The Estimation of

Parameter Curves

for

Locally Stationary Processes

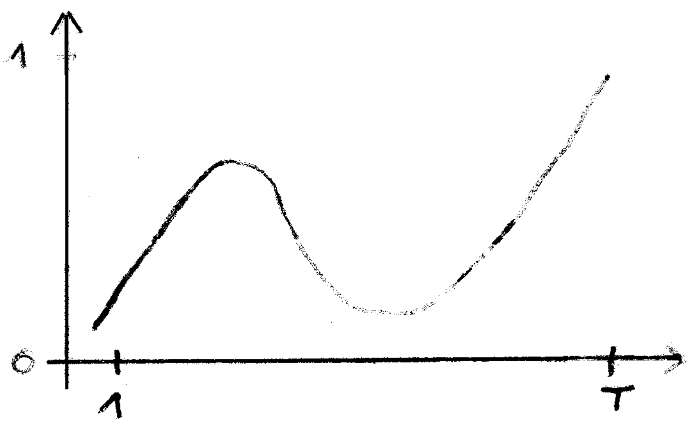
Rainer Dahlhaus

Universität Heidelberg

Example

OBSERVATIONS $X_1, \dots, X_T$

MODEL $X_t = g(t)X_{t-1} + \varepsilon_t, \quad g(t) = a + bt + ct^2 + dt^3$



EXAMPLE OF ESTIMATES

– MLE / Kalman Filter

– LS-Estimate

– Local AR(1) - Fit + LS – estimate for coefficients

GOALS

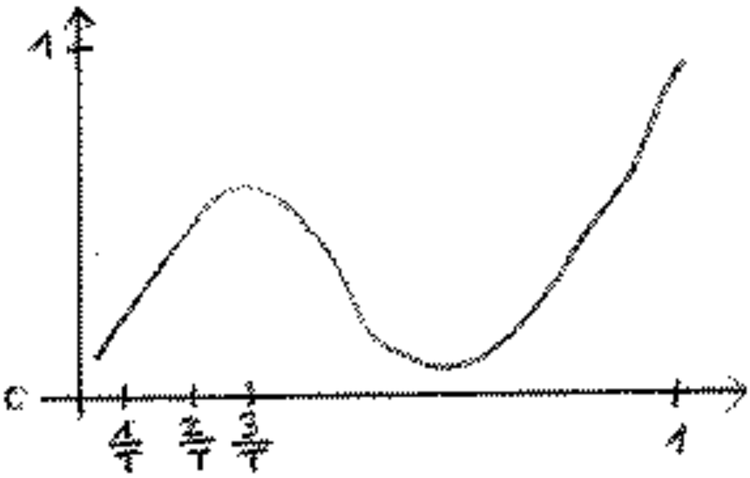
– compare different estimates

$\rightarrow$ reasonable asymptotics

– model selection

– model misspecification

Reasonable Asymptotics



"Observe" g(t) on a finer grid, i.e.

observe $X_{t,T} = g(\frac{t}{T})\, X_{t-1,T} + \varepsilon_t \qquad t = 1, \ldots , T$

(g is now rescaled).

– triangular array of observations

– g constant $\rightarrow X_{t,T}$ independent of T

$\rightarrow$ classical asymptotics

– $T \rightarrow \infty$ does not mean looking into the future
but: theory of prediction is still possible

NOTATION

t regular time

u = t/T rescaled time $u \in [0,1]$

Locally Stationary Processes

DEFINITION

$X_{t,T}$ is called locally stationary with mean function $\mu(u)$ and transfer function $A(u,\lambda)$ <u>iff</u>

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t)\, A\left(\frac{t}{T},\lambda\right)\, d\xi(\lambda).$$

$f(u,\lambda) := |A(u,\lambda)|^2$ is called time-frequency spectrum (definition simplified).

