

# Nichtparametrische Ansätze in der Zeitreihen- und Bildanalyse

## **AG Angewandte Mathematische Statistik**

Prof. Dr. Jürgen Franke, Dr. Jean-Pierre Stockis, Dr. Joseph Tadjuidje  
(FB Mathematik)

Dr. Gerald Kroisandt, Dr. Alex Sarishvili (ITWM)

Dr. Siana Halim (Petra University), Dr. Peter Mwita (Jomo Kenyatta University of  
Technology)

# 1. Allgemeine Regressions- und Zeitreihenmodelle

## Regressionsmodell:

$$Y_t = m(X_t) + \varepsilon_t, \quad t = 1, \dots, N$$

$Y_t \in \mathbb{R}, X_t \in \mathbb{R}^d, t=1, \dots, N$ , unabhängig

Residuen  $\varepsilon_t, t=1, \dots, N$ , unabhängig, identisch verteilt,  $E\varepsilon_t = 0$

## Autoregressionsmodell der Ordnung $p$ :

$$Y_t = m(Y_{t-1}, \dots, Y_{t-p}) + \varepsilon_t, \quad t = 1, \dots, N$$

$$X_t = (Y_{t-1}, \dots, Y_{t-p})$$

## Modell für verrauschtes Bild:

$$Y_z = m(z) + \varepsilon_z, \quad z \in I_n \subset \mathbb{Z}^2$$

stochastischer Prozess auf ganzzahligem Gitter  $\mathbb{Z}^2$

Regressionsmodell: 
$$Y_t = m(X_t) + \varepsilon_t$$

### Ziel I:

**Schätze** Funktion  $m(x)$  ( $= E\{Y_t | X_t = x\}$ ) aus den Daten ohne spezielle Annahmen an die Form von  $m$  (außer allgemeinen Regularitätsannahmen, z.B.  $m \in C^2, L^2(\mu), \dots$ )

Unter dieses allgemeine Schätzproblem fällt:

- **Vorhersage** von Zeitreihen aus den vergangenen Beobachtungen
- **Entrauschen** von Bildern
- **Risikomanagement** von Zeitreihen

Verallgemeinerungen:

- 1)  $Y_t = m(X_t) + \sigma(X_t)\varepsilon_t$       Heteroskedastizität
- 2) Residuen  $\varepsilon_t$  stationärer stochastischer Prozess
- 3) ARMA oder Feedback:  $Y_t = m(Y_{t-1}, \varepsilon_{t-1}) + \varepsilon_t$

Regressionsmodell:  $Y_t = m(X_t) + \varepsilon_t, \quad Y_t^* = m^*(X_t^*) + \varepsilon_t^*$

## Ziel II:

**Teste**, ob  $m=m^*$  auf der Grundlage nichtparametrischer Schätzer

## Anwendungen:

- Veränderungen in Bildern incl. Lokalisation
- Strukturbrüche in Zeitreihen
- Klassifikation anhand von unabhängigen Zeitreihenbeobachtungen

Beispiel: Detektion und Lokalisation von Defekten in textilen Geweben

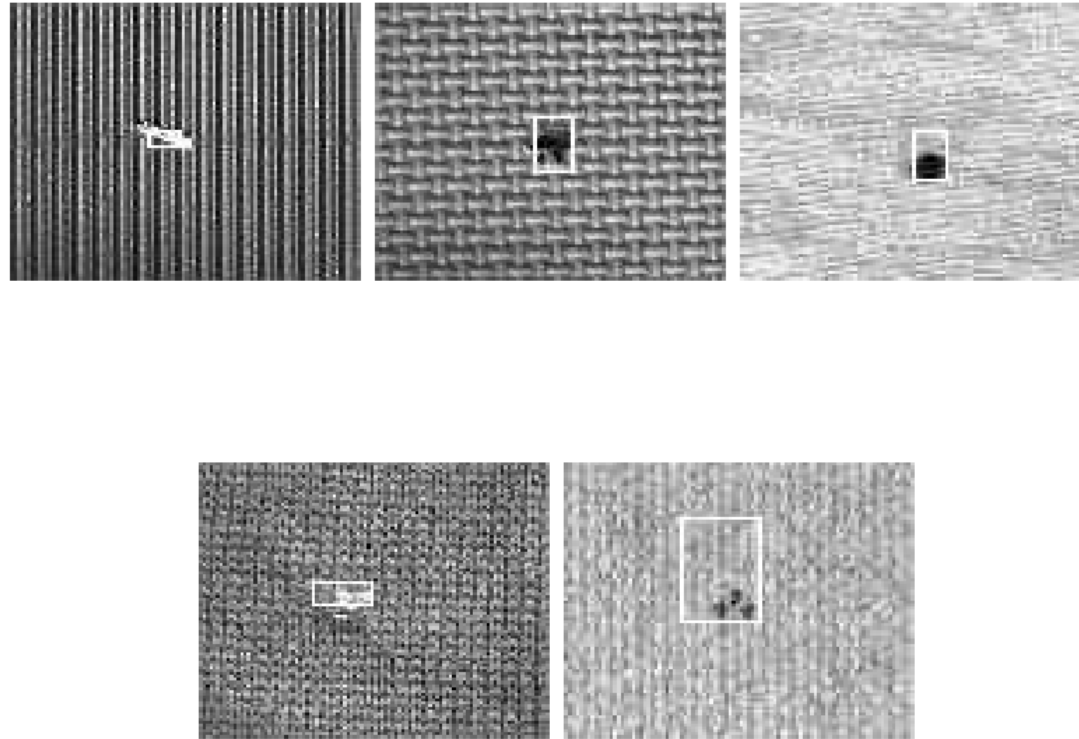


Figure 5.3: Some examples of defect detection

## 2. Lokale Glättungsverfahren

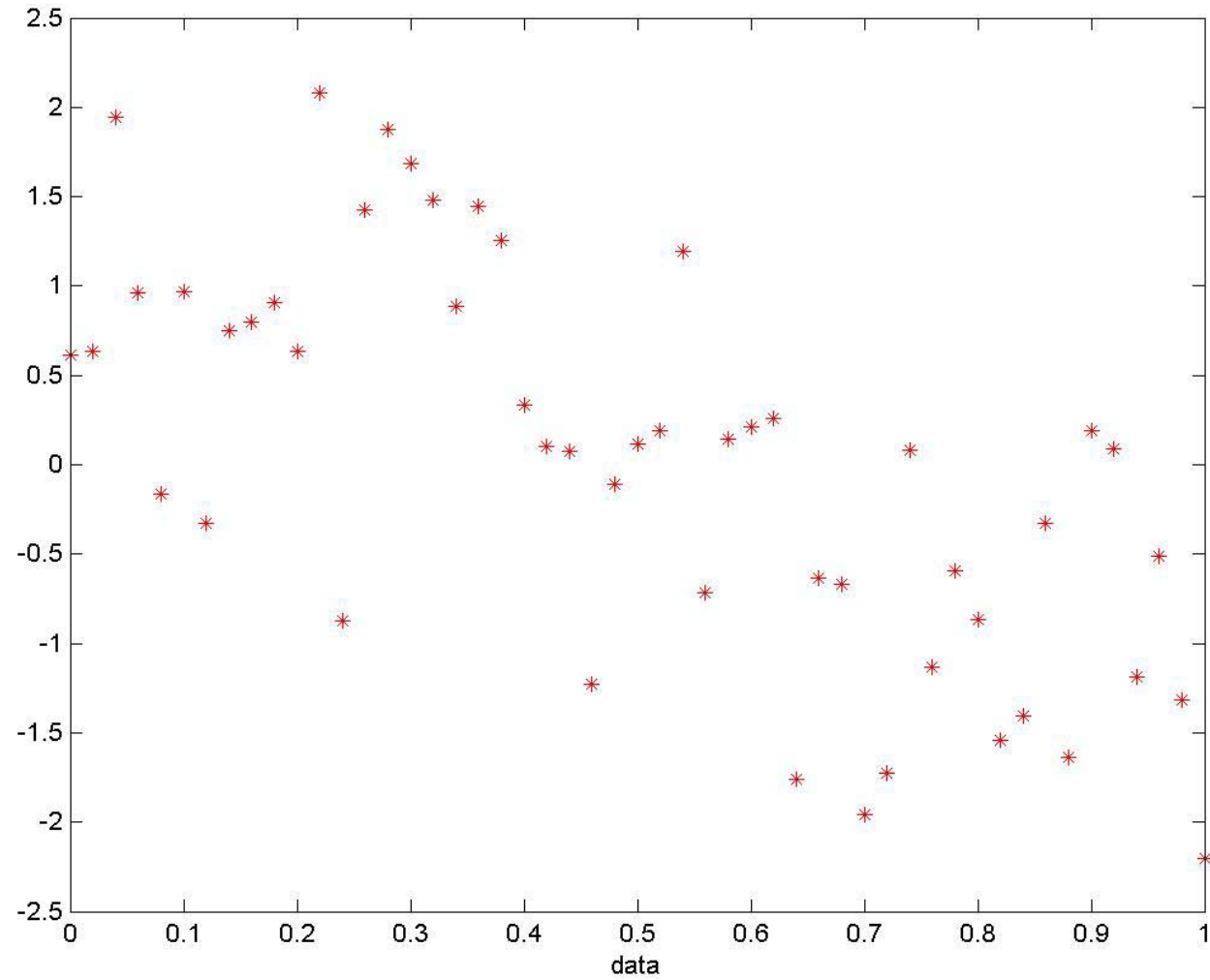
Regressionsmodell:  $Y_t = m(X_t) + \varepsilon_t, t = 1, \dots, N$

