Adaptive Bandwidth Selection for Locally Stationary Processes

Stefan Richter (joint work with Rainer Dahlhaus, Jan Johannes)

February 16th, 2016 at CIRM

Heidelberg University Research Training Group 1653

1. Locally stationary processes
   - Motivation and Examples
   - Local conditional maximum likelihood estimators
2. Global bandwidth selection via Cross Validation
   - Description of method
   - Results
   - Simulations
3. Work in progress: Local bandwidth selection via contrast minimisation and Lepski's method

Locally stationary processes - definition

Definition, cf. Dahlhaus, Polonik (2009)

Assume that $\varepsilon_t, t \in \mathbb{Z}$ are iid with $\mathbb{E}\varepsilon_t = 0$, $\mathrm{Var}(\varepsilon_t) = 1$ and

$$X_{t,n} = \sum_{k=0}^{\infty} a_{t,n}(k)\varepsilon_{t-k}, \qquad t = 1, ..., n$$

with $\sup_{t,n} |a_{t,n}(k)| \leq \frac{C}{\ell(k)}$, where $\ell(k) = |k| \log^{1+\kappa} |k|$, $\kappa > 0$.

Locally stationary processes - definition

Definition, cf. Dahlhaus, Polonik (2009)

Assume that $\varepsilon_t, t \in \mathbb{Z}$ are iid with $\mathbb{E}\varepsilon_t = 0$, $\mathrm{Var}(\varepsilon_t) = 1$ and

$$X_{t,n} = \sum_{k=0}^{\infty} a_{t,n}(k)\varepsilon_{t-k}, \qquad t = 1, ..., n$$

with $\sup_{t,n} |a_{t,n}(k)| \le \frac{C}{\ell(k)}$, where $\ell(k) = |k| \log^{1+\kappa} |k|$, $\kappa > 0$.
Assume there are $a(\cdot, k) : [0,1] \to \mathbb{R}$ with

$$\sup_k \sum_{t=1}^{n} |a_{t,n}(k) - a(t/n, k)| \le C.$$

Then $(X_{t,n})_{t=1,\dots,n}$ is called a locally stationary process.

Locally stationary processes - corresp. stationary process

$$X_{t,n} = \sum_{k=0}^{\infty} a_{t,n}(k)\varepsilon_{t-k}, \qquad a_{t,n}(k) \approx a(t/n, k)$$

Parametric assumption

The time dependence of $a(u,k)$ on $u \in [0,1]$ is solely via a finite dimensional parameter curve $\theta_0 : [0,1] \to \Theta \subset \mathbb{R}^d$, i.e.

$$a(u,k) = a_{\theta_0(u)}(k) \quad \text{for all } u \in [0,1].$$

with some $a_{\cdot}(k) : \Theta \to \mathbb{R}$ $(k \in \mathbb{Z})$.
We assume θ_0 is Hoelder-continuous with some exponent $\beta > 0$ and is of bounded variation.

Remark: This assumption ensures that the locally stationary process is obtained by a stationary process by replacing the constant parameters θ through parameter curves θ_0.

Example 1: tvARMA processes

tvARMA(p, q) processes

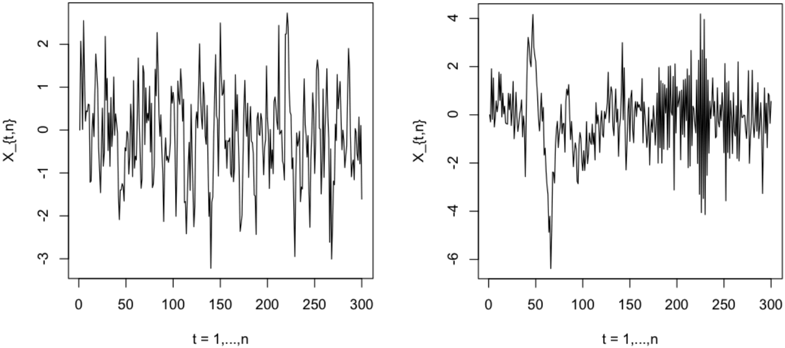
Let $\varepsilon_t \overset{iid}{\sim} (0,1)$. If $a_k, b_l : [0,1] \to \mathbb{R}$ have bounded variation and $\alpha(z) := \sum_{k=0}^{p} a_k(u) z^k$, $\beta(z) := \sum_{l=0}^{q} b_l(u) z^l$ have only zeros outside the unit circle uniformly in $u \in [0,1]$, then solutions $(X_{t,n})$ of

$$\sum_{k=0}^{p} a_k\left(\frac{t}{n}\right) X_{t-k,n} = \sum_{l=0}^{q} b_l\left(\frac{t}{n}\right) \varepsilon_{t-l}$$

are locally stationary processes.

Here, we assume $a_0 \equiv b_0 \equiv 1$. Then $\theta_0(u) = (a_1(u), ..., a_p(u), b_1(u), ..., b_q(u))' : [0,1] \to \Theta \subset \mathbb{R}^{p+q}$.

