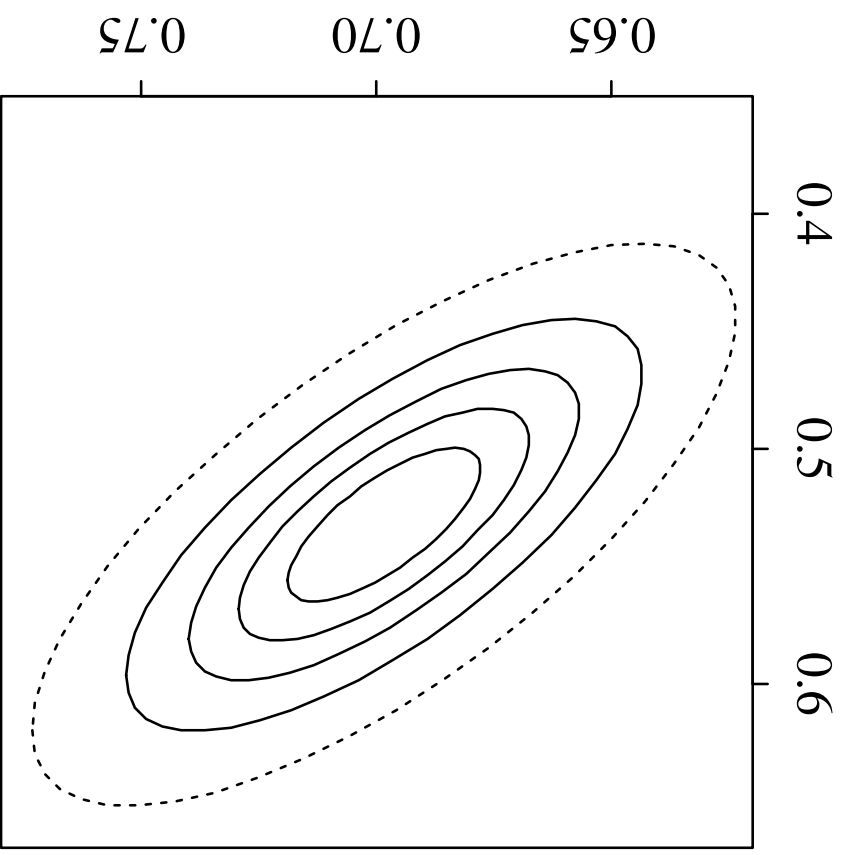
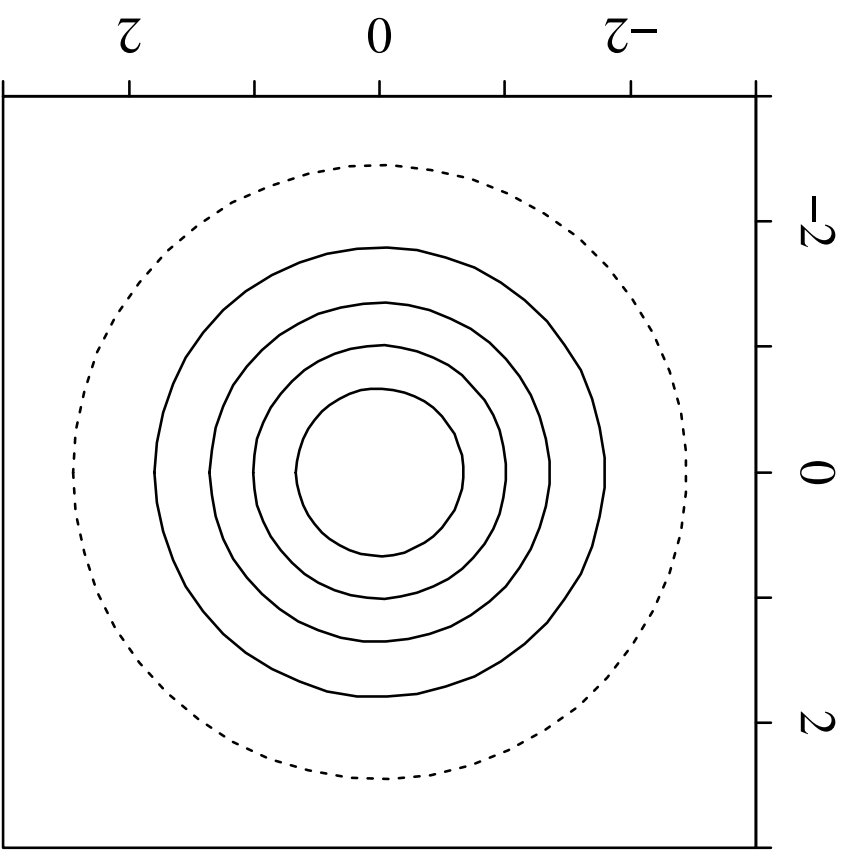
d **Räumliche Korrelation**