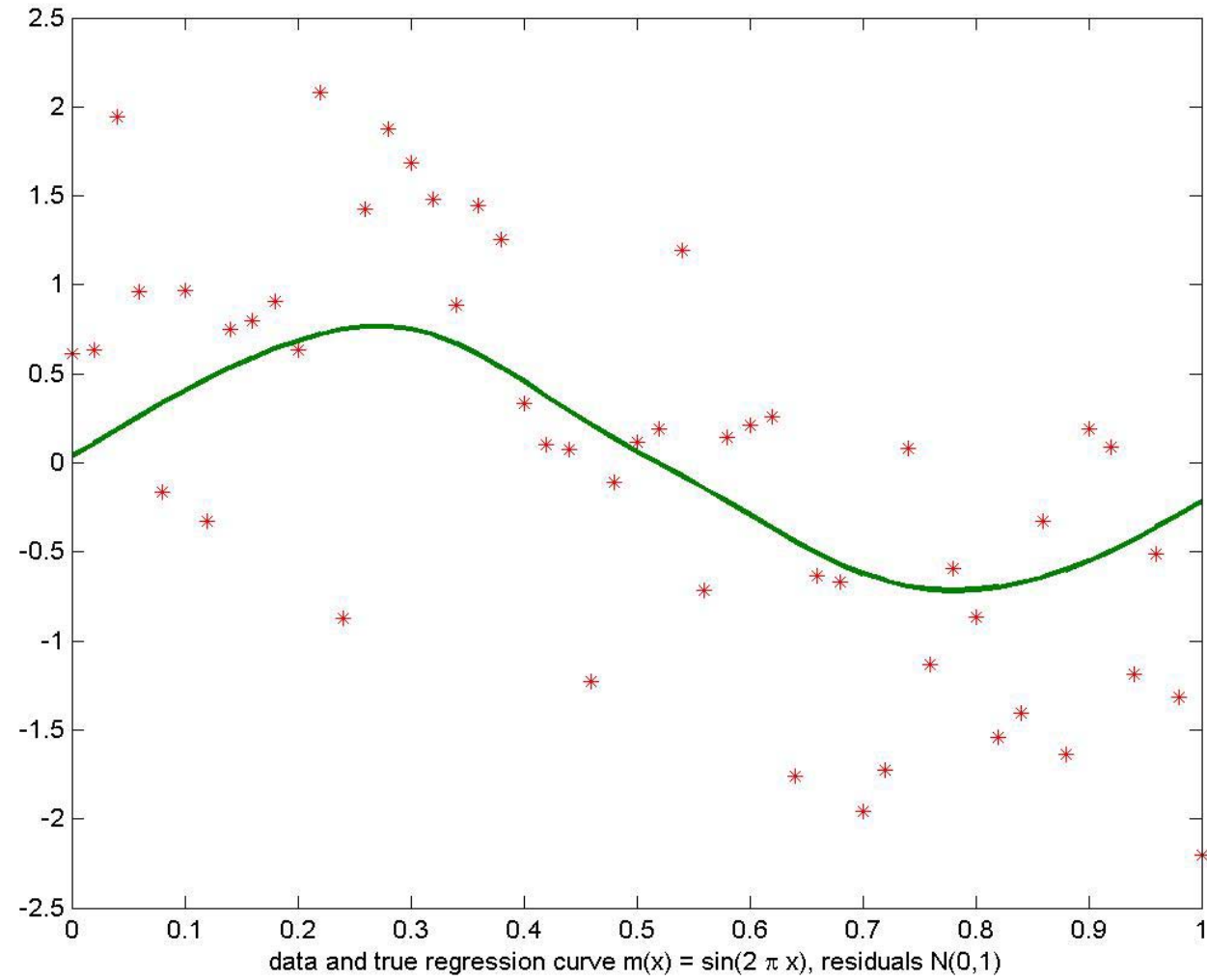
$X_t$  äquidistantes Gitter auf Intervall  $[0,1]$ :

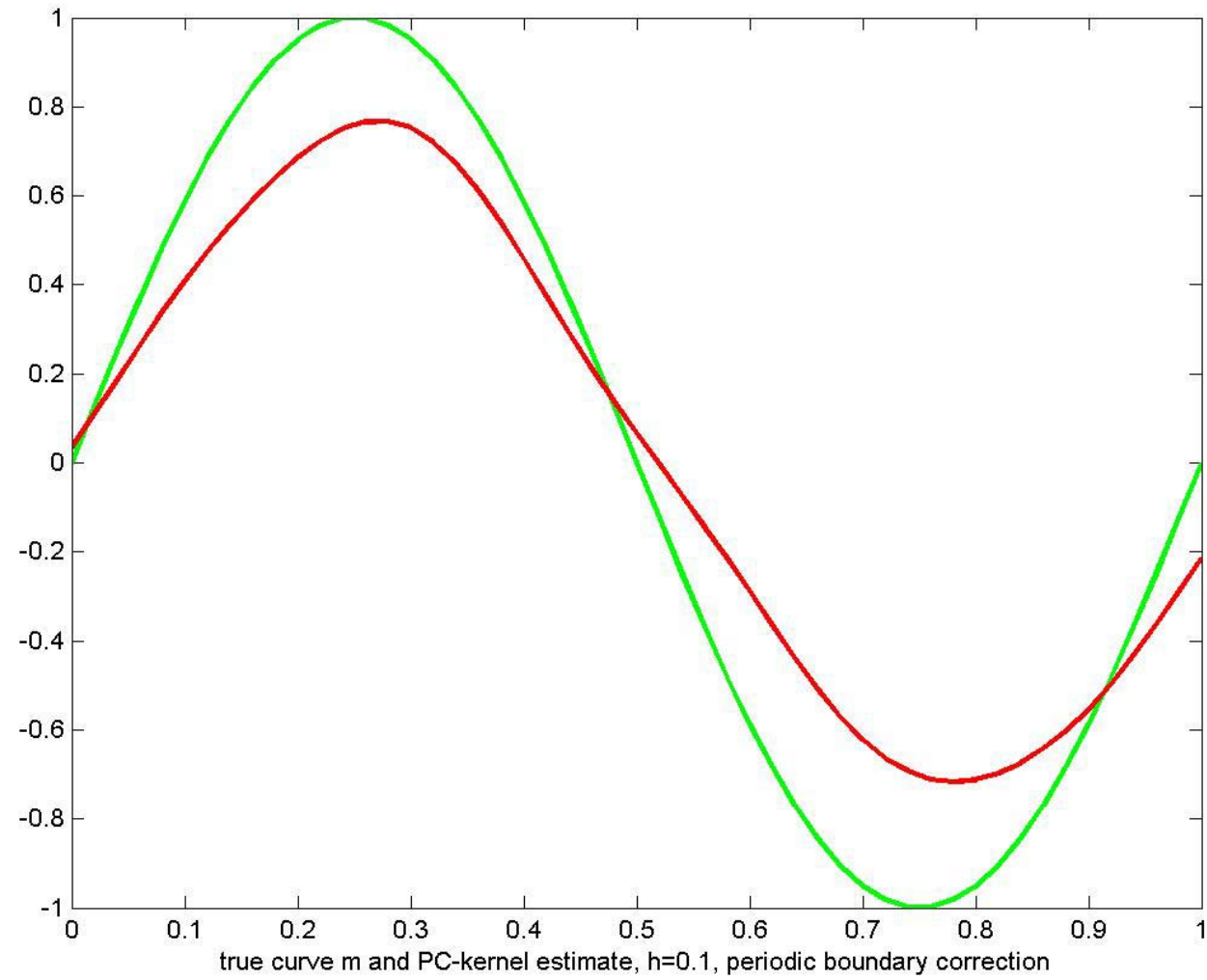
$$\hat{m}(x, h) = \frac{1}{N} \sum_t K_h(x - X_t) Y_t, \quad K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$$

lokales gewichtetes Mittel der Daten:  $K_h(x - X_t) \approx 0$  für  $|x - X_t| > h$

Intuition:  $m$  glatt  $\rightarrow m(u) \approx m(x)$  für alle  $u \approx x \rightarrow EY_t \approx m(x)$  für alle  $t$   
mit  $X_t \approx x \rightarrow \hat{m}(x, h) \rightarrow m(x)$  für  $N \rightarrow \infty, h \rightarrow 0$   
(Gesetz der großen Zahlen)







Allgemein Korrektur für variierende lokale Dichte der  $X_t$  nötig.

