

Analyse von Diabetes-Daten

B. Spangl und R. Dutter

Institut für Statistik und Wahrscheinlichkeitstheorie

Technische Universität Wien

Wiedner Hauptstr. 8-10, A-1040 Wien

Inhalt

- **Motivation**
- **Daten**
- **Hauptkomponentenanalyse**
 - Ergebnisse
- **Zeitreihenanalyse**
 - Deskriptive Methoden (ACF, CCF, PACF)
 - Ergebnisse
- **Ausblick**

Motivation

Das Ziel ist, ein besseres Verständnis der Zusammenhänge verschiedener medizinischer Variablen und deren Einflüsse auf Spätfolgen bei Diabetikern zu bekommen.

Daten

- **Medizinische Diabetes-Datenbank**

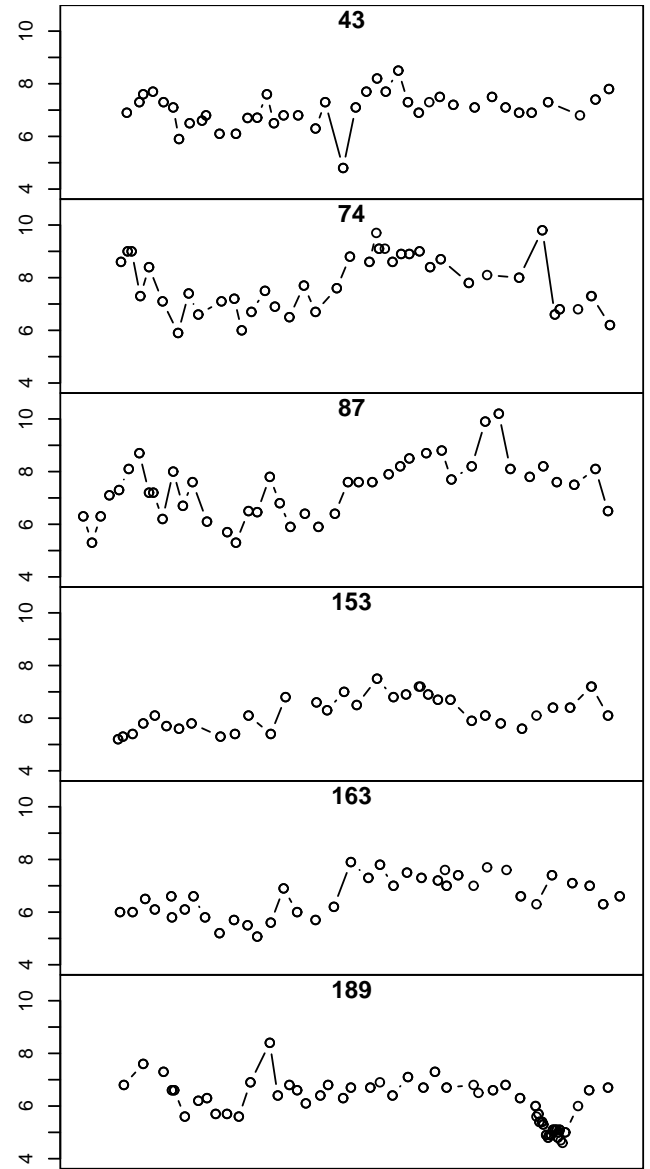
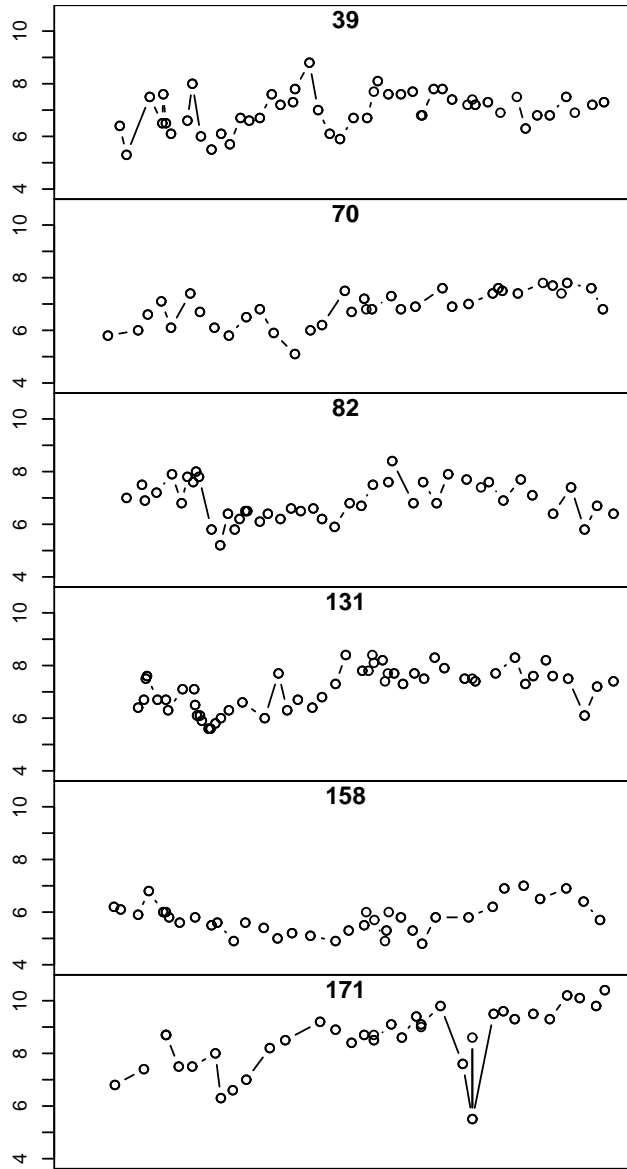
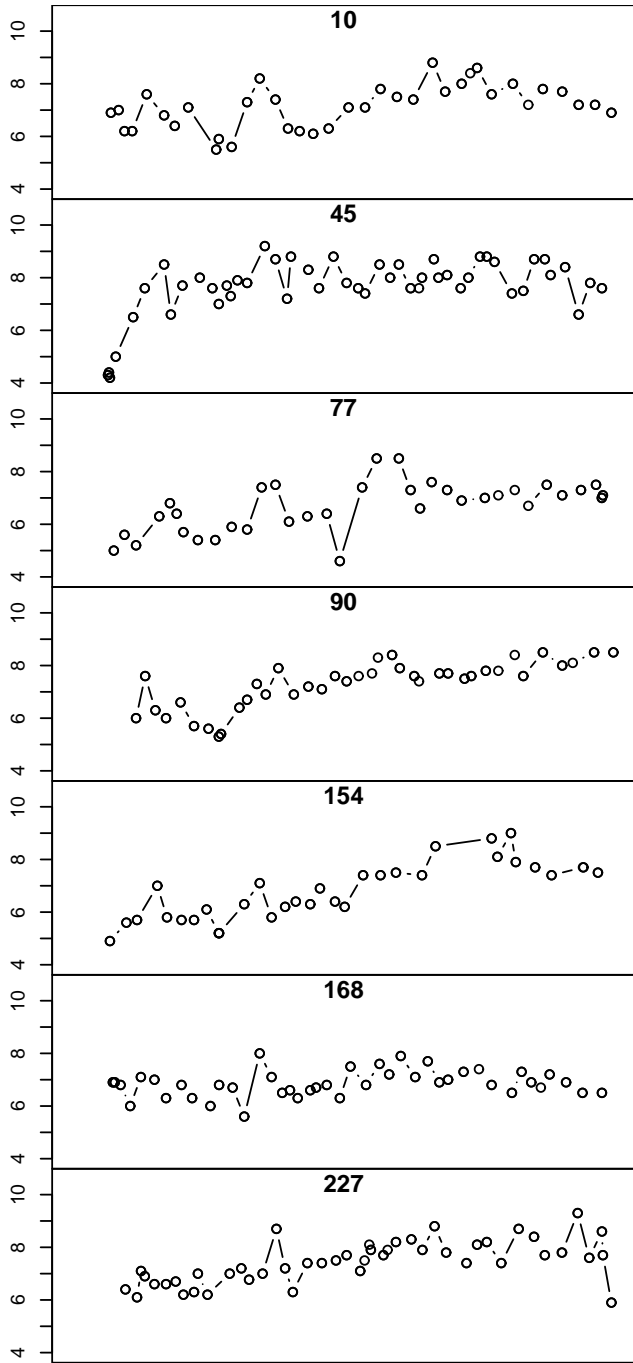
Univ.-Prof. Dr. Howorka, Inst. f. biomed. Technik und Physik, AKH