EXAMPLE

– $X_{t,T} = a\left(\frac{t}{T}\right) X_{t-1,T} + \sigma\left(\frac{t}{T}\right) \varepsilon_t$ AR(1)

– $X_{t,T} = b\left(\frac{t}{T}\right) \varepsilon_t + c\left(\frac{t}{T}\right) \varepsilon_{t-1}$ MA(1)

– $X_{t,T} = \mu\left(\frac{t}{T}\right) + \varepsilon_t$ (nonparametric regression)

– $X_{t,T} = \mu\left(\frac{t}{T}\right) + \sigma\left(\frac{t}{T}\right) Y_t$, Y_t stationary

Locally Stationary Processes

EXAMPLES

$$-\sum_{j=0}^{p} a_j\left(\frac{t}{T}\right) X_{t-j,T} = \sigma\left(\frac{t}{T}\right)\varepsilon_t, \qquad \varepsilon_t \text{ iid}$$

(time varying autoregression)

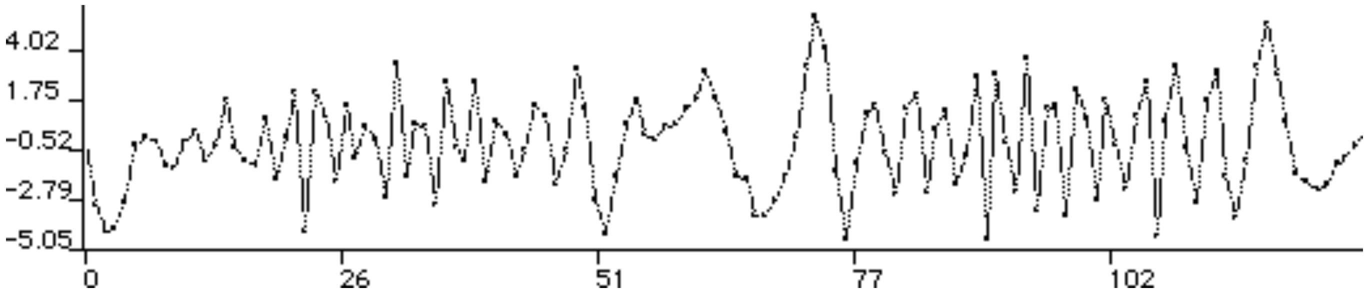


Figure 1. T=128 realisations of a time varying AR-model

Local Covariance Estimation

THEORETICAL COVARIANCE

$$c(u,k) := \int f(u,\lambda) \exp(i \lambda k)\, d\lambda$$

$$= \int |A(u,\lambda)|^2 \exp(i \lambda k)\, d\lambda$$

PROPERTY

If $c^*(u,k) := \mathrm{cov}(X_{[uT-k/2]}, X_{[uT+k/2]})$

then

$$|c(u,k) - c^*(u,k)| = O(T^{-1})$$

Local Covariance Estimation

EXAMPLE $(E\, X_t = 0)$

$$c_T(u,k) := \frac{1}{b_T T} \sum_t K\left(\frac{u - t/T}{b_T}\right) X_{t,T}\, X_{t+k,T}\ .$$

$$\textbf{BIAS}\ \big(c_T(u,k)\big) = \frac{1}{2}\, b_T^2 \int \alpha^2\, K(\alpha)\, d\alpha \left[\frac{\partial^2}{(\partial u)^2}\, c(u,k)\right] + o(b_T^2)\ .$$

CONSEQUENCE

– Influence of Nonstationarity can be quantified

$\longrightarrow$ optimal b_T , K

EQUIVALENT ESTIMATES

– ordinary covariance estimate on segment of length N (corresponds to above estimate with rectangular kernel and $b_T = N/T$)

– tapered covariance estimate on segment of length N (corresponds to $K(x) = h(x)^2$ and $b_T = N/T$)

Semiparametric AR - estimation

MODEL

$$X_{t,T} = a\left(\frac{t}{T}\right) X_{t-1,T} + \varepsilon_t$$

LOCAL YULE-WALKER EQUATIONS

$$\operatorname{cov}(X_{t,T}, X_{t-1,T}) = a\left(\frac{t}{T}\right) \operatorname{cov}(X_{t-1,T}, X_{t-1,T})$$

LOCAL AR-ESTIMATE

$$\hat{a}\left(\frac{t}{T}\right) = \frac{\hat{c}_T\left(\frac{t}{T},1\right)}{\hat{c}_T\left(\frac{t}{T},0\right)} \quad \text{with } \hat{c}_T(u,k) \text{ as above}$$