**Nadaraya-Watson Kernschätzer:**  $\hat{m}(x, h) = \frac{\sum_t K_h(x - X_t) Y_t}{\sum_t K_h(x - X_t)}$

**Lokal-konstanter Kleinster-Quadrate-Schätzer:**

$$\hat{m}(x, h) = \arg \min_m \sum_t (Y_t - m)^2 K_h(x - X_t)$$

**Lokal-linearer Schätzer**

$$\min_{m_0, m_1} \sum_t (Y_t - m_0 - m_1(x - X_t))^2 K_h(x - X_t)$$

$$m_0 = \hat{m}(x, h) \rightarrow m(x), \quad m_1 \rightarrow m'(x)$$

$$\hat{m}(x, h) = \arg \min_m \sum_t (Y_t - m)^2 K_h(x - X_t)$$

Kleinste-Quadrate: Schätzer für **bedingten Erwartungswert**, optimal (da lokal Maximum Likelihood für **normalverteilte Residuen**)

Alternative Verlustfunktionen:

$|Y_t - m|_\alpha \rightarrow$  Lösung schätzt **bedingtes  $\alpha$ -Quantil**  $q_\alpha(x)$ , i.e.

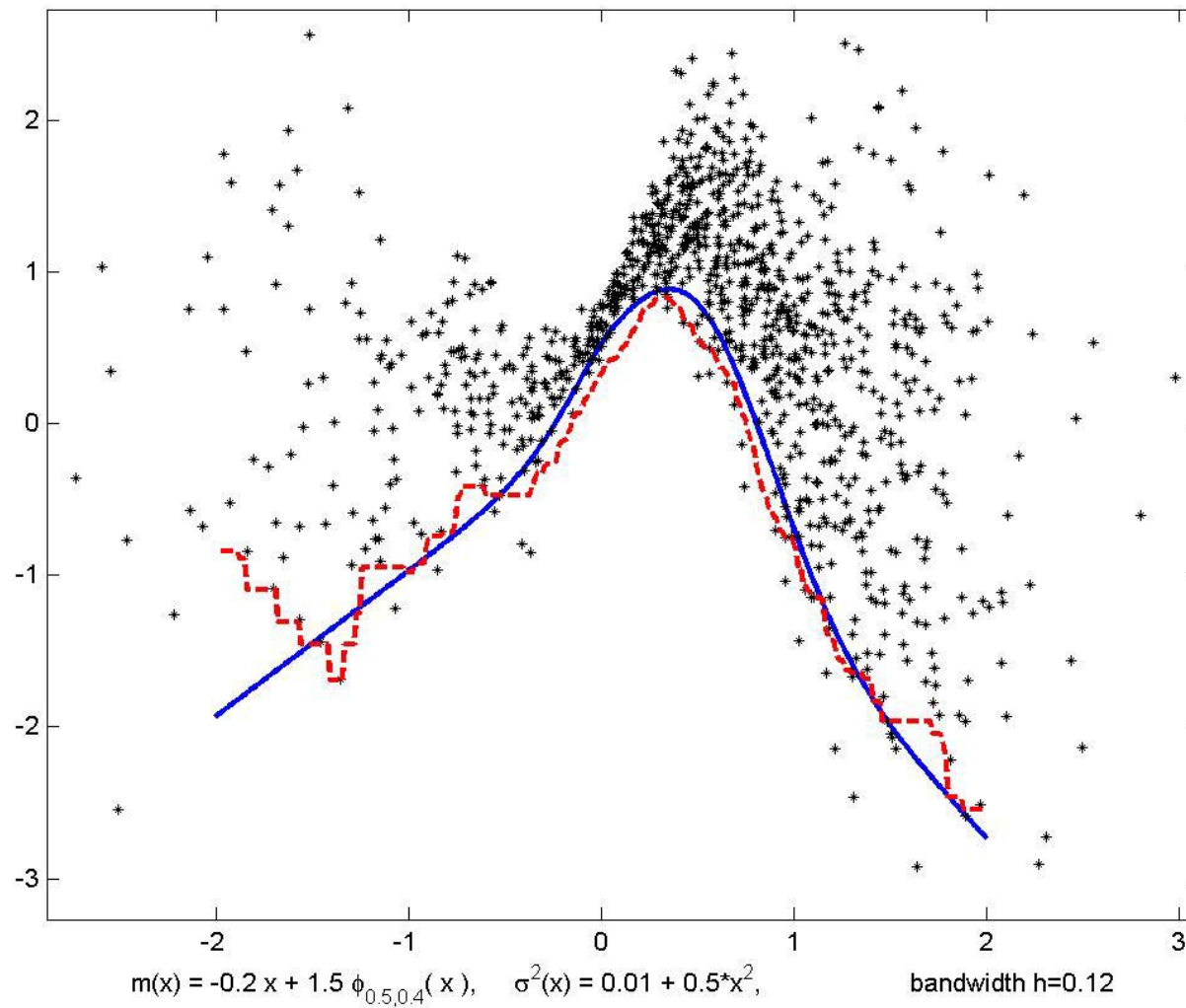
$$\text{pr}(Y_t \leq q_\alpha(x) | X_t = x) = \alpha$$

Beispiel:  $Y_{t+1} = m(Y_t) + \sigma(Y_t)\varepsilon_{t+1}$  (AR-ARCH)

$$m(r) = a + br + \frac{1}{\sqrt{2\pi d}} \exp\left(-\frac{(r-c)^2}{d^2}\right), \quad \sigma^2(r) = \omega + \phi r^2$$

$N=1000$ ,  $\varepsilon_t$   $N(0,1)$ -verteilt

Conditional 0.05-quantile estimation for NLAR(1)-ARCH(1)-process with Gaussian innovations,  $N=1000$ , incl. scatter plot ( $X_t$ ,



## Lokale Maximum-Likelihood Schätzer

$$\hat{m}(x, h) = \arg \max_m \sum_t l_t(Y_t, m) K_h(x - X_t)$$

$l_t(Y_t, m)$  bedingte log-Likelihood von  $Y_t$  gegeben  $X_t$ , Funktion von Parameter  $m$

Theorie:

- **Asymptotische Verteilung** von  $\hat{m}(x, h)$  für  $N \rightarrow \infty, h \rightarrow 0$
- Alternative: **Bootstrap-Approximation** der Verteilung von  $\hat{m}(x, h)$
- **Konfidenzbänder** für  $m(x)$ :  
 $\text{pr}(\hat{m}_-(x, h) \leq m(x) \leq \hat{m}_+(x, h) \text{ für alle } x \in [a, b]) \geq \gamma$
- automatische **Wahl des Glättungsparameters**: Kreuzvalidierung, Plug-in- oder Bootstrap-Approximation von  $e(h) = E d(\hat{m}(\cdot, h), m)$  und Minimierung

- automatische **lokale** Wahl des Glättungsparameters  $h=h(x)$ , so dass z.B.  
 $e(x, h) = E (\hat{m}(x, h), m(x))^2 = \min!$
- **Tests** von  $H_0: m=m^*$ , typischerweise von der Form  $d(\hat{m}, \hat{m}^*) \geq c_\alpha \rightarrow$   
ablehnen
- **Tests parametrischer Modelle**:  $H_0: m(x) = g(x; \theta)$  für ein  $\theta \in \Theta$

Beziehung zu anderen lokalen Verfahren:

**Spline-Glätter, nichtlineare Diffusionsfilter, Regularisierungsverfahren, ...**

(mit J. Weickert, Saarbrücken)

### 3. Globale nichtparametrische Funktionsschätzer

Regressionsmodell:  $Y_t = m(X_t) + \varepsilon_t, t = 1, \dots, N$

Parametrisierte **Funktionsklassen**  $F_N$ , deren Dimension mit  $N$  wächst, mit universeller Approximationseigenschaft, z.B.  $F_1 \cup F_2 \cup \dots$  dicht in  $L^2(\mu)$

**Siebschätzer**  $\hat{m}_N = \arg \min_{f \in F_N} \sum_{t=1}^N (Y_t - f(X_t))^2$

Wenn die Parameterdimension von  $F_N$  mit der richtigen Rate wächst (abhängig von  $N$ , der „Größe“ der Funktionsklassen und den stochastischen Eigenschaften von  $(X_t, \varepsilon_t), t = 1, \dots, N$ ) :

$$\|\hat{m}_N - m\|_2^2 = \int (\hat{m}_N(x) - m(x))^2 \mu(dx) \rightarrow 0 \quad \text{fast sicher}$$

Regressionsmodell:  $Y_t = m(X_t) + \varepsilon_t, t = 1, \dots, N$

Unterschied zu lokalen Glättungsverfahren:

- $\hat{m}(x, h)$  hängt im wesentlichen nur von den Daten  $(Y_t, X_t)$  mit  $X_t \approx x$  ab
- $\hat{m}_N(x)$  hängt von allen Daten ab

Beispiele:

**Orthogonalreihen-** (z.B. Fourier-) **Entwicklungen** mit  $M=M(N)$ :

$$F_N = \{f(x) = \sum_{k=-M}^M a_k e^{ikx}; a_0, \dots, a_M \in \mathbb{R}, a_{-k} = a_k\}$$

**Neuronale Netze**, z.B. Feedforward mit einer verborgenen Schicht aus  $H=H(N)$  Neuronen und logistisch-sigmoider Aktivierungsfunktion  $\psi(u) = (1 + e^{-u})^{-1}$  :

$$f \in F_N, \quad \text{falls } f(x; \theta) = v_0 + \sum_{h=1}^H v_h \psi(w_h^{(0)} + w_h^T x), \quad \sum_{h=0}^H |v_h| \leq \Delta$$