- **852 Patienten**

- **Zeitraum 1988 bis 2000**

- patientenbezogene **Untersuchungsergebnisse** und **Laborbefunde**

- **Variable:** Blutwerte (z.B.: HbA_{1c} , ...), Harnwerte, Blutdruck, Größe, Gewicht, ...



Multivariater Datensatz

- verwendete Variable:

1	trans_HbA1c	4	BG	7	RR_syst
2	Cholesterin	5	Mikroalbuminurie	8	RR_diast
3	Triglyceride	6	Albuminexkr_rate	9	bmi

→ **Datenmatrix** mit **fehlenden Werten**

Nonlinear Iterative Partial Least Squares (NIPALS)

- Alternative zur klassischen **Hauptkomponentenanalyse**
- **Modell:** $X = TP^T + \varepsilon$
- **Grundidee:** Alternierende Regression der Spalten bzw. Zeilen

$$X_{h-1} = t_h p_h^T + \varepsilon_1$$

↑↓

$$X_{h-1}^T = p_h t_h^T + \varepsilon_2^T$$

- für **fehlende Werte** adaptierbar (→ **minimaler Informationsverlust**)

Zeitreihenanalyse

- **Diabetes-Daten = “cross-sectional” Daten**, d.h.,

$$\mathbf{y}_{tl} = (y_{1,t,l}, \dots, y_{i,t,l}, \dots, y_{m,t,l})^T, \quad (1)$$

mit $t = 1, \dots, n$ und $\ell = 1, \dots, N$, wobei

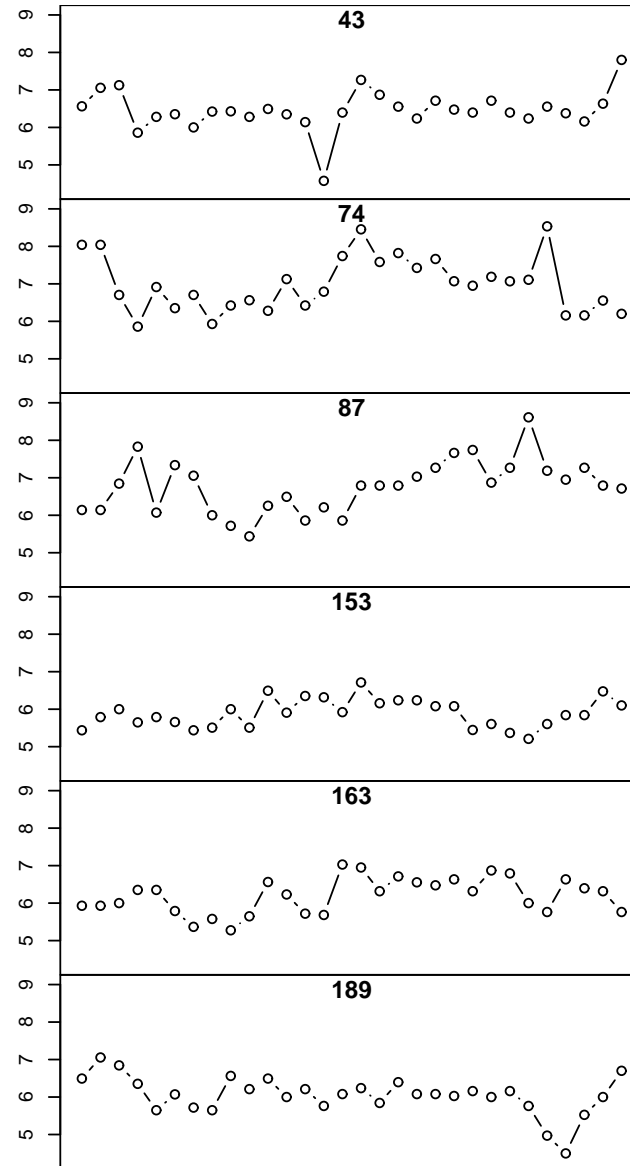
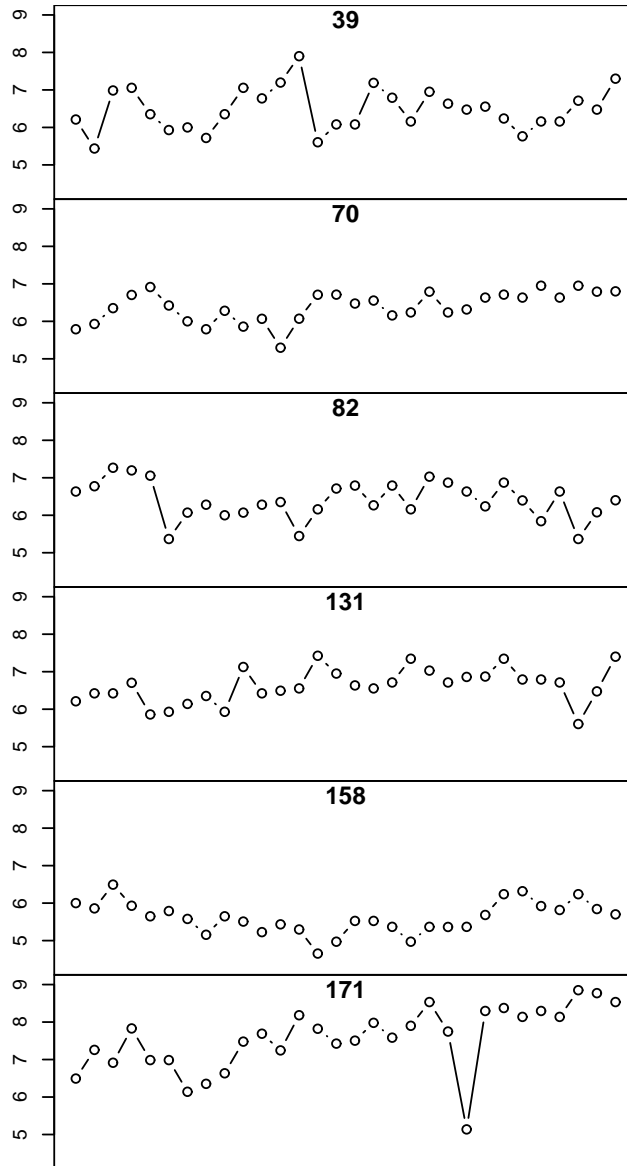
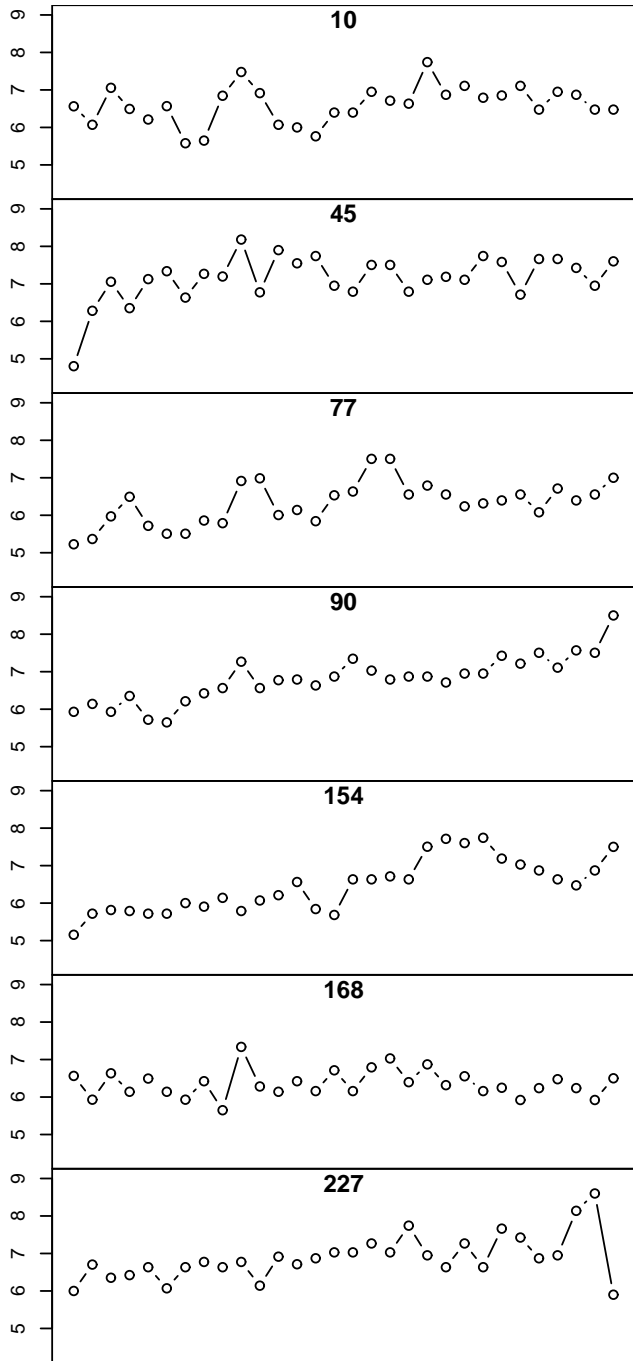
N ... Anzahl der Patienten