BIAS and **VARIANCE**

$$E\left(\hat{a}(u) - a(u)\right) = c_1 b_T^2 + c_2 \frac{1}{b_T T} \qquad \text{(constants can be}$$

$$\operatorname{var}(\hat{a}(u)) = c_3 \frac{1}{b_T T} \qquad \text{explicitly calculated)}$$

MEAN (INTEGRATED) SQUARED ERROR

minimize $E\left(\hat{a}(u) - a(u)\right)^2$ with respect to b_T and K

$\rightarrow b_T^{opt} = c(a(u))\, T^{-1/5}, K^{opt}(x) = \ldots$

$\rightarrow$ MSE, MISE = const. $T^{-4/5}$ (with L. Giraitis)

Estimation of Curves

SPECTRAL REPRESENTATION

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int \exp\ (i\lambda t)\ A\left(\frac{t}{T},\lambda\right) d\,\xi\,(\lambda)$$

with $A\,(u,\lambda) = A^{*}_{a\,(u)}(\lambda)$

GOAL

estimate $\theta(u) = (\mu(u), a(u))$

EXAMPLES

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \sigma\left(\frac{t}{T}\right)\varepsilon_t$$

$$X_{t,T} = b\left(\frac{t}{T}\right)\varepsilon_t + c\left(\frac{t}{T}\right)\varepsilon_{t-1}$$

$$X_{t,T} = a\left(\frac{t}{T}\right)X_{t-1,T} + \varepsilon_t$$

STATIONARY METHOD

Estimate $\theta(u)$ by choosing a small segment around u and proceed as if the process were stationary on that segment with parameter $\theta(u)$ (i.e. apply a classical method).

Parametric Models

$a(u) = a_\theta(u)$ (e. g. polynomials in u).

QUASI-LIKELIHOOD (Whittle-type)

$$\mathcal{L}_T(\theta) = \frac{1}{M}\sum_{j=1}^{M} \frac{1}{4\pi}\int \left\{ \log f_\theta(u_j,\lambda) + \frac{I_N(u_j,\lambda)}{f_\theta(u_j,\lambda)} \right\} d\lambda$$

$$\hat{\theta}_T = \operatorname{argmin} \mathcal{L}_T(\theta)$$

THEOREM

$\hat{\theta}_T$ is asympt. normal and efficient.

A Simulation Example

TRUE DATA **T = 128**

$$X_{t,T} = a_1\left(\frac{t}{T}\right) X_{t-1,T} + a_2\left(\frac{t}{T}\right) X_{t-2,T} + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0,1)$$

$$a_1(u) = 1.8 \cos(1.5 - \cos 4\pi u)$$
$$a_2(u) = -\,0.81$$

roots: $\frac{1}{0.9} \exp[\pm i\,(1.5 - \cos 4\pi u)$

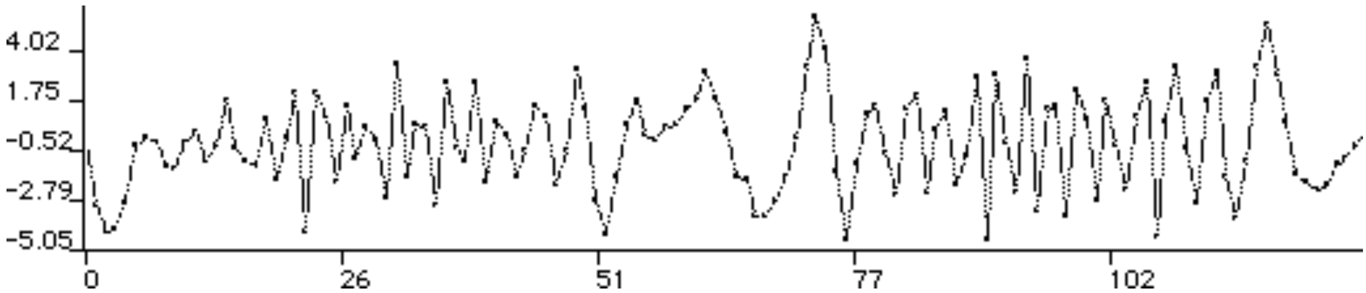


Figure 1. T=128 realisations of a time varying AR-model

ESTIMATION MODEL

$$X_{t,T} = \sum_{j=1}^{p} a_j\left(\frac{t}{T}\right) X_{t-j,T} + \varepsilon$$

$$a_j(u) = \sum_{k=0}^{K_j} b_{jk}\, u^k$$

$$\text{AIC}(p; K_1, \ldots, K_p) = \log \hat{\sigma}^2 + 2\left(p + 1 + \sum_{j=1}^{p} K_j\right) / T$$