$\Delta = \Delta_N$  mit  $N$  wachsende Schranke an die Output-Gewichte,

$\theta \in R^{(d+1)H+H+1}$  Parametervektor aus allen Gewichten  $v_h, w_h^{(0)} \in R, w_h \in R^d$

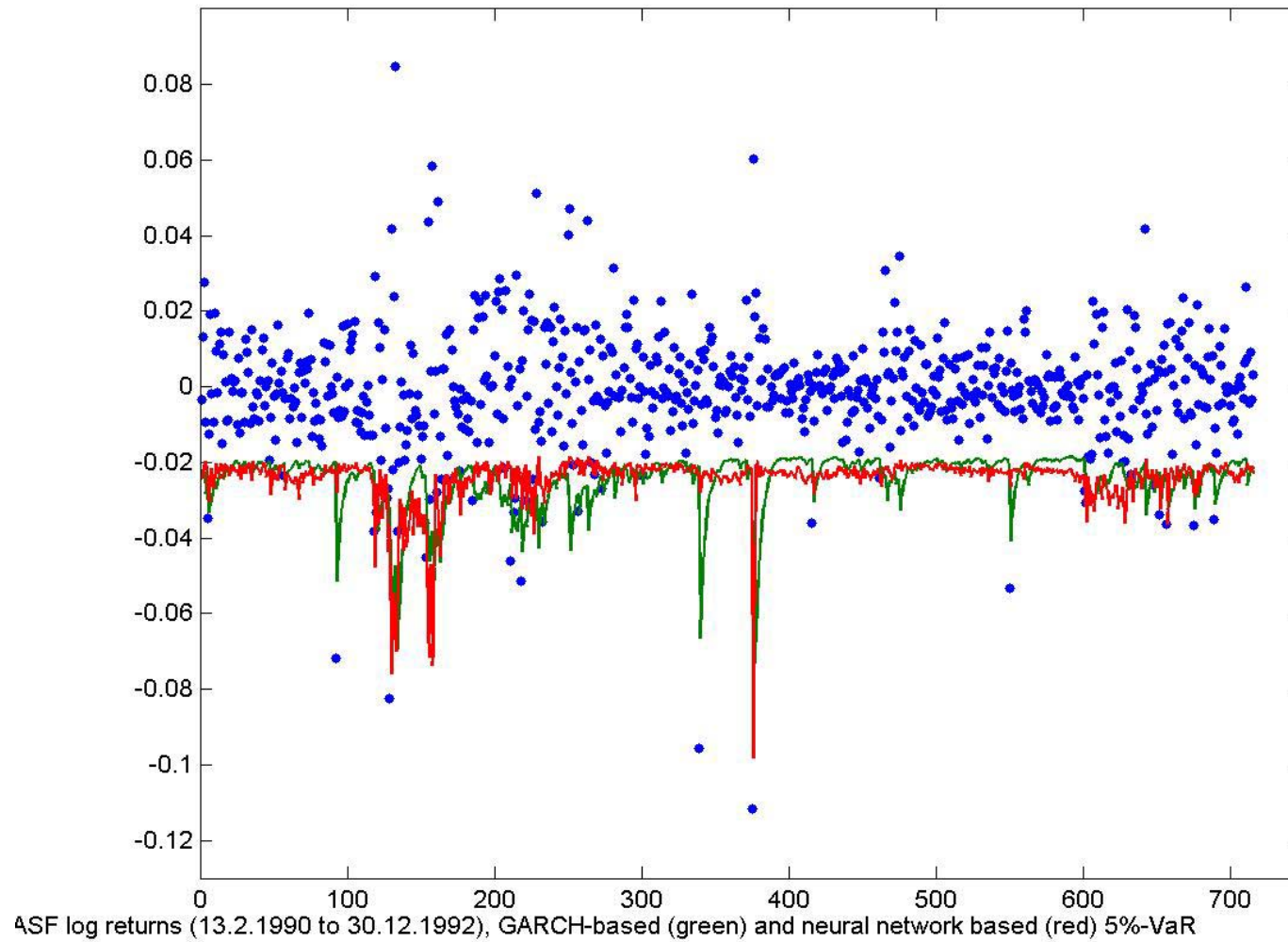
Alternative Verlustfunktionen: Quantilsschätzer, Varianzschätzer etc.

Beispiel:

$Y_{t+1} = \sigma(X_t)\varepsilon_t$	(ARCHX)
$Y_{t+1}$	Rendite der BASF-Aktie (1990-1992)
$X_{t1} = Y_t$	
$X_{t2}$	Rendite des DAX
$X_{t3}$	30 Tage-Mittel der DAX-Renditen
$X_{t4}$	exponentiell gewichtete 30 Tage Stichprobenstandardabweichung (Volatilität) von BASF

Vergleiche Value-at-Risk ( $\alpha = 5\%$ -Quantil) geschätzt über

- GARCH(1,1)-Modell
- netzbasierter Schätzer für  $\sigma(x)$  mit standard-normalverteilten Innovationen



## 4. Mischungen stationärer Zeitreihen

### Detektion von Strukturänderungen

Beobachtete Zeitreihe

$$Z_t = \mu_t + Y_t$$

$$EY_t = 0$$

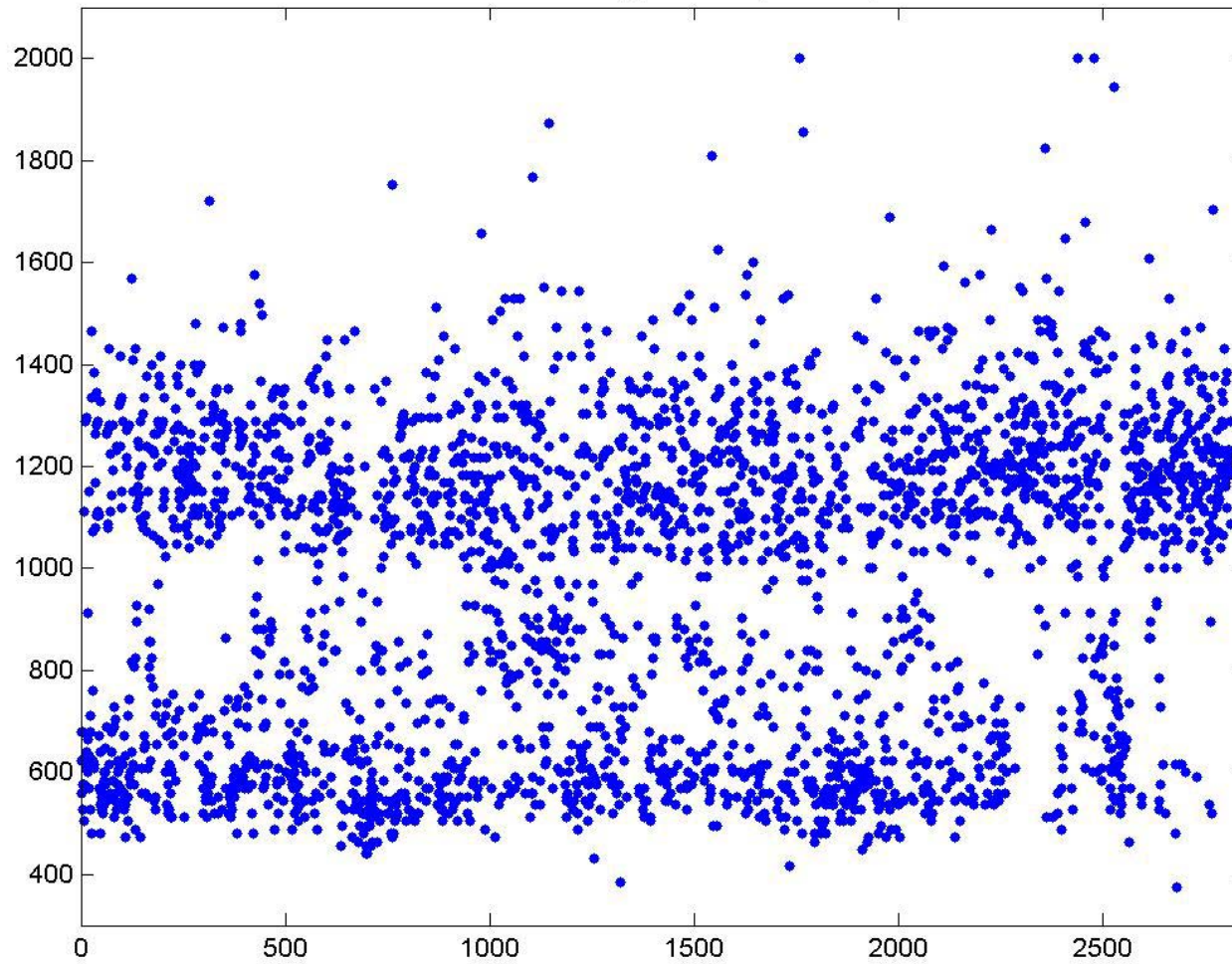
Stationarität:  $\mu_t \equiv \mu$ ,  $Y_t$  stationär

Verschiedene Phasen:

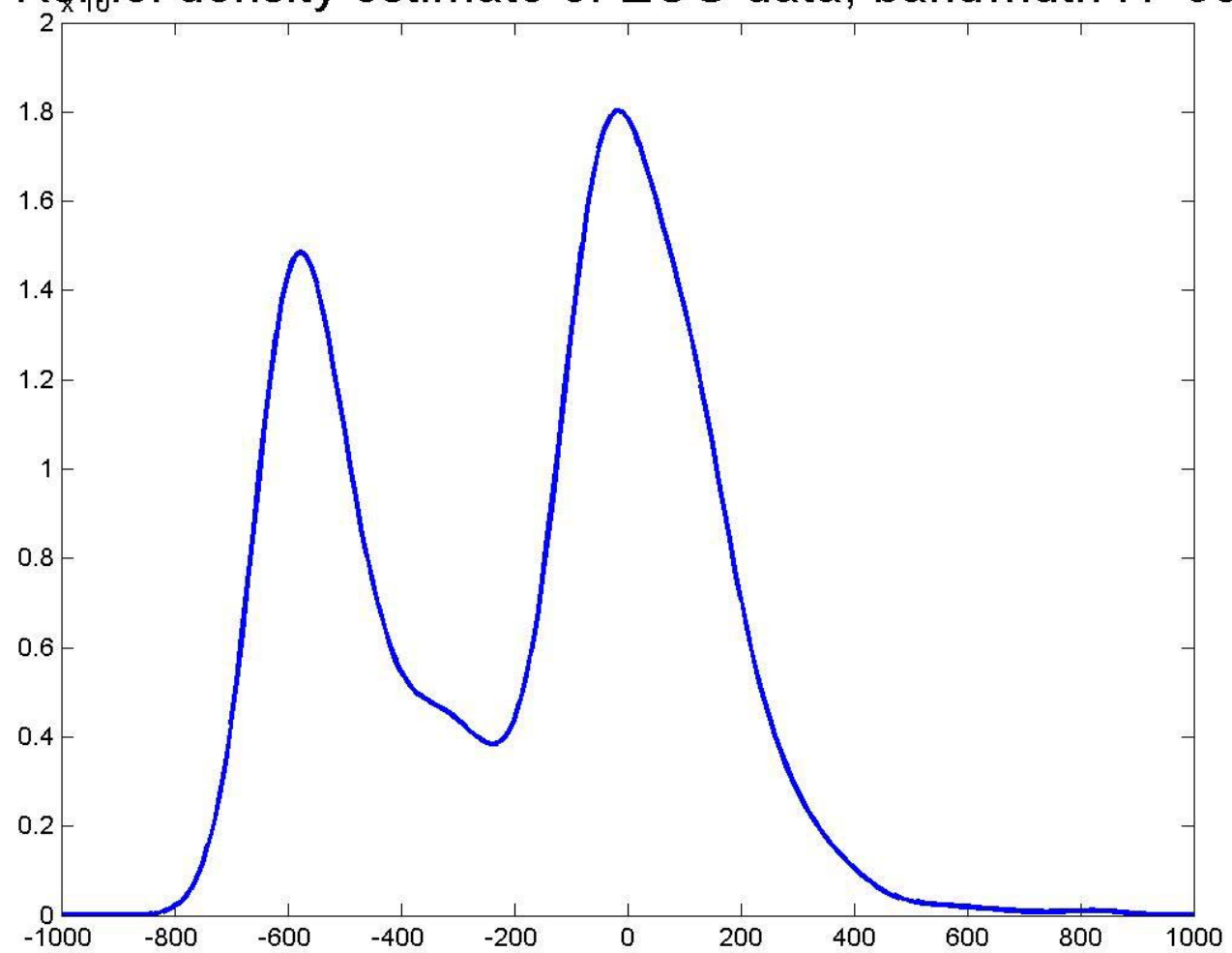
- zeitabhängiges Mittel (Sprünge, beginnende Trends)
- Änderungen in der stochastischen Struktur von  $Y_t$  (Modellparameter, Spektralzerlegung, Kovarianzstruktur)

Anzeichen: multimodale Dichte, verschiedene Cluster im Scatterplot  $(Z_t, Z_{t+1})$

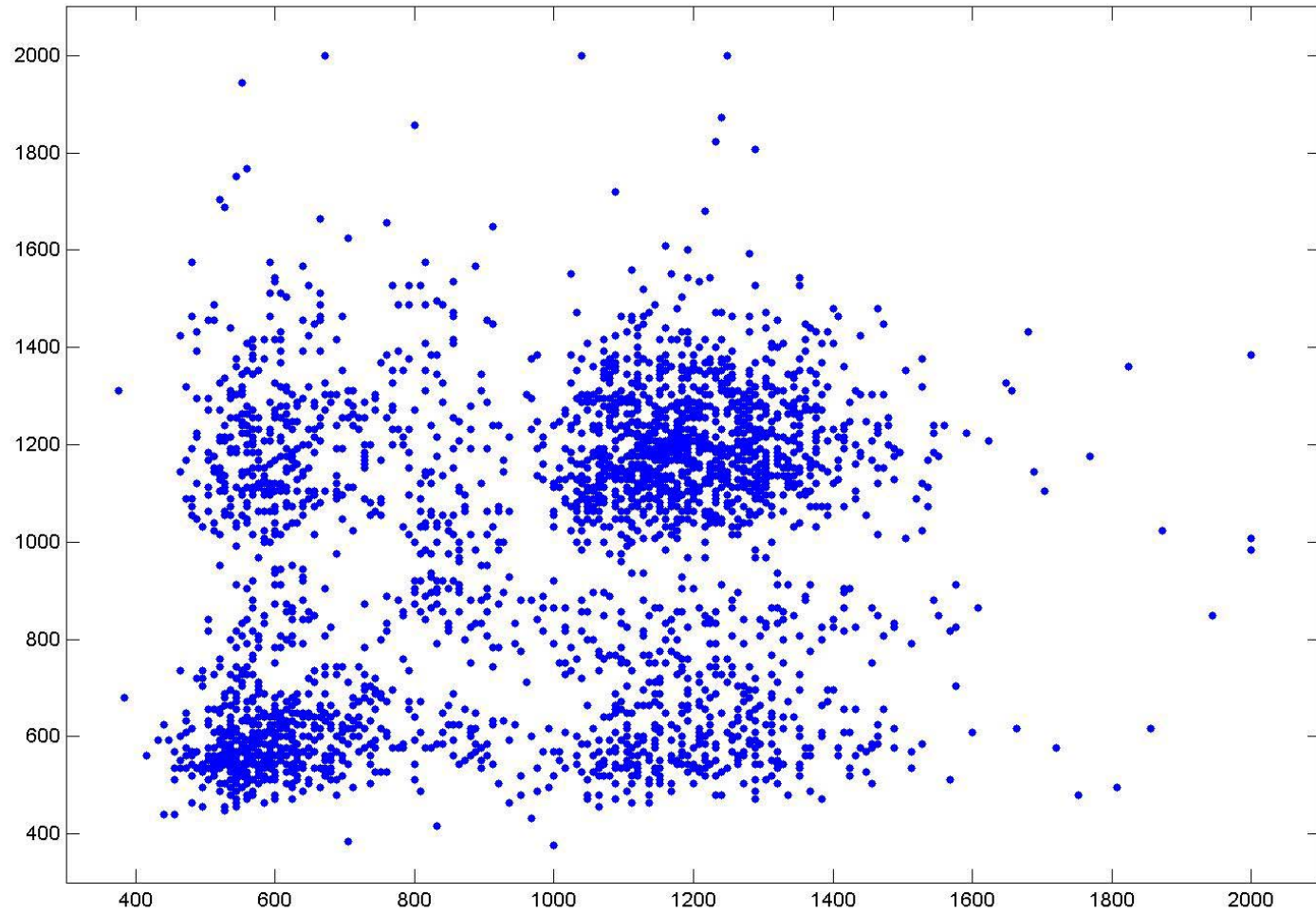
### Electrocardiogram (ECG) data



Kernel density estimate of ECG data; bandwidth  $H=50$



ECG data - scatterplot ( $Z_t, Z_{t+1}$ )



**Der Phasen- oder Zustandsprozess:**

$K$  Phasen oder Zustände, festgelegt durch verborgenen **Zustandsprozess**  $Q_t$ ,

z.B. Markowkette mit Zustandsraum  $\{1, \dots, K\}$ . Alternative Darstellung:

**Zustandsindikator**  $S_t \in \{e_1, \dots, e_K\}$ ,  $e_j$  Basisvektoren des  $R^K$ , mit

$$S_{t,k} = \begin{cases} 1 & \text{if } Q_t = k \\ 0 & \text{else} \end{cases}, \quad k = 1, \dots, K.$$

**Modell (Mischung autoregressiver Prozesse):**

Ordnung 1:

$$Z_t = \sum_{k=1}^K S_{t,k} \{m_k(Z_{t-1}) + \sigma_k \varepsilon_t\}$$

Residuen  $\varepsilon_t \sim N(0,1)$ , Residuenvarianzen  $\sigma_k^2 > 0$ ,  $m_k(z)$  autoregressive Funktionen

Allgemein:

**CHARME** (conditional heteroscedastic autoregressive mixtures of experts):

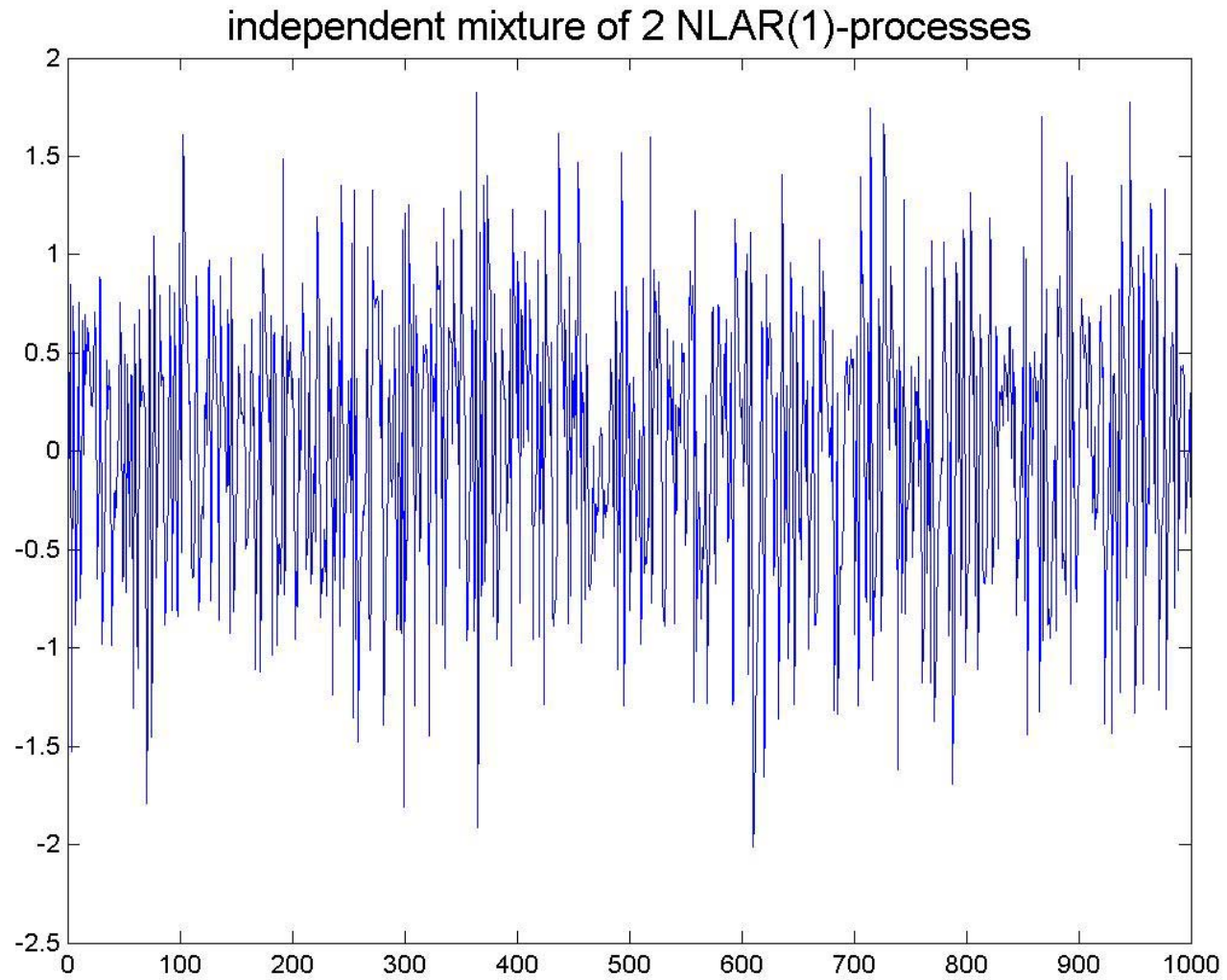
$$Z_t = \sum_{k=1}^K S_{t,k} \left\{ m_k(Z_{t-1}, \dots, Z_{t-p}) + \sigma_k(Z_{t-1}, \dots, Z_{t-p}) \varepsilon_t \right\}$$

$pr(S_t \neq S_{t+1}) \leq p \ll 1$  , so dass Zustandsänderungen nicht zu oft auftreten