Korrelation zwischen  $E_i$ 's hängt vom Abstand der Beobachtungsorte ab.

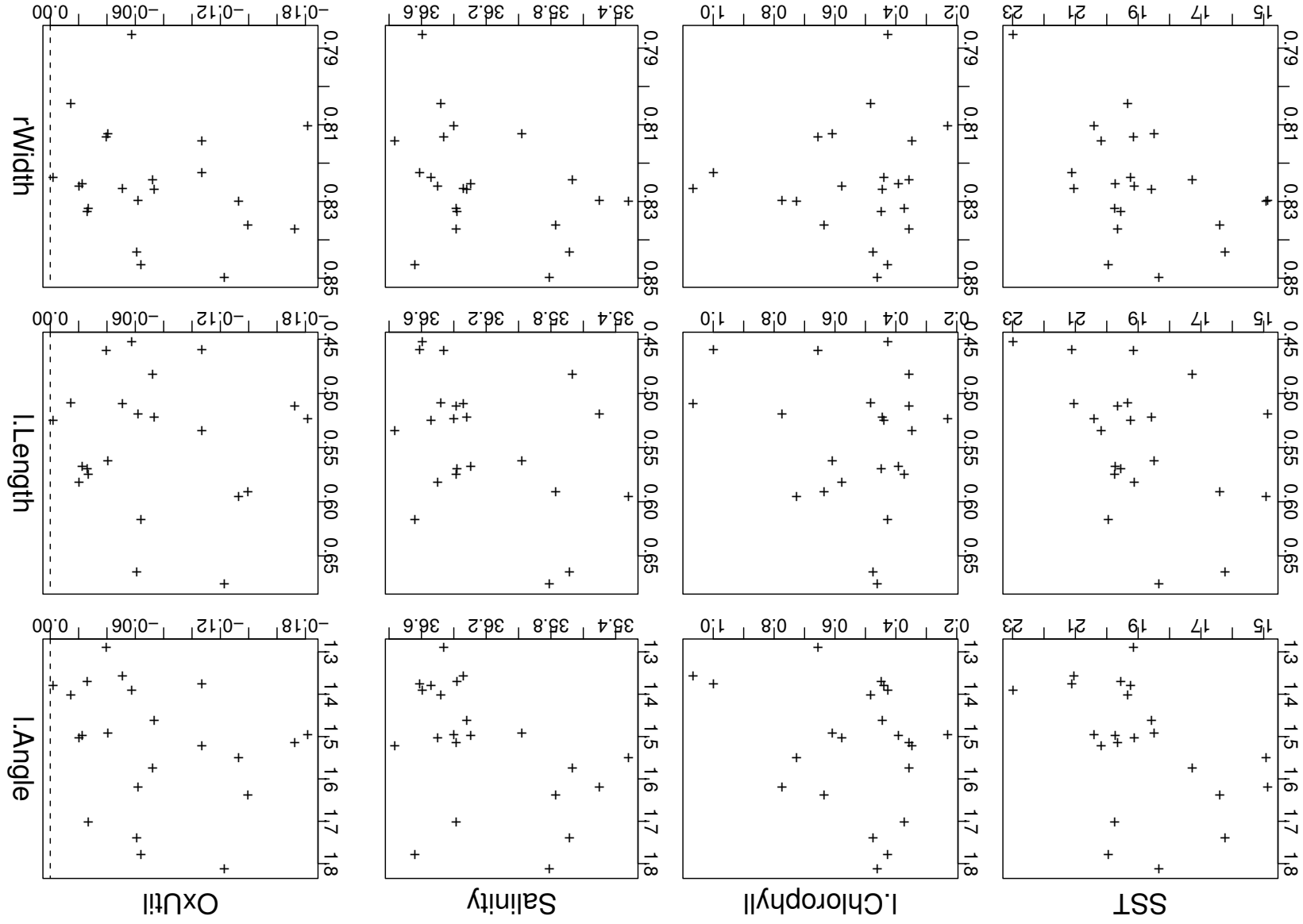
## 16.3 Multivariate Regression

a Mehrdimensionale Normalverteilung

charakterisiert durch Erwartungswerts-**Vektor** und Varianz-**Matrix**  
 $\bar{E} \sim \mathcal{N}_m(\bar{\mu}, \mathbb{K})$