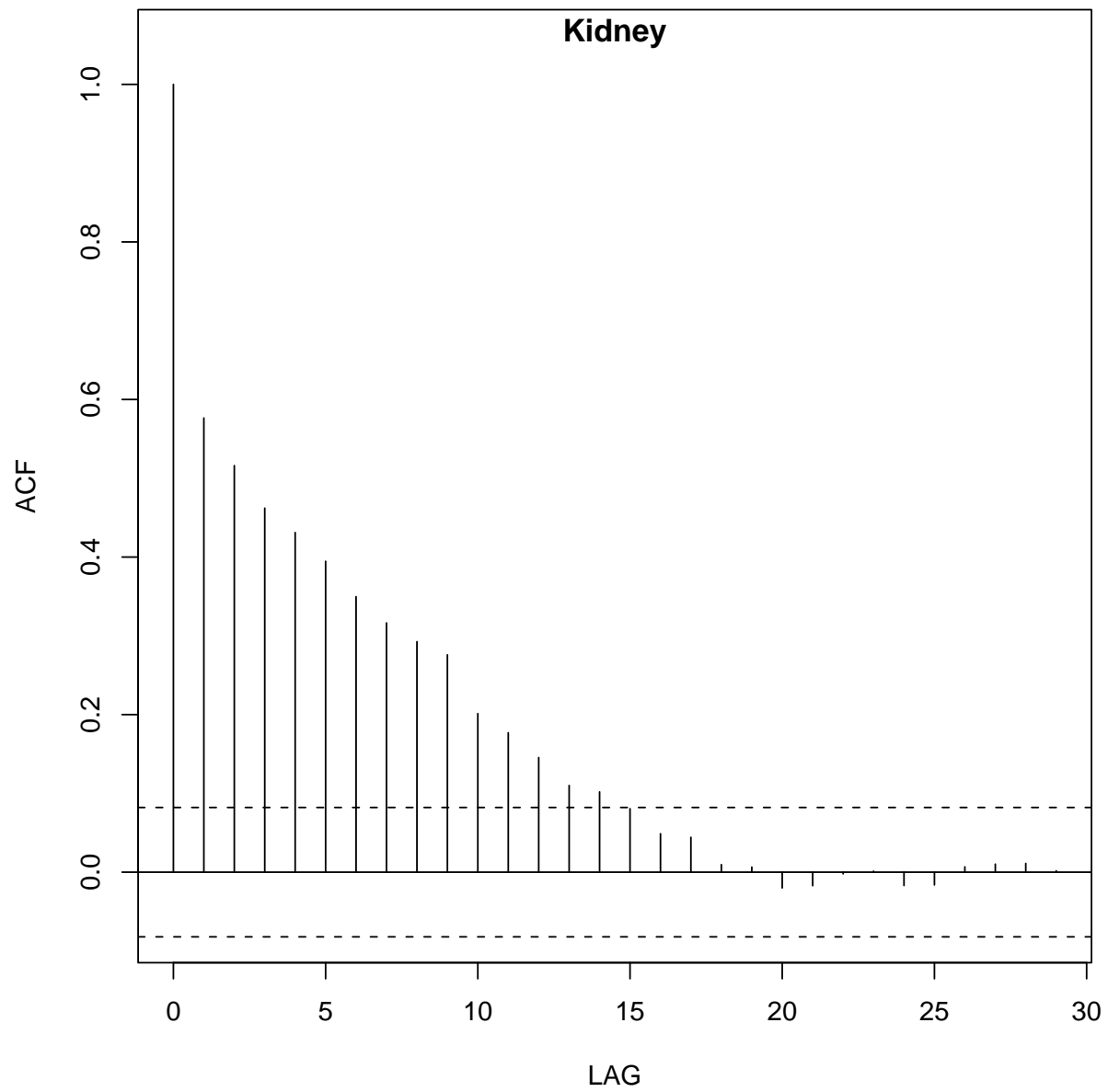
n ... Anzahl der Beobachtungen

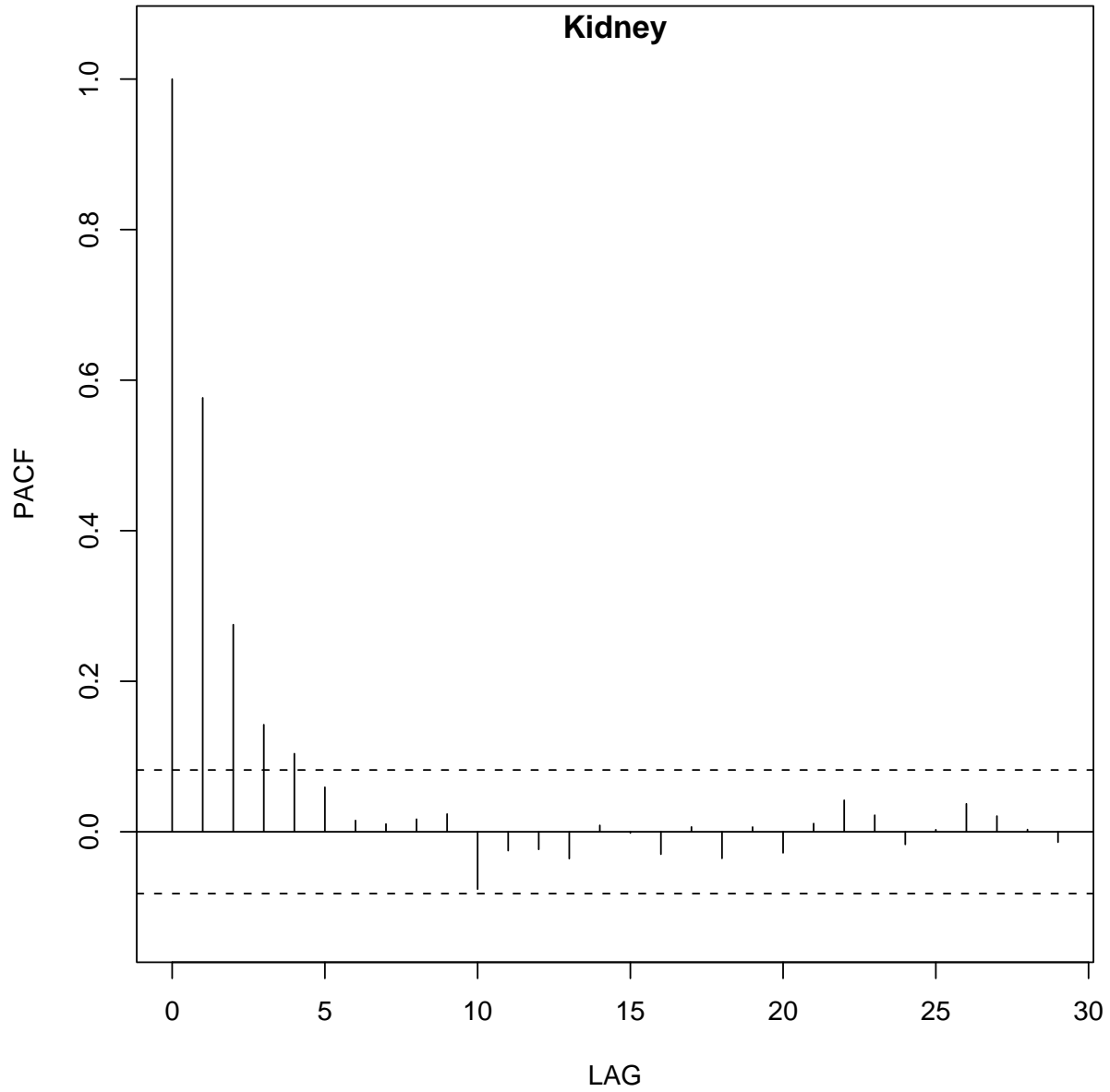
m ... Anzahl der medizinischen Variablen