K_2 \ K_1	4	5	6	7	8	9
0	0.929	0.888	0.669	0.685	0.673	0.689
1	0.929	0.901	0.678	0.694	0.682	0.698
2	0.916	0.888	0.694	0.709	0.697	0.712

Table 1. Values of AIC for p = 2 and different polynomial orders

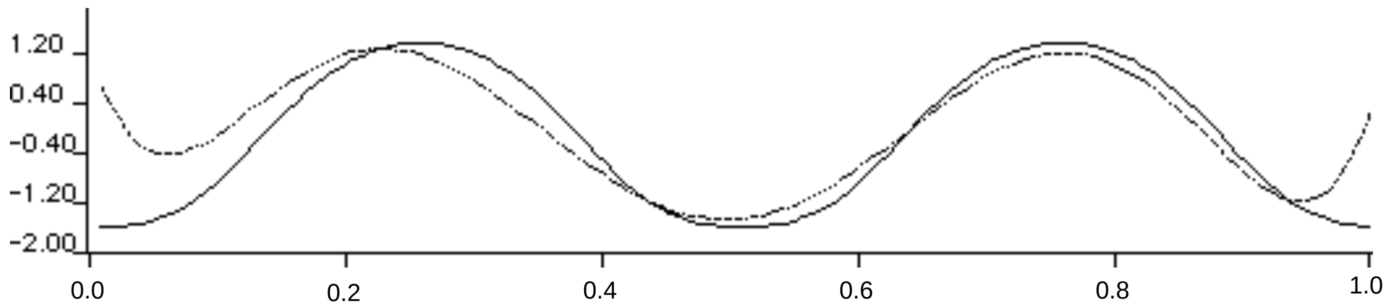


Figure 4. True and estimated time varying coefficient a (u)