## b Mehrere Zielgrößen, mehrere gewöhnliche Regressionen

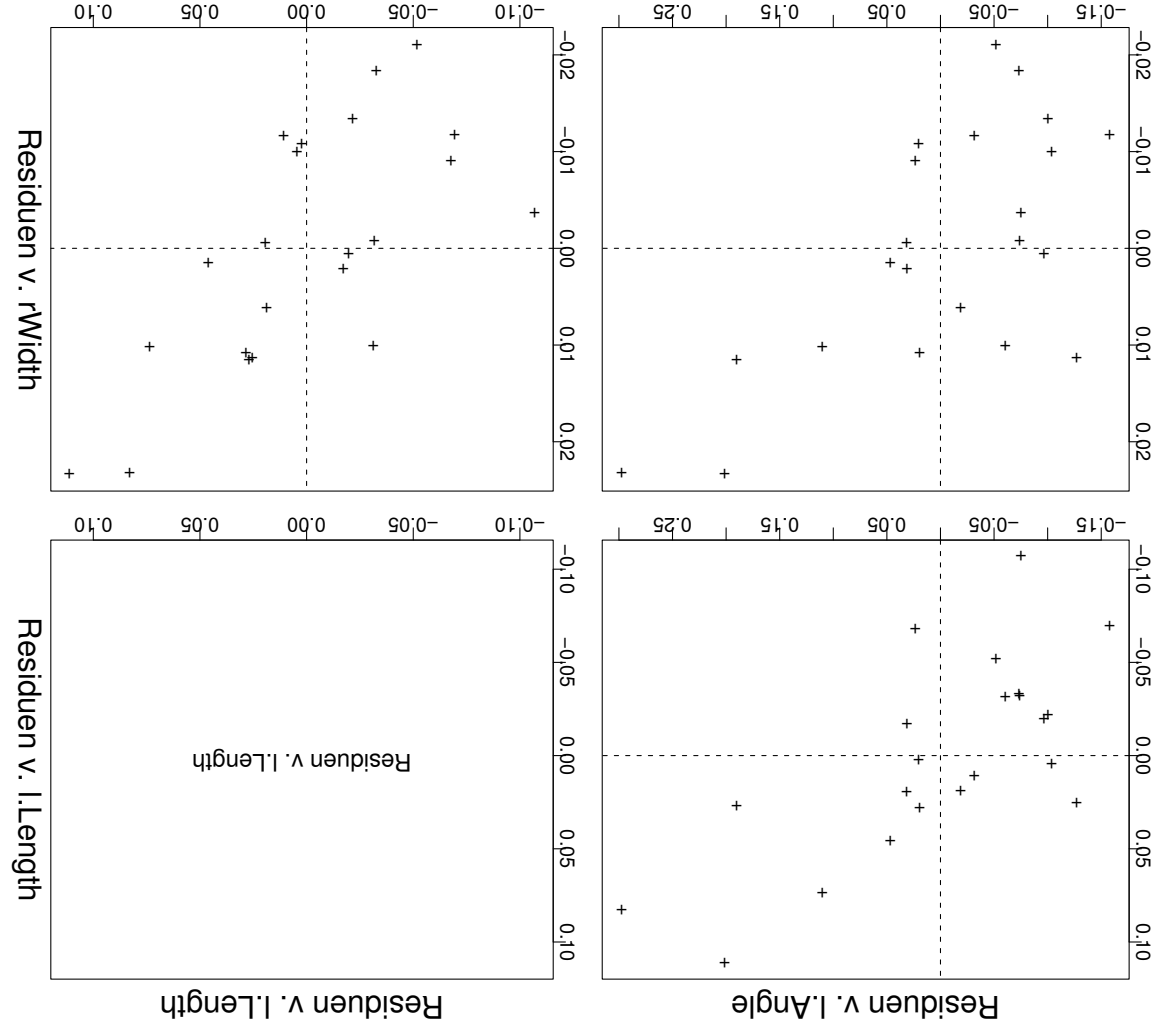


c **Modell**  $\bar{Y}_T^i = \bar{x}_T^i \beta + \bar{E}_T^i$