Example 1: Special case tvAR(1)



$X_t = 0.5X_{t-1} + \varepsilon_t$

$X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t,$
$\theta_0(u) = 0.9 \cdot \sin(2\pi u)$

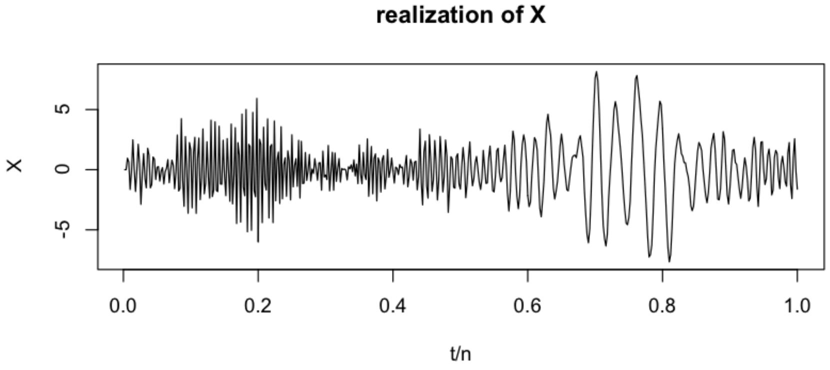
$\varepsilon_t \overset{iid}{\sim} N(0,1).$

Example 2: special parametrised tvAR(2) process

$$X_{t,n} = \frac{2}{r_0} \cos\left(\phi\left(\frac{t}{n}\right)\right) X_{t-1,n} - \frac{1}{r_0^2} X_{t-2,n} + \sigma(t/n)\varepsilon_t,$$

$$\phi(u) = 1.4 + \sin(2\pi u), \quad r_0 = 1.05, \quad \sigma(u) = 1.0, \quad \varepsilon_t \overset{iid}{\sim} N(0,1).$$

Here, $r_0 \cdot e^{\pm i\phi(u)}$ are the two roots of the characteristic polynomial of the process, and $\theta_0(u) = (\phi(u), \sigma(u))' : [0,1] \to \Theta \subset \mathbb{R}^2$.



Estimation of parameter curves

Definition of estimator $\hat{\theta}_h(\cdot)$

We define

$$\hat{\theta}_h(u) := \text{argmin}_{\theta\in\Theta} L_{n,h}(u,\theta),$$

where

$$L_{n,h}(u,\theta) := \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{\frac{t}{n}-u}{h}\right) l_{t,n}(\theta)$$

is a local likelihood (kernel K, bandwidth h) at time $u \in [0,1]$ and

$$l_{t,n}(\theta) := \log p_{\theta_0(\cdot)=\theta}(X_{t,n}|X_{t-1,n},...,X_{1,n}, X_{0,n}=0, X_{-1,n}=0,...)$$

is a negative log infinite past conditional Gaussian Likelihood, where the curve $\theta_0(\cdot)$ is replaced by a constant parameter θ.

Discussion: $l_{t,n}(\theta)$

Proposition

Define

$$A_\theta(\lambda) := \sum_{k=0}^{\infty} a_\theta(k) e^{-i\lambda k}, \qquad \gamma_\theta(k) := \int_{-\pi}^{\pi} A_\theta(-\lambda)^{-1} e^{i\lambda k} d\lambda,$$

$$d_{t,n}(\theta) := \frac{1}{2\pi} \sum_{k=0}^{t-1} \gamma_\theta(k) X_{t-k,n}$$

If $|A_\theta(\lambda)| \geq \delta > 0$ uniformly in λ, θ, then

$$l_{t,n}(\theta) = -\frac{1}{2} \log\left(\frac{\gamma_\theta(0)^2}{2\pi}\right) + \frac{1}{2}\left[d_{t,n}(\theta)\right]^2.$$

Discussion: $l_{t,n}(\theta)$

Proposition

Define

$$A_\theta(\lambda) := \sum_{k=0}^{\infty} a_\theta(k) e^{-i\lambda k}, \qquad \gamma_\theta(k) := \int_{-\pi}^{\pi} A_\theta(-\lambda)^{-1} e^{i\lambda k} d\lambda,$$

$$d_{t,n}(\theta) := \frac{1}{2\pi} \sum_{k=0}^{t-1} \gamma_\theta(k) X_{t-k,n}$$

If $|A_\theta(\lambda)| \geq \delta > 0$ uniformly in λ, θ, then

$$l_{t,n}(\theta) = -\frac{1}{2} \log\left(\frac{\gamma_\theta(0)^2}{2\pi}\right) + \frac{1}{2}\Big[d_{t,n}(\theta)\Big]^2.$$

- A finite past likelihood would have coefficients which are harder to calculate and to investigate theoretically (that is why we use infinite past).
- $tvAR(1) : X_{t,n} = \theta_0(t/n) X_{t-1,n} + \varepsilon_t$ has $l_{t,n}(\theta) \sim (X_{t,n} - \theta X_{t-1,n})^2$.
- $tvMA(1) : X_{t,n} = \varepsilon_t + \theta_0(t/n)\varepsilon_{t-1}$ has $l_{t,n}(\theta) \sim \left(\sum_{k=0}^{t-1} \theta^k X_{t-k,n}\right)^2$.