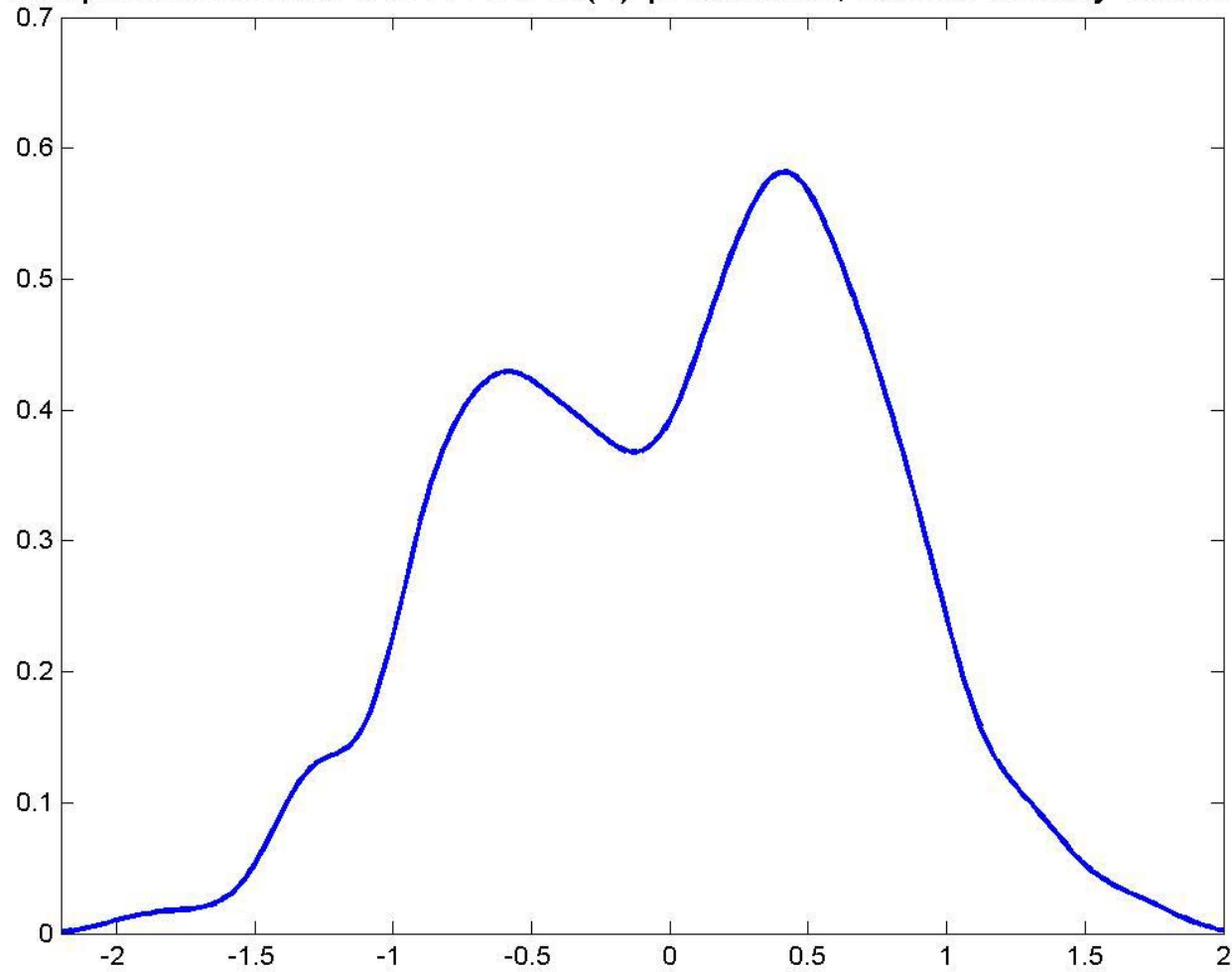
**Ziel:**

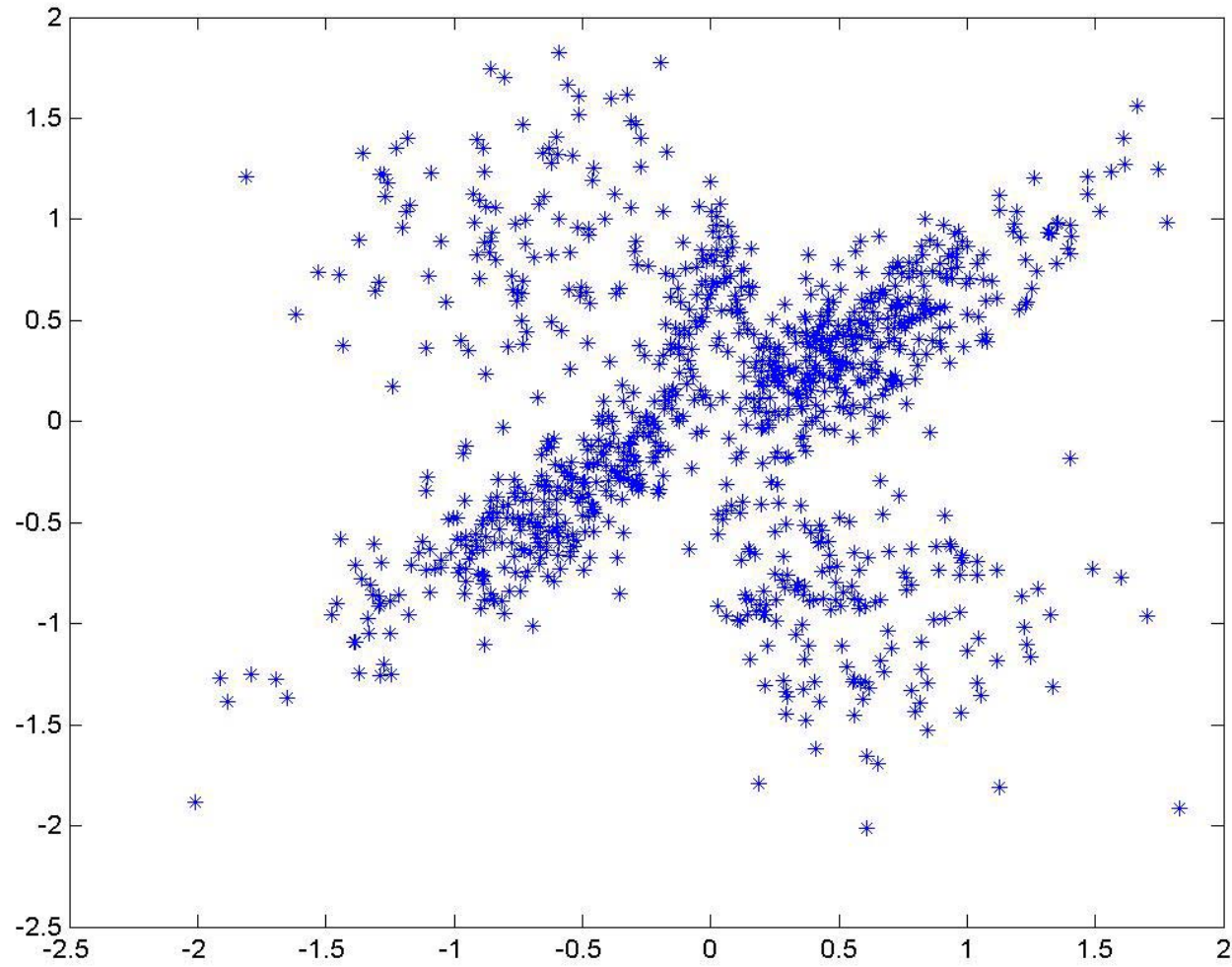
- Simultane nichtparametrische Schätzung (z.B. mit Nadaraya-Watson-Kernschätzer oder neuronalen Netzen) der  $m_k(z), \sigma_k^2(z), k = 1, \dots, K$ .
- Schätzen der Zustandsindikatoren  $S_t$  und damit auch von Zustandsänderungen

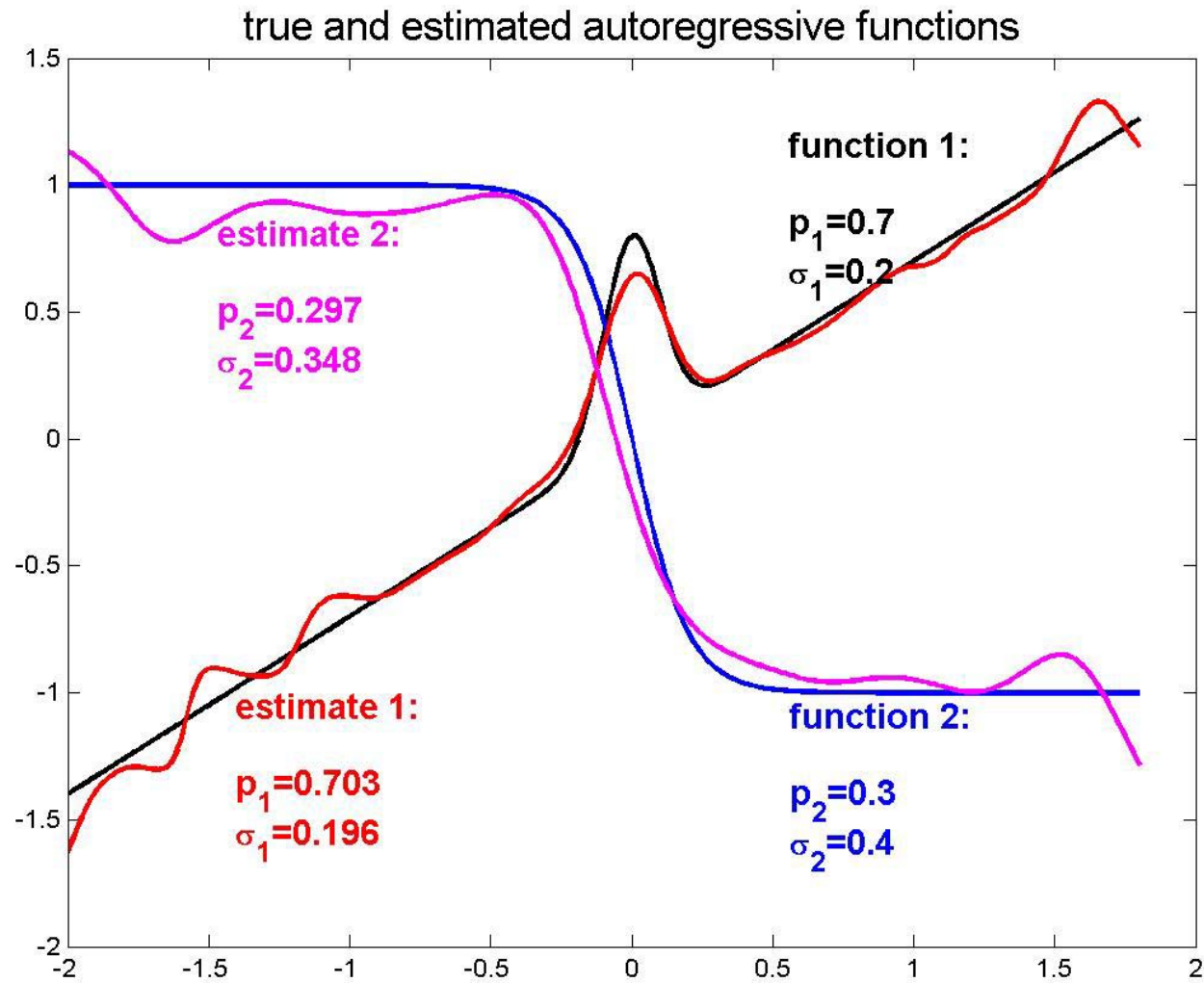
**Expectation-Maximization- (EM) Algorithmen**



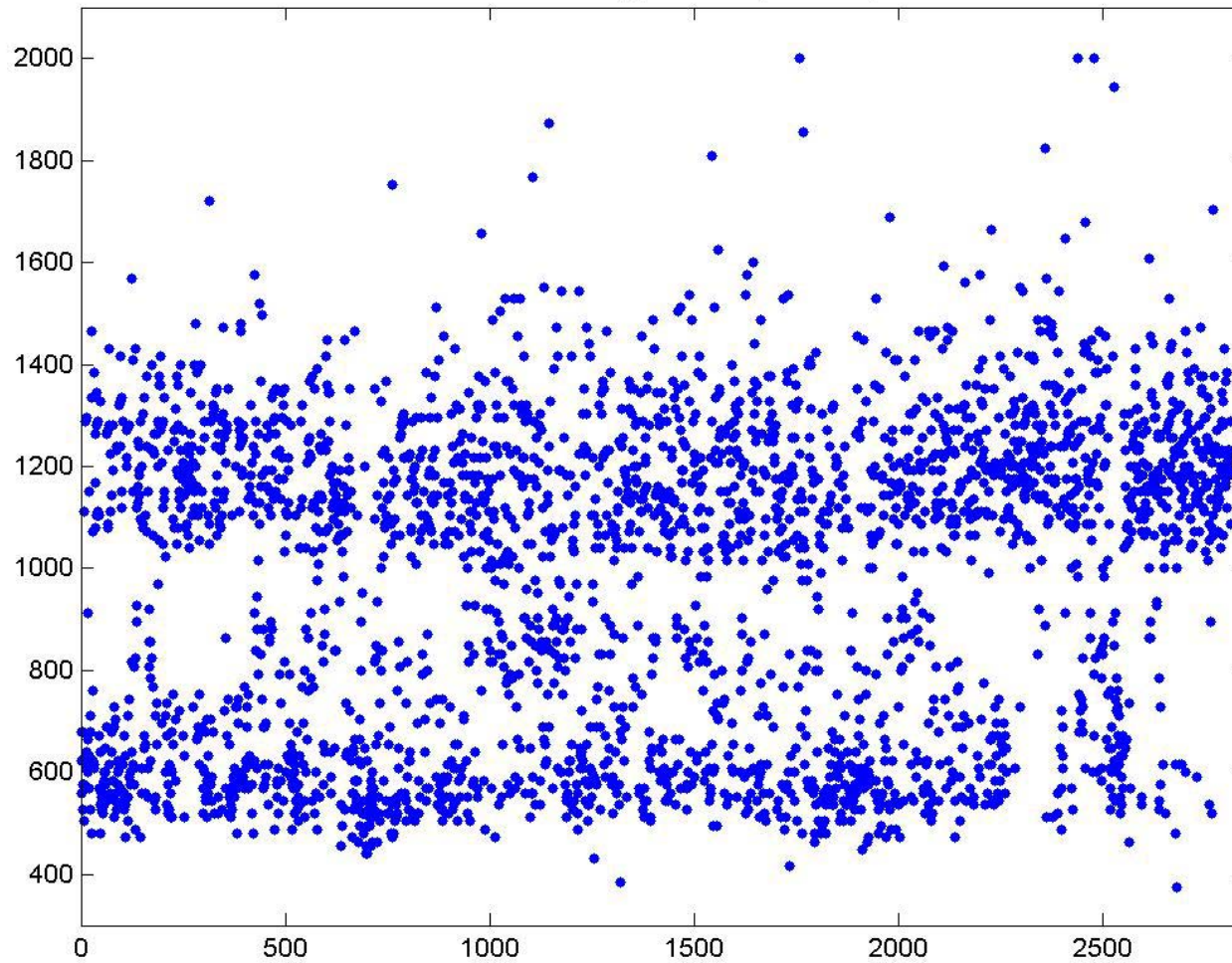
independent mixture of 2 NLAR(1)-processes, kernel density estimate