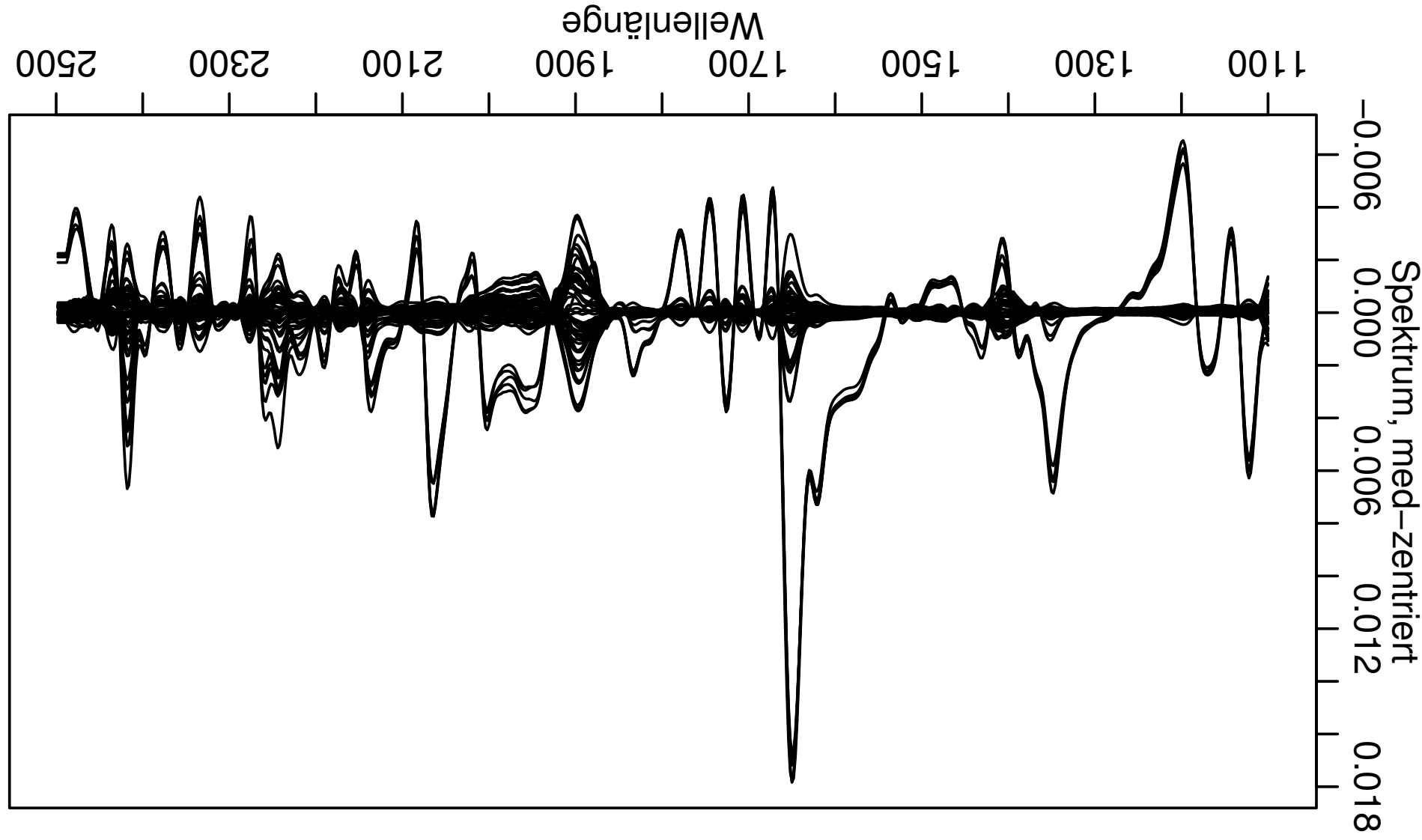
d **Schätzung der Koeffizienten**  $\beta$   
 durch „univariate“ Regressionen.