Adaptive bandwidth selection of h

- There is a huge literature about adaptive bandwidth selection in the iid regression case, but up to our knowledge no general theoretical results are available for locally stationary processes

Adaptive bandwidth selection of h

- There is a huge literature about adaptive bandwidth selection in the iid regression case, but up to our knowledge no general theoretical results are available for locally stationary processes
- The main difficulty: Statistical properties of the locally stationary process (like variance) usually are strongly connected to the unknown parameter curve θ_0. Example: $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$,

Adaptive bandwidth selection of h

- There is a huge literature about adaptive bandwidth selection in the iid regression case, but up to our knowledge no general theoretical results are available for locally stationary processes
- The main difficulty: Statistical properties of the locally stationary process (like variance) usually are strongly connected to the unknown parameter curve θ_0.
  Example: $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$,
- Estimators usually have a more complicated structure which forces us to use (Taylor) approximations and carefully inspect convergence rates of the remaining terms
  Example: $\hat{\theta}_h(u) = \frac{\hat{c}_{1,h}(u)}{\hat{c}_{0,h}(u)}$, $\mathbb{E}[(\hat{\theta}_h(u) - \theta_0(u))^2]$ harder to calculate.

Adaptive bandwidth selection of h

- There is a huge literature about adaptive bandwidth selection in the iid regression case, but up to our knowledge no general theoretical results are available for locally stationary processes
- The main difficulty: Statistical properties of the locally stationary process (like variance) usually are strongly connected to the unknown parameter curve θ_0.
  Example: $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$,
- Estimators usually have a more complicated structure which forces us to use (Taylor) approximations and carefully inspect convergence rates of the remaining terms
  Example: $\hat{\theta}_h(u) = \frac{\hat{c}_{1,h}(u)}{\hat{c}_{0,h}(u)}$, $\mathbb{E}[(\hat{\theta}_h(u) - \theta_0(u))^2]$ harder to calculate.
- Plugin estimators need strong smoothness assumptions on the unknown parameter curve θ_0 and good pre-estimators for the unknown terms occuring in the MSE-asymptotic optimal bandwidth

Adaptive bandwidth selection of h

- There is a huge literature about adaptive bandwidth selection in the iid regression case, but up to our knowledge no general theoretical results are available for locally stationary processes
- The main difficulty: Statistical properties of the locally stationary process (like variance) usually are strongly connected to the unknown parameter curve θ_0.
  Example: $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$,
- Estimators usually have a more complicated structure which forces us to use (Taylor) approximations and carefully inspect convergence rates of the remaining terms
  Example: $\hat{\theta}_h(u) = \frac{\hat{c}_{1,h}(u)}{\hat{c}_{0,h}(u)}$, $\mathbb{E}[(\hat{\theta}_h(u) - \theta_0(u))^2]$ harder to calculate.
- Plugin estimators need strong smoothness assumptions on the unknown parameter curve θ_0 and good pre-estimators for the unknown terms occuring in the MSE-asymptotic optimal bandwidth
- We will focus on adaptive methods which work under Hoelder-continuity assumptions on θ_0

Cross validation in the general setting

Cross validation

Now define $L_{n,h,-t}(u,\theta) := \frac{1}{nh}\sum_{s=1,s\neq t}^{n} K\left(\frac{\frac{s}{n}-u}{h}\right) l_{s,n}(\theta)$ and the leave-one-out estimate

$$\hat{\theta}_{h,-t}(u) := \mathrm{argmin}_{\theta\in\Theta} L_{n,h,-t}(u,\theta).$$

The global cross validation function is defined via

$$CV(h) := \frac{1}{n}\sum_{t=1}^{n} l_{t,n}(\hat{\theta}_{h,-t}(t/n)),$$

with a set of bandwidths H_n, and

$$\hat{h} = \mathrm{argmin}_{h\in H_n} CV(h).$$

Final estimator of $\theta_0(u)$: $\hat{\theta}_{\hat{h}}(u)$.

Results

Distance measure

Define a Kullback-Leiber-type distance measure ($\|x\|_A^2 := \langle x, Ax \rangle$)

$$d_A(\hat{\theta}_h, \theta_0) := \frac{1}{n} \sum_{t=1}^{n} \|\hat{\theta}_h(t/n) - \theta_0(t/n)\|_{I(\theta_0(t/n))}^2$$

with a weight matrix $I(\cdot)$, the local Fisher information

$$I(\theta) := \frac{1}{4\pi} \int_{-\pi}^{\pi} (\nabla_\theta \log f_\theta(\lambda)) \cdot (\nabla_\theta \log f_\theta(\lambda))' d\lambda, \qquad f_\theta(\lambda) := \frac{1}{2\pi} |A_\theta(\lambda)|^2.$$

Remark: $d_A(\theta_1, \theta_0)$ is the leading term of a taylor expanded empirical Kullback-Leibler divergence $\log \frac{p_{\theta_1}(X_{1,n}, \ldots, X_{n,n})}{p_{\theta_1}(X_{1,n}, \ldots, X_{n,n})}$.