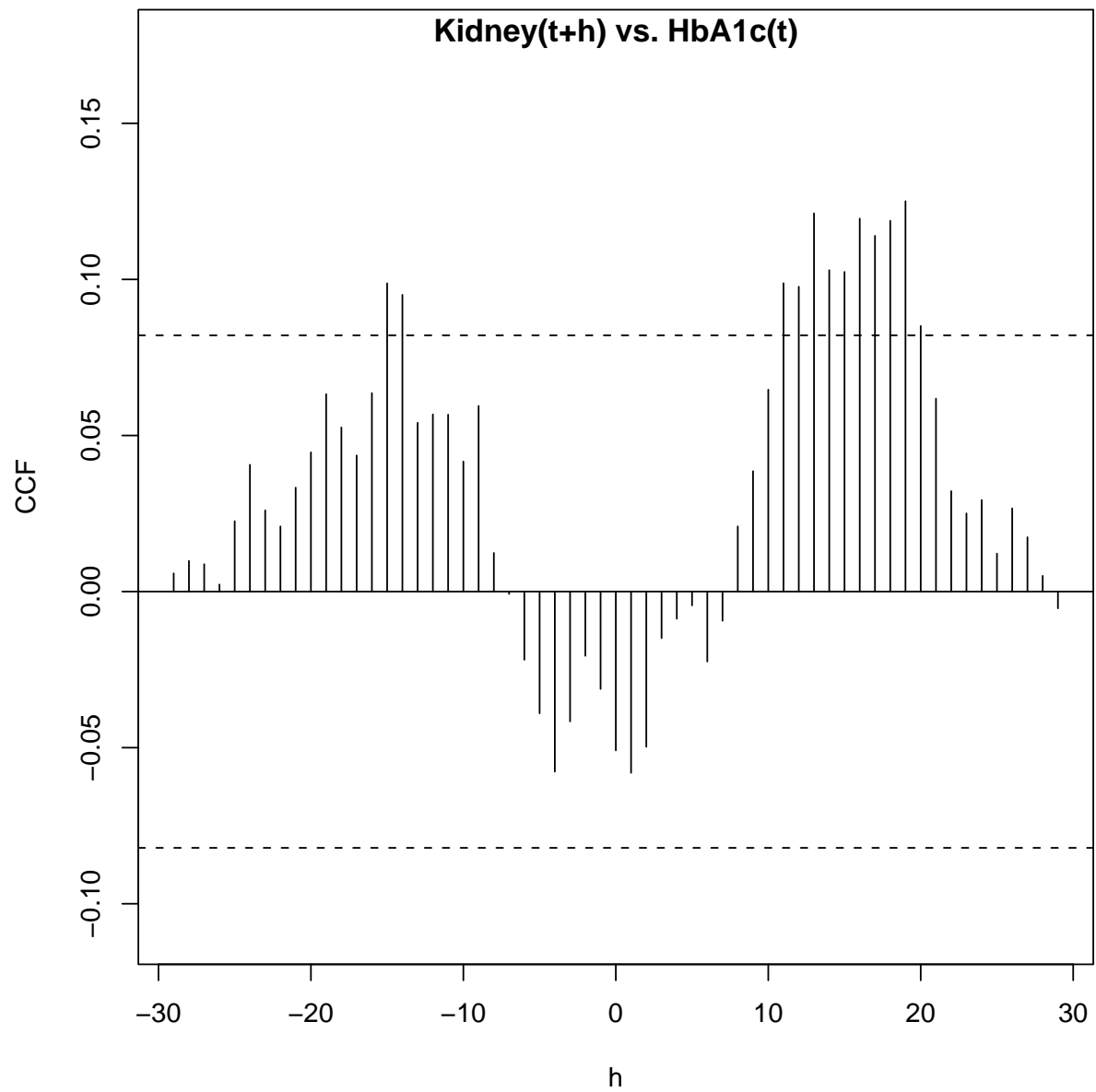
- **verwendete Variable:**

1	HbA1c	3	Triglyceride	5	mittlerer RR
2	Cholesterin	4	Mikroalbuminurie	6	bmi









Autoregressives Modell

- **AR(3)-Modell für die medizinische Variable Niere:**

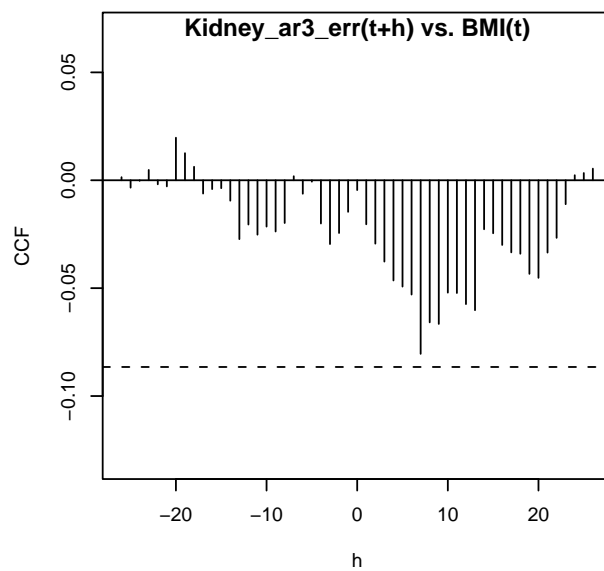