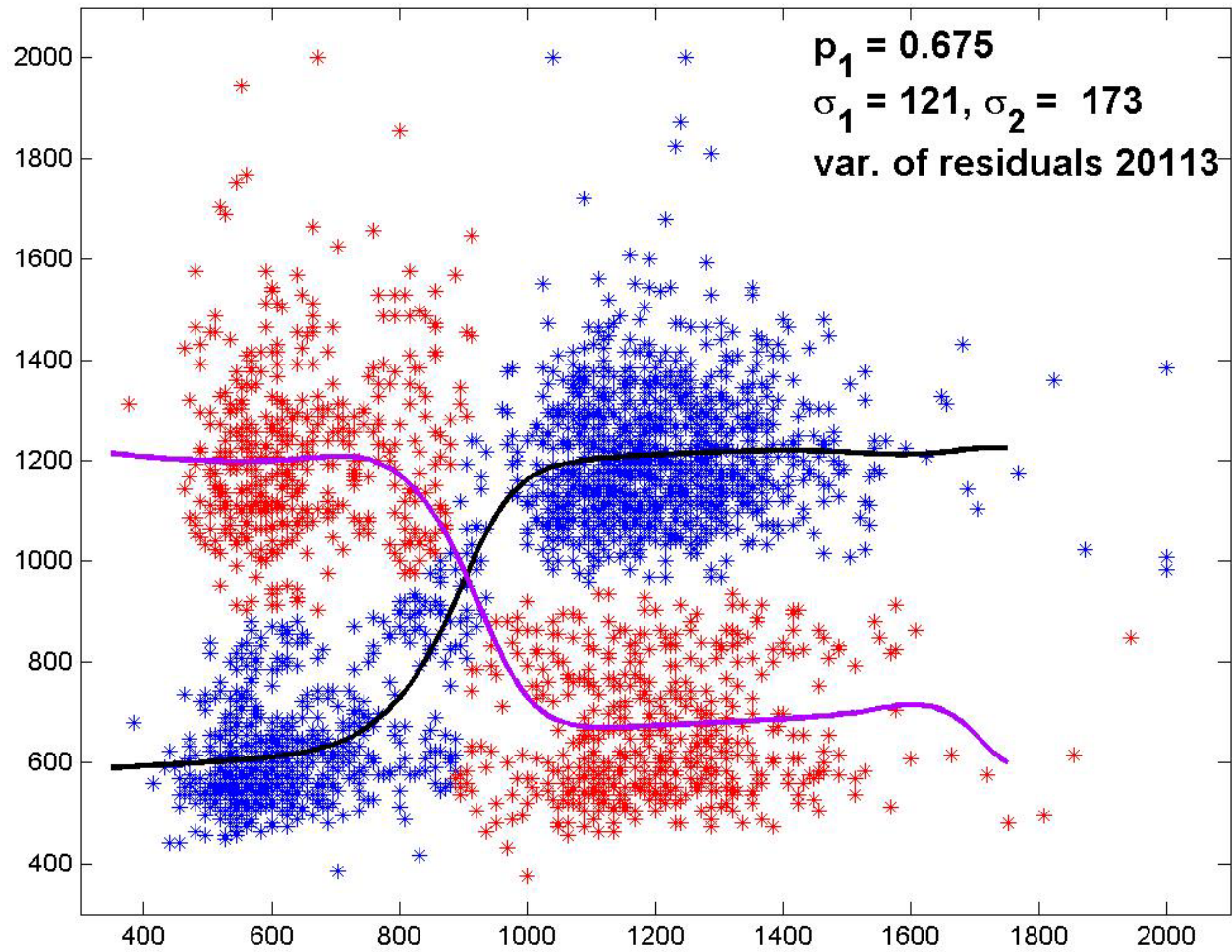




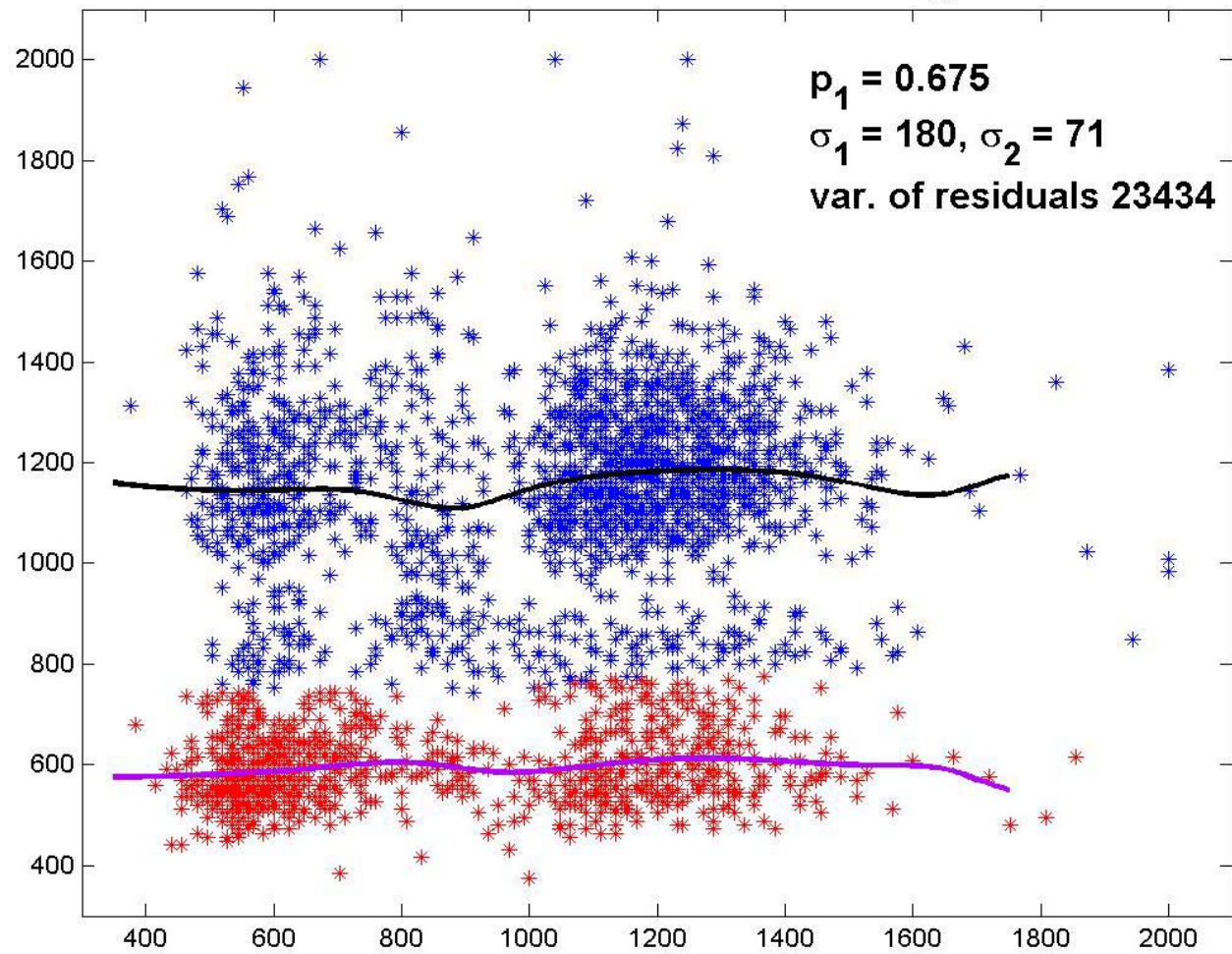
### Electrocardiogram (ECG) data



### ECG data - fitted mixture of two autoregressions I



## ECG data - fitted mixture of two autoregressions II



Forecasting of Daily Transformed BASF Values: Validation Set