Result

Theorem 1

Assume that the kernel $K \geq 0$ is Lipschitz continuous, and

- Parameter space Θ is compact, θ is identifiable from A_θ, $\theta_0(u) \in \text{int}(\Theta)$ for all $u \in [0,1]$.
- $A_\theta(\lambda)$ is twice differentiable in θ. $A_\theta(\lambda)^{-1}$ and the components of the first two derivatives are uniformly bounded in λ, θ and are uniformly Hoelder-continuous in λ with some exponent $\beta > 1$.
- The minimal eigenvalue of $I(\theta)$ is bounded away from 0 uniformly in θ.
- All moments of ε_0 exist.
- $H_n = [\underline{h}_n, \overline{h}_n]$ with $\underline{h}_n \geq c_0 n^{\delta-1}$, $\overline{h}_n \leq c_1 n^{-\delta}$ for some $c_0, c_1, \delta > 0$.
- θ_0 has bounded variation and θ_0 is β-Hoelder-continuous with $\beta > 0$

Then

$$\lim_{n\to\infty} \frac{d_A(\hat{\theta}_{\hat{h}}, \theta_0)}{\inf_{h\in H_n} d_A(\hat{\theta}_h, \theta_0)} = 1 \quad a.s.$$

i.e., $\hat{h}$ is asymptotically optimal with respect to d_A.

Result

Theorem 2

Let the assumptions of Theorem 1 hold. Assume that θ_0 is twice continuously differentiable. Define

$$d_M^{**}(h) := \frac{V_0}{nh} + \frac{h^4}{4} B_0,$$

B_0, V_0 complicated terms, dependent on θ_0. If $B_0 > 0$, we have

$$\sup_{h \in H_n} \left| \frac{d_A(\hat{\theta}_h, \theta_0) - d_M^{**}(h)}{d_M^{**}(h)} \right| \to 0, \qquad \frac{\hat{h}}{h^*} \to 1,$$

where $h^* = \left(\frac{V_0}{B_0}\right)^{1/5} n^{-1/5}$ is the unique minimizer of $d_M^{**}(h)$ and can be seen as the MSE-asymptotic optimal bandwidth.

Remark: This means that $d_A(\hat{\theta}_h, \theta_0)$ asymptotically allows for a usual bias-variance-decomposition.

Proof idea of Theorem 1:

One main step of the proof is to show $\sup_{h \in H_n} |\mathrm{diff}(h)| \to 0$ a.s. (*), where

$$\mathrm{diff}(h) := \frac{CV(h) - \frac{1}{n}\sum_{t=1}^{n} l_{t,n}(\theta_0(t/n)) - \overline{d}_A(\hat{\theta}_h, \theta_0)}{d_M^*(\hat{\theta}_h, \theta_0)},$$

$$\overline{d}_A(\hat{\theta}_h, \theta_0) = \frac{1}{n}\sum_{t=1}^{n} \left\| \hat{\theta}_{h,-t}(t/n) - \theta_0(t/n) \right\|_{I(\theta_0(t/n))}^2,$$

$$d_M^*(\hat{\theta}_h, \theta_0) = \mathbb{E}\frac{1}{n}\sum_{t=1}^{n} \| \nabla_\theta L_{n,h}(t/n, \theta_0(t/n)) \|_{I(\theta_0(t/n))^{-1}}^2$$

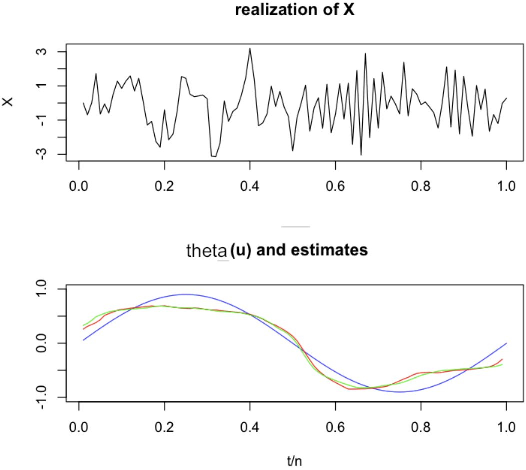
By using Hoelder-continuity in h we can discretize H_n. By the Borel Cantelli lemma, (*) is proved if (fixed $K > 0$, $\varepsilon > 0$)

$$n^K \cdot \sup_{h \in H_n} \mathbb{P}(|\mathrm{diff}(h)| > \varepsilon) \to 0. \quad (**)$$

To verify (**), we use Markov's inequality, Taylor expansions of $l_{t,n}$ and bounds for moments of martingale difference sequences by Rio (2009).