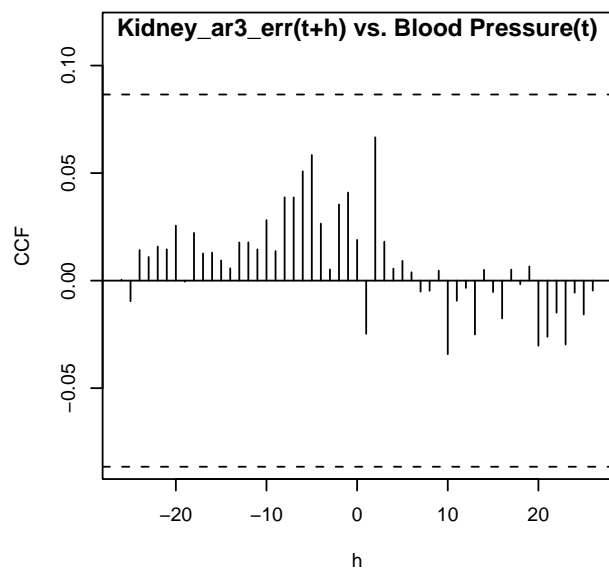
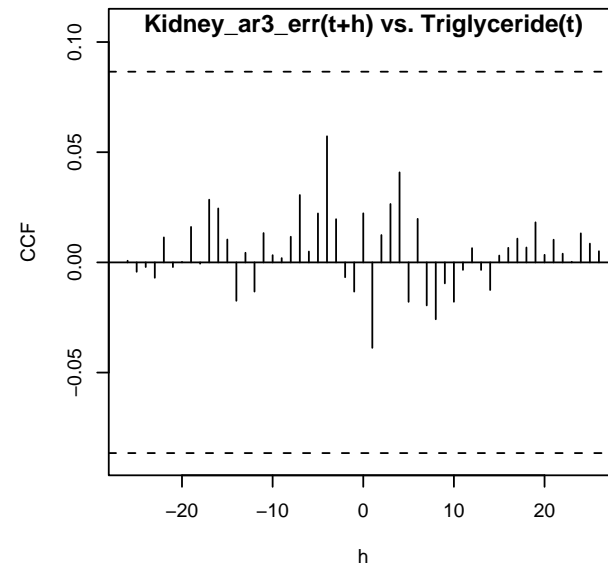
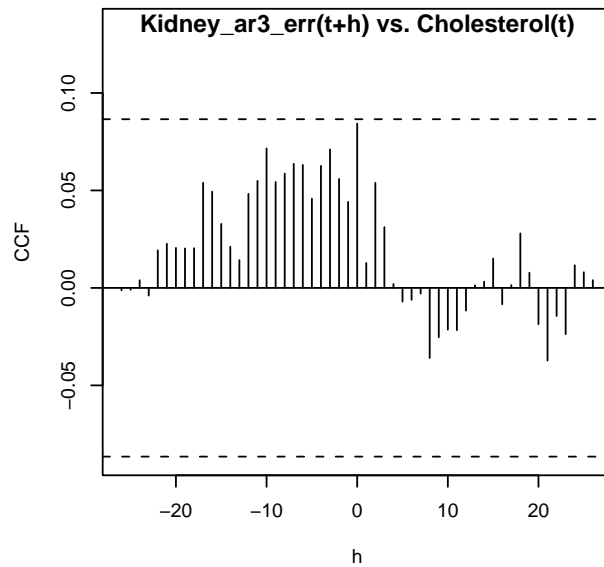
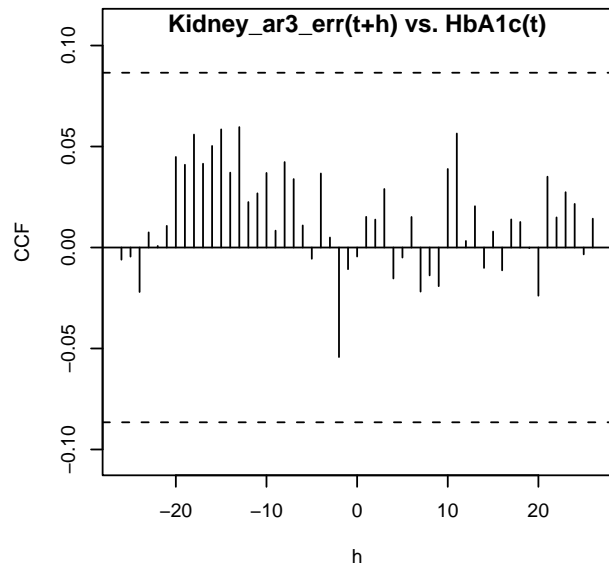
$$K_{tl} = \alpha + \phi_1 K_{t-1,l} + \phi_2 K_{t-2,l} + \phi_3 K_{t-3,l} + \omega_{tl} \quad , \quad (2)$$