true $a_2(u) = 0.81$

estimated $\hat{a}_2(u) = 0.79$

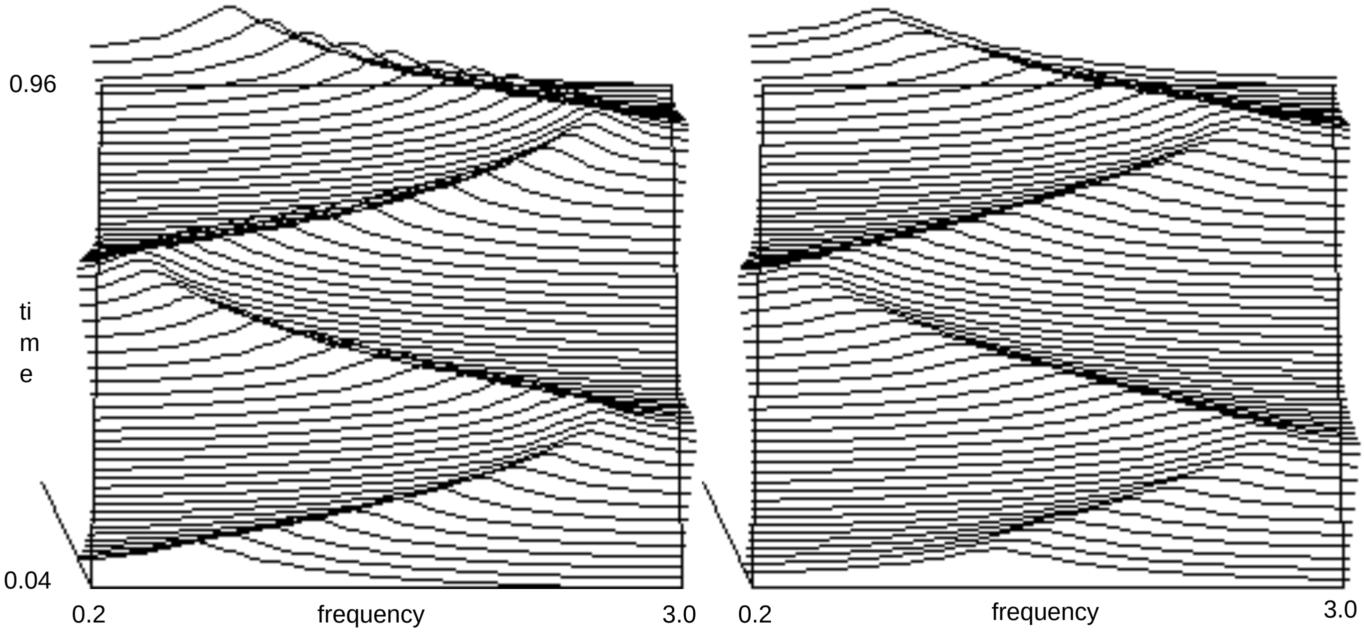


Figure 2. True and estimated spectrum of a time varying AR - process

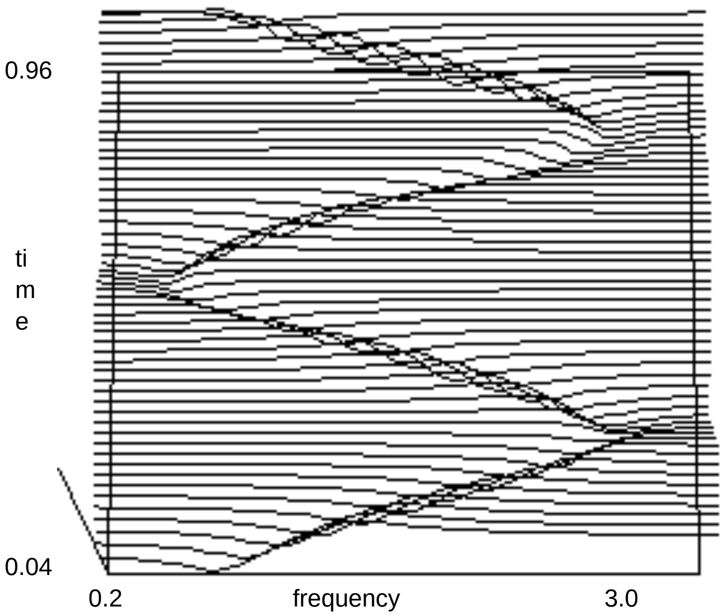


Figure 3. Difference of estimated and true spectrum

Model Selection

IDEA

– Consider the model as an approximation to the true unknown process
– estimate the best approximation (here: in the sense of the Kullback-Leibler distance), i.e. estimate θ_0.
– estimate goodness of fit, i.e. estimate $L(\theta_T^{MLE})$

NONSTATIONARY INFORMATION CRITERION (AIC -type)

$$\text{NIC} := \widehat{L}_T(\theta_T^{MLE}) + \frac{p}{T}$$

p = dimension of parameter space

$$\text{NIC} := \widehat{L}_T(\theta_T^{QE}) + \frac{p}{T}$$

(= AIC for stationary processes).

ADVANTAGES

– can select between stationary and nonstationary models

– can select between deterministic trends and "long range dependence" (in the sence of e.g. an AR - root close to the unit circle).

A Local Likelihood Approximation

LOCAL LIKELIHOOD

$$\mathcal{L}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \underbrace{\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log\ 4\pi^2 f_\theta\left(\frac{t}{T},\lambda\right) + \frac{\tilde{I}_T\left(\frac{t}{T},\lambda\right)}{f_\theta\left(\frac{t}{T},\lambda\right)}\right\} d\lambda}_{\ell_T\left(\theta,\frac{t}{T}\right)}$$

where

$$\tilde{I}_T(u,\lambda) = \frac{1}{2\pi}\sum_k X_{[uT+k/2]}\, X_{[uT-k/2]}\ \exp(i\lambda k)$$

(local periodogram)

($\tilde{I}_T$ was introduced for spectral estimation by Neumann and von Sachs, 1997).

A LOCAL LIKELIHOOD APPROXIMATION

$$\theta_T^L = \arg\min_{\theta} L_T^L(\theta)$$

THEOREM

(i) $\theta_T^L \xrightarrow{P^*} \theta_0$

(ii) $\sqrt{T}(\theta_T^L - \theta_0) \xrightarrow{D} N(0, V^{-1})$

(iii) If the model is correct, then θ_T^L is as. efficient (LAM).