Simulations: tvAR(1)

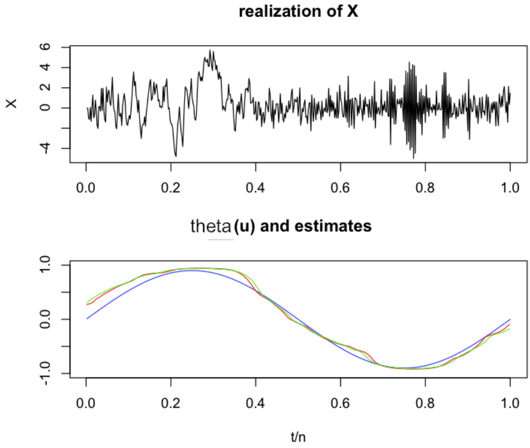
$n = 100$, $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$ with $\varepsilon_t \overset{iid}{\sim} N(0,1)$, $\theta_0(u) = 0.9\sin(2\pi u)$.



red: $\hat{\theta}_{\hat{h}}$ (CV), green: $\hat{\theta}_{h^*}$ (MSE optimal bandwidth), blue: true θ_0

Simulations: tvAR(1)

$n = 500$

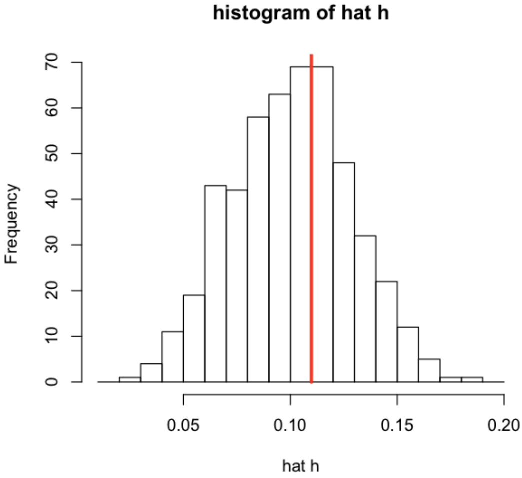


red: $\hat{\theta}_{\hat{h}}$ (CV), green: $\hat{\theta}_{h^*}$ (MSE optimal bandwidth), blue: true θ_0

Simulations: tvAR(1)

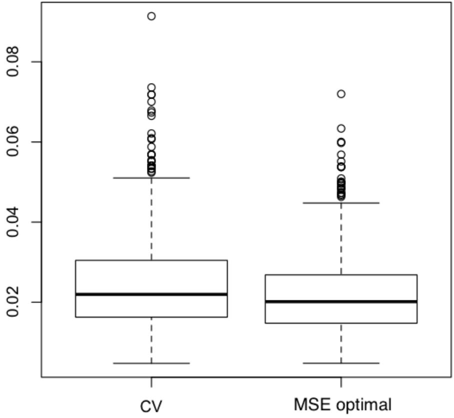
$n = 500$.

Plot: Cross validation bandwidth $\hat{h}$ (histogram, 500 reps), the bandwidth h^* (red) which minimizes $d_M^{**}(h)$



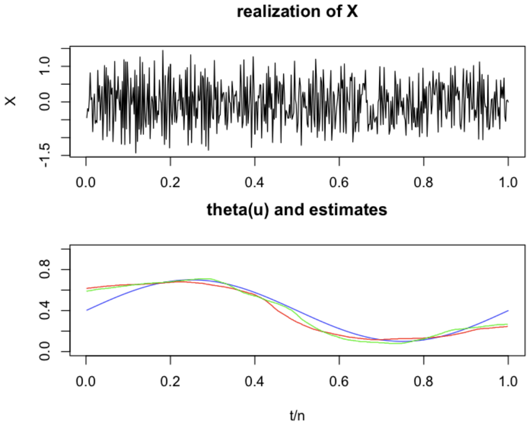
Simulations: tvAR(1)

$n = 500$.
Boxplot (500 reps) of $d_A(\hat{\theta}_h, \theta) = \frac{1}{n}\sum_{t=1}^{n}(\hat{\theta}_h(t/n) - \theta_0(t/n))^2 \cdot I(\theta_0(t/n))$



Simulations: tvMA(1)

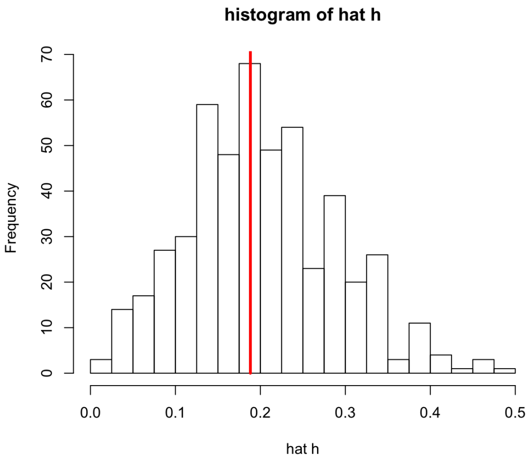
$n = 500$, $X_{t,n} = \varepsilon_t - \theta_0(t/n)\varepsilon_{t-1}$ with $\varepsilon_t \overset{iid}{\sim} N(0,1)$, $\theta_0(u) = 0.4 + 0.3 \cdot \sin(2\pi \cdot u)$.