mit $t = 4, \dots, 30$ und $l = 1, \dots, 19$.

- **Geschätztes Modell:**

$$\begin{aligned} \widehat{K}_t = & 0.156 + 0.337_{(0.043)} K_{t-1,l} + 0.220_{(0.044)} K_{t-2,l} \\ & + 0.174_{(0.043)} K_{t-3,l} + \widehat{\omega}_{tl} \quad , \end{aligned} \quad (3)$$

mit $\widehat{\sigma}_\omega^2 = 0.094$.



Autoregressives Modell mit exogenen Variablen

- **ARX-Modell für die medizinische Variable Niere:**

$$K_{tl} = \Gamma \mathbf{u}_{tl} + \phi_1 K_{t-1,l} + \phi_2 K_{t-2,l} + \phi_3 K_{t-3,l} + \omega_{tl} \quad , \quad (4)$$

mit $t = 12, \dots, 30$, $l = 1, \dots, 19$, $\Gamma = (\alpha, \beta)$ und $\mathbf{u}_{tl} = (1, H_{t-11,l})$.

- **Geschätztes Modell:**

$$\begin{aligned} \widehat{K}_t = & -0.200 + 0.057_{(0.025)} H_{t-11,l} + 0.348_{(0.051)} K_{t-1,l} \\ & + 0.179_{(0.051)} K_{t-2,l} + 0.161_{(0.049)} K_{t-3,l} + \widehat{\omega}_{tl} \quad , \quad (5) \end{aligned}$$

mit $\widehat{\sigma}_\omega^2 = 0.098$.

Ausblick

- Clusteranalyse, Diskriminanzanalyse
- weitere exogene Variablen (Geschlecht, Alter, ...)
- Modellierung anderer Spätfolgen (Retinopathie, Neuropathie)