e **Zusätzlich: Gesamtests** für alle Zielgrößen gleichzeitig  
 Einfluss eines Regressors / Terms : alle  $\beta^{(j)} = 0$

- f **Zusätzlich: Korrelierte Fehler** → Kovarianzmatrix der Residuen  
**Partielle Korrelation:** Korr. der  $Y^{(j)}$ , gegeben  $\bar{x}$ .



9 **Spektrien, Functional Data Analysis:** Kurven als Eingangs- od. Zielgrößen



h **Simultane Gleichungen** Ökonometrie

! **Strukturgleichungs-Modelle und grafische Modelle**

Ganzes Netz von Variablen mit Regressions-Zusammenhängen  
Variablen sind nur teilweise beobachtbar.

**Latente Variable** zur Formulierung eines Modells

Vgl. Geordnete Zielgr., Faktoranalyse, Zustandsraum- (Zeitreihen-) Modelle

## 16.4 Nichtlineare Regression

- a **Modell**  $Y_i = h(\bar{x}_i; \bar{\theta}) + E_i$  Parameter  $\bar{\theta}$ .
- b **Methodik analog zur linearen Regression.** Kleinste Quadrate
- c **Näherung durch lineare Regression.** Iterativer Algorithmus.  
Statistik: Standard-Fehler der gesch. Parameter  $\rightarrow$  Tests, Vertrauensint.
- d **Schwierigkeiten:** Startwerte, ungenaue Vertrauensintervalle  
 $\rightarrow$  Kunst der Parameter-Transformation
- e **„Systemanalyse“** Nichtlin. Regr. mit Differentialgleichungen.  
Berechnung von  $h(\bar{x}_i; \bar{\theta})$  braucht numerische Lösung  
eines Diff.-gl.-systems

## 16.5 Nichtparametrische Regression

a „Glättung“

b Mehrere Dimensionen: zu wenige Beobachtungen

c General Additive Model

d Einsatz in gewöhnlicher Regression

## 16.6 Überlebenszeiten, zensierte Zielgrößen

- a **Verteilungen:** Exponential, Weibull, Gamma, Lognormal
- b **Verteilungen der logarithm. Grösse:** Gumbel, Normal (ausser Gamma)
- c **Zensierte Grössen**
- d **Regression:**
  - Parametrisch mit Weibull-Verteilung
  - Ohne Annahme über die Fehler-Verteilung mit **Cox-Regression**

## Regressionsmodelle im Überblick

### Merkpunkte

- Modelle für eine **kontinuierliche** Zielgröße:
  - lineare R., Kl.Qu. oder robust
  - nichtlineare R.
  - GLM mit Gamma-Verteilung, ...
  - Überlebenszeiten: parametrisch oder Cox (mit zensierten Daten)
  - nichtparametrische R. (Glättung), general additive model
  - Zeitreihen-R.
  - multivariate R.
  - Strukturgleichungsmodelle, grafische Modelle
- Modelle für andere Zielgrößen:
  - **logistische R., kumulative Logits, multinomiale R.**
  - Poisson-R. (log-lineare Modelle)

- **Schätzungen:** meistens „problemlos“, da programmiert
- **Robuste Varianten?**
  - ↳ **Tests:** meistens nur über asymptotische Verteilung! (o.k.?)
  - ↳ Simulation, „bootstrap“! Robustheit?
- **Vorhersage:** Man braucht nicht unbedingt die „richtigen“ Terme.
  - Beurteilung der Vorhersagefehler (Abweichung / Fehlklassifikation) mit Trainings-Datensatz zu optimistisch
  - ↳ Test-Datensatz, Kreuzvalidierung
- **Modell-Entwicklung** je nach Fragestellung und Grösse des Datensatzes.
  - Automatisierte Modellwahl
  - Strategie der Datenanalyse!
  - **Fachwissen und Kreativität verwenden!**

( Viel Vergnügen! )