red: $\hat{\theta}_{\hat{h}}$ (CV), green: $\hat{\theta}_{h^*}$ (MSE optimal bandwidth), blue: true θ_0

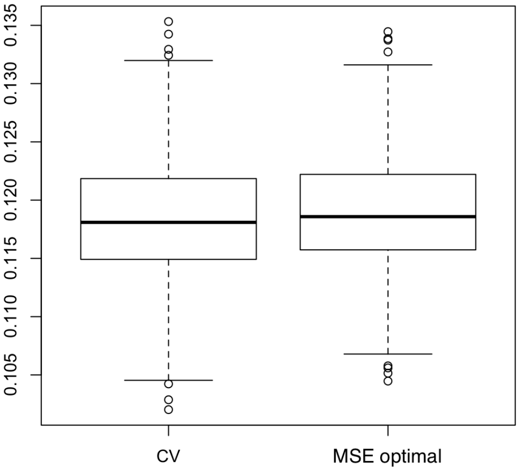
Simulations: tvMA(1)

$n = 500$.
Plot: Cross validation bandwidth $\hat{h}$ (histogram, 500 reps), the bandwidth h^* (red) which minimizes $d_M^{**}(h)$



Simulations: tvMA(1)

$n = 500$.
Boxplot (500 reps) of $d_A(\hat{\theta}_h, \theta)$ for $h = \hat{h}$, $h = h^*$



Conclusion - Crossvalidation

- We have proposed kernel density estimators and an adaptive bandwidth selection procedure via cross validation for unknown parameter curves in locally stationary processes
- We have shown that the proposed estimators are asymptotically optimal
- Simulations show good behaviour of the estimators, in particular in the tvAR case

Future work:

- Nonlinear processes and local bandwidth selection

Part 2: Locally adaptive bandwidth selection via contrast minimisation

- Work in progress (joint with Jan Johannes)
- Instead of linear processes we assume a recursive Markov structure of the process $X_{t,n}$, again depending on an unknown parameter curve θ_0
- We use a general method which was first applied to the iid regression model in Lepski, Mammen and Spokoinij (1997) and Goldenshluger and Lepski (2011) to get locally adaptive bandwidth selectors for kernel estimators of θ_0

Model

Model

Assume that

$$X_{t,n} = G_{\varepsilon_t}\Big(X_{t-1,n}, ..., X_{t-p,n}, \theta_0\big(\frac{t}{n} \vee 0\big)\Big), \qquad t \leq n \tag{1}$$

where $G_\varepsilon(x,\theta) = \mu(x,\theta) + \sigma(x,\theta)\varepsilon_t$ with

- Deterministic functions $\mu, \sigma : \mathbb{R}^p \times \Theta \to \mathbb{R}$,
- Parameter curve $\theta_0 : [0,1] \to \Theta$ (Hoelder-continuous with exponent $\beta > 0$)
- $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an iid sequence with $\mathbb{E}\varepsilon_t = 0$ and $\mathrm{Var}(\varepsilon_t) = 1$.

Remarks:

- G has to fulfil some (weak) assumptions such that $X_{t,n}$ exists.
- Examples: as tvTAR, tvexpAR, tvARCH, ...

Estimation of parameter curves

Definition of estimator $\hat{\theta}_h(\cdot)$

We define

$$\hat{\theta}_h(u) := \text{argmin}_{\theta \in \Theta} L_{n,h}(u, \theta),$$

where

$$L_{n,h}(u, \theta) := \frac{1}{nh} \sum_{t=p+1}^{n} K\left(\frac{\frac{t}{n} - u}{h}\right) l_{t,n}(\theta)$$

is a local likelihood (kernel K, bandwidth h) at time $u \in [0, 1]$ and

$$\begin{aligned} l_{t,n}(\theta) &:= \log p_{\theta_0(\cdot)=\theta}(X_{t,n} | Y_{t-1,n}) \\ &= \frac{\left(X_{t,n} - \mu(Y_{t-1,n}, \theta)\right)^2}{2\sigma(Y_{t-1,n}, \theta)^2} + \frac{1}{2} \log \sigma(Y_{t-1,n}, \theta)^2 \end{aligned}$$

is a negative log conditional Gaussian Likelihood ($Y_{t-1,n} := (X_{t-1,n}, ..., X_{t-p,n})$).

Contrast minimisation / Lepski's method

Local bandwidth selection procedure

Let H_n be a (discrete) set of bandwidths. Define Lepski's contrast function

$$Y(u,h) := \max_{h' \in H_n, h' \leq h} \Big\{ \|\hat{\theta}_h(u) - \hat{\theta}_{h'}(u)\|_2^2 - \mathrm{pen}(u,h') \Big\},$$

and the local bandwidth selector

$$\tilde{h}(u) := \arg\min_{h \in H_n} \{ Y(u,h) + \mathrm{pen}(u,h) \}.$$

The penalisation term $\mathrm{pen}(u,h)$ has still to be defined.