Nonparametric Regression

MODEL

$$X_{t,T} = m(\tfrac{t}{T}) + \varepsilon_t \quad , \qquad \varepsilon_t \text{ Gaussian}$$

PARAMETRIC FIT

$$m(u) = m_\theta(u)$$

$$\hat{\theta}_T = \arg\min_\theta \frac{1}{T} \sum_{t=1}^{T} (X_{t,T} - m_\theta(\tfrac{t}{T}))^2$$

KERNEL ESTIMATE

$$\hat{m}(u) = \arg\min_m \frac{1}{b_T T} \sum_t K\left(\frac{u - t/T}{b_T}\right) \left(X_{t,T} - m\right)^2$$

LOCAL LINEAR FIT

$$\begin{pmatrix} \hat{m}(u) \\ \hat{b}(u) \end{pmatrix} = \arg\min_{m,b} \frac{1}{b_T T} \sum_t K\left(\frac{u - t/T}{b_T}\right)\left(X_{t,T} - m - b(\tfrac{t}{T} - m)\right)^2$$

ORTHOGONAL SERIES (WAVELETS)

$$\underline{\alpha} = \arg\min \frac{1}{T} \sum_t \left(X_{t,T} - \sum_{j=1}^{J} \alpha_j \psi_j(\tfrac{t}{T})\right)^2$$

+ shrinkage of $\underline{\alpha}$.

Local Likelihood Methods

MODEL

$$X_{t,T} = \mu(\tfrac{t}{T}) + \int \exp(i\lambda t)\, A(\tfrac{t}{T}, \lambda)\, d\xi(\lambda)$$

PARAMETRIC FIT $A = A_\theta, \quad \mu = \mu_\theta$

$$\theta_T^L = \arg\min_\theta \frac{1}{T} \sum_{t=1}^{T} \ell_T(\theta, \tfrac{t}{T})$$

KERNEL ESTIMATE $A(u,\lambda) = A_{\theta(u)}(\lambda), \ \mu(u) = \theta_1(u)$

$$\hat{\theta}(u) = \arg\min_\theta \frac{1}{b_T T} \sum_t K(\frac{u - t/T}{b_T})\, \ell_T(\theta, \tfrac{t}{T})$$

LOCAL LINEAR FIT

$$\begin{pmatrix} \hat{\theta}^{(0)}(u) \\ \hat{\theta}^{(1)}(u) \end{pmatrix} = \arg\min \frac{1}{b_T T} \sum_t K(\frac{u - t/T}{b_T})\, \ell_T(\theta^{(0)} + \theta^{(1)}(\tfrac{t}{T} - u), \tfrac{t}{T})$$

ORTHOGONAL SERIES (WAVELETS)

$$\underline{\alpha} = \arg\min \frac{1}{T} \sum_t \ell_T\ (\sum_{j=1}^{J} \alpha_j \psi_j(\tfrac{t}{T}), \tfrac{t}{T})$$

+ shrinkage of $\underline{\alpha}$

REFERENCES

Dahlhaus, R. (1997a) Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1-37.

Dahlhaus, R. (1997b) The Analysis of Nonstationary Time Series and Curve Estimation with Locally Stationary Models. Lecture Notes. Academia Sinica, Taiwan.

Dahlhaus, R. (1996) Asymptotic statistical inference for non-stationary processes with evolutionary spectra. In: Athens Conference on Applied Probability and Time Series, vol. II (eds. P.M. Robinson and M. Rosenblatt). Springer Lecture Notes in Statistics **115**, 145-159.

Dahlhaus, R. and Giraitis, L. (1997) On the optimal segment length for parameter estimates for locally stationary time-series. *J. Time Ser. Analysis, to appear*.

Dahlhaus, R., Neumann, M. and v. Sachs, R. (1997) Nonlinear wavelet estimation of time-varying processes. *Bernoulli, to appear*.

Neumann, M.H. und von Sachs, R. (1997) Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.*, **25**, 38-76.

Künsch, H.R. (1995) A note on causal solutions for locally stationary AR-processes. ETH Zürich.