Remark: $Y(u,h)$ mimics the bias part, $\mathrm{pen}(u,h)$ the variance part of the MSE decomposition $\mathbb{E}\|\hat{\theta}_h(u) - \theta_0(u)\|_2^2$

Contrast minimisation / Lepski's method

Key argument

Assume that $\text{pen}(u, h)$ is monotone in $h \in H_n$. For all $h \in H_n$, it holds

$$\|\hat{\theta}_{\tilde{h}(u)}(u) - \theta_0(u)\|_2^2 \quad \leq \quad 85 \cdot \max\{\text{bias}(u,h)^2, \text{pen}(u,h)\} \\ +42 \max_{h' \in H_n, h' \leq h} \left\{ \|\hat{\theta}_{h'}(u) - \theta_{h'}(u)\|_2^2 - \text{pen}(u, h') \right\}$$

Here,

- $\theta_h(u)$ is a theoretical approximation of $\theta_0(u)$,
- $\text{bias}(u,h) := \sup_{h' \in H_n, h' \leq h} \|\theta_{h'}(u) - \theta_0(u)\|$ is the approximation error.

Special case: tvAR(1)

Special case: $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$

We choose $H_n = \{a^{-k} : a^{-k} \geq C_1 \log^2(n)n^{-1}, k \in \mathbb{N}_0\}$ and

$$\hat{\theta}_h(u) := \frac{\hat{c}_{1,h}(u)}{\hat{c}_{0,h}(u)} \mathbb{1}_{\{\hat{c}_{0,h}(u) \geq n^{-1}\}},$$

$$\hat{c}_{i,h}(u) := \frac{1}{nh} \sum_{t=1}^{n} K\left(\frac{t/n - u}{h}\right) X_{t-1,n}X_{t-1+i,n}.$$

We set

$$\text{bias}(u,h) := \frac{\mathbb{E}[\hat{c}_{1,h}(u) - \theta_0(u)\hat{c}_{0,h}(u)]}{c_{0,h}(u)} = O(h^{\beta}), \qquad c_{0,h}(u) := \mathbb{E}[\hat{c}_{0,h}(u)].$$

$$\theta_h(u) := \theta_0(u) + \text{bias}(u,h),$$

$$\text{pen}(u,h) := C_2 \cdot \log^2(n) \cdot \frac{\text{Var}\big(\hat{c}_{1,h}(u) - \theta_0(u)\hat{c}_{0,h}(u)\big)}{c_{0,h}(u)^2}.$$

Special case: tvAR(1)

Theorem 3 (for $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$)

Assume that $\varepsilon_t \overset{iid}{\sim} (0,1)$, $\mathbb{E}\varepsilon_t^k \le C_\varepsilon^k$ for all $k \in \mathbb{N}$, $\theta_0 : [0,1] \to [-1+\delta, 1-\delta]$ ($\delta > 0$) is Hoelder-continuous with exponent $\beta > 0$. Then for some $C_0 > 0$ independent of n and θ_0,

$$\mathbb{E}|\hat{\theta}_{\tilde{h}(u)}(u) - \theta_0(u)|^2 \le C_0(\log^2(n)n^{-1})^{\frac{\beta}{2\beta+1}},$$

i.e. it is asymptotically minimax up to a log factor since (Arkoun et al. 2014):

$$\inf_{\hat{\theta}} \sup_{\theta_0 \in \Sigma(\beta,L)} \mathbb{E}_{\theta_0}|\hat{\theta}_h(u) - \theta_0(u)|^2 \ge c_0 \cdot n^{-\frac{\beta}{2\beta+1}}.$$

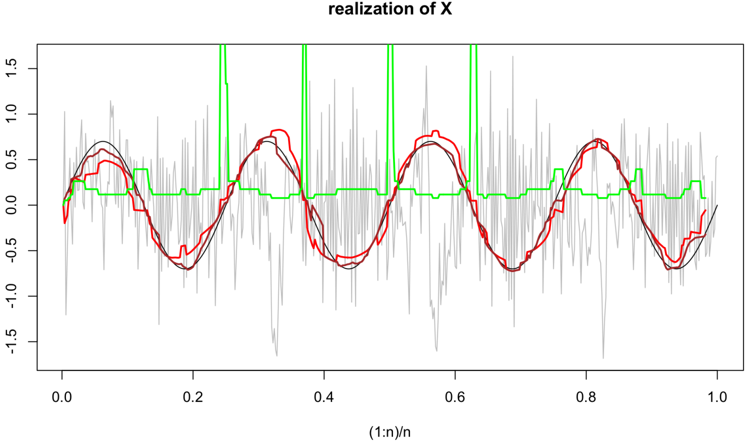
Conjecture 3 (for $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$)

Theorem 3 holds true if we replace $\text{pen}(u,h)$ by its empirical counterpart

$$\widehat{\text{pen}}(u,h) := \frac{1}{nh} \cdot \frac{\frac{1}{nh}\sum_{t=1}^n K^2\left(\frac{t/n-u}{h}\right)X_{t-1,n}^2}{\hat{c}_{0,h}(u)^2}.$$

Simulations: tvAR(1)

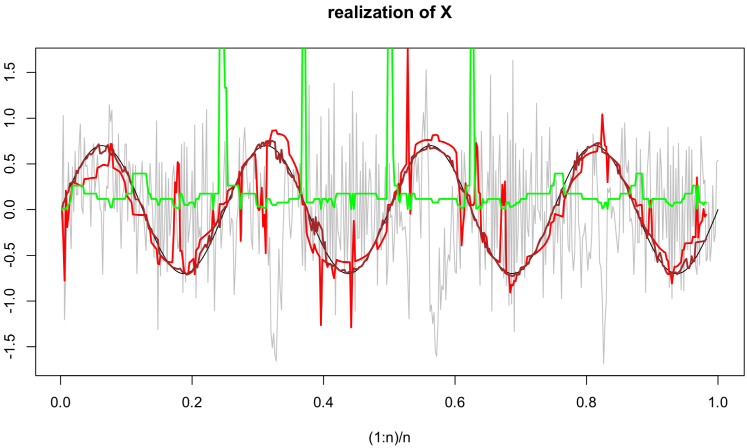
$n = 500$, $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$ with $\varepsilon_t \overset{iid}{\sim} N(0, 0.5^2)$, $\theta_0(u) = 0.7\sin(8\pi u)$.



red: $\hat{\theta}_{\tilde{h}(u)}$ (adaptive bandwidth estimate), brown: $\hat{\theta}_{h^*(u)}$ (Optimal estimate), black: true θ_0, green: $2\tilde{h}(u) \in H_n = \{1.5^{-k} : 1.5^{-k} \geq 0.02\}$.

Simulations: tvAR(1)

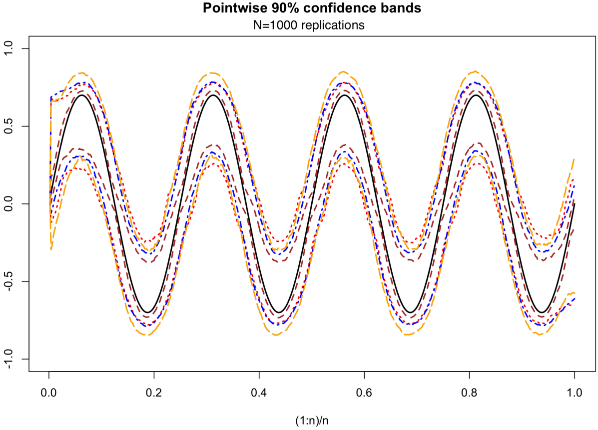
$n = 500$, $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$ with $\varepsilon_t \overset{iid}{\sim} N(0, 0.5^2)$, $\theta_0(u) = 0.7\sin(8\pi u)$.



red: $\hat{\theta}_{\tilde{h}(u)}$ (adaptive bandwidth estimate), brown: $\hat{\theta}_{h^*(u)}$ (Optimal estimate),
black: true θ_0, green: $2\tilde{h}(u) \in H_n = \{1.5^{-k} : 1.5^{-k} \geq 0.002 = n^{-1}\}$.

Simulations: tvAR(1) - Comparison of methods

$n = 500$, $X_{t,n} = \theta_0(t/n)X_{t-1,n} + \varepsilon_t$ with $\varepsilon_t \overset{iid}{\sim} N(0, 0.5^2)$, $\theta_0(u) = 0.7\sin(8\pi u)$.



red: $\hat{\theta}_{\tilde{h}(u)}$ (contrast local), orange: $\hat{\theta}_{\tilde{h}}$ (contrast global),
blue: $\hat{\theta}_{\hat{h}}$ (crossval. global), brown: $\hat{\theta}_{h^*(u)}$ (optimal), black: true θ_0.

Conclusion - Contrast Minimisation

- We proposed an adaptive bandwidth selection procedure via a combination of contrast minimisation and Lepski's method for unknown parameter curves in locally stationary processes
- We conjecture that under mild assumptions on the unknown parameter curve, the proposed estimators are asymptotically minimax in the tvAR(1) case
- Simulations show good behaviour in the case of tvAR processes
- Contrary to the Crossvalidation method, the contrast minimisation method depends on constants C_1, C_2 which have to be chosen suitably.

References

- R. Dahlhaus, W. Polonik (2009): Empirical spectral processes for locally stationary time series, Bernoulli 15(1), 1-39
- O. V. Lepski, E. Mammen, V. G. Spokoiny (1997): Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors, Annals of Statistics, Vol. 25. No. 3, 929-947
- A. Goldenshluger and O. Lepski: Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality, Annals of Statistics, Vol. 39, 1608-1632
- W. Härdle, J. S. Marron (1985): Optimal bandwidth selection in nonparametric regression function estimation, Annals of Statistics, Vol. 13, No. 4, 1465-1481
- O. Arkoun, S. Pergamenchtchikov (2014): Nonparametric estimation for an autoregressive model, arXiv:0806.3012v1
- E. Rio (2009): Moment Inequalities for Sums of Dependent Random Variables under Projective Conditions, J Theor Probab 22: 146-163