Wavelets and their Applications

Yurii Golubev

Wolfgang Härdle

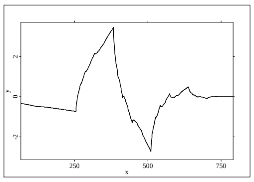
Stefan Sperlich

Sonderforschungsbereich 373

Wirtschaftswissenschaftliche Fakultät

Humboldt–Universität zu Berlin

Germany



Introduction

Orthogonal series expansions of functions, respectively transformations in the frequency domain provide several useful applications:

- alternative way of describing a signal
- provide a parsimonious (approximate) representation
- serve for (nonparametric) estimation

Classical approach:

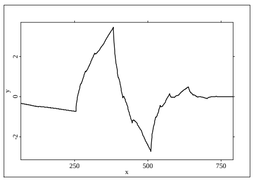
Fourier series expansion

global transform in the frequency domain
=> only good for periodic functions.

Local Fourier transform:

windowed Fourier transform

The window size is not local.



Wavelet transform

offers a simultaneous localization in time and scale domain, (Y.Meyer 1990)

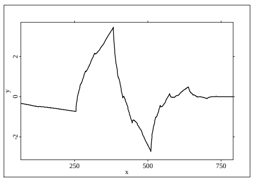
A further advantage is that, using

fast wavelet transform,

it is computationally very fast $= O(n)$.

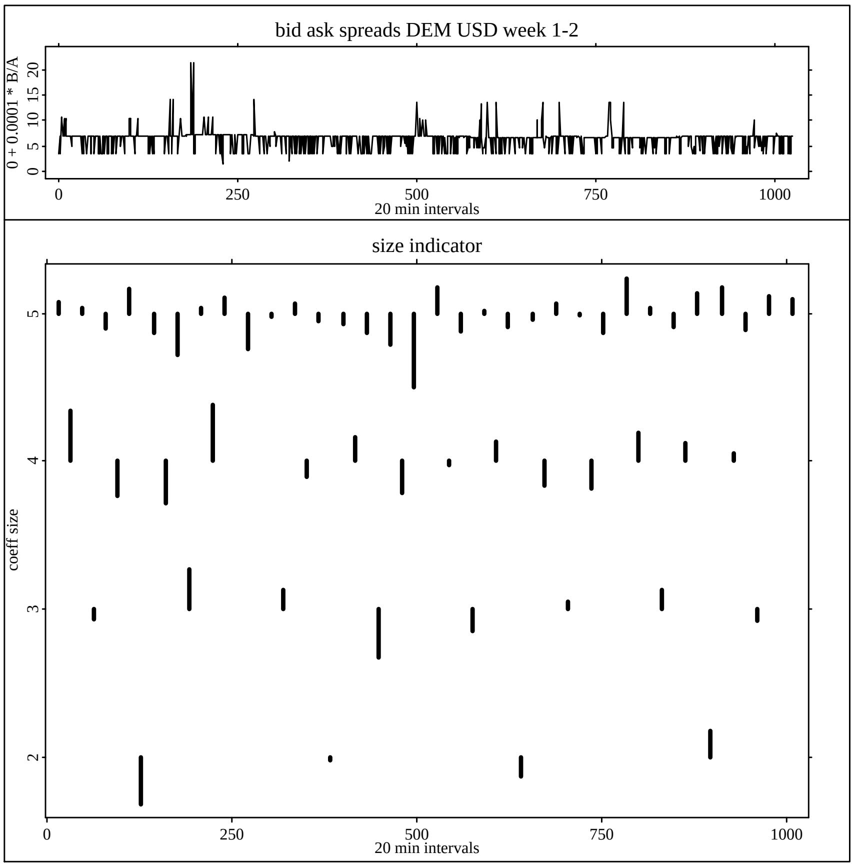
Typical applications are

- data compression
- nonparametric estimation
- smoothing and image denoising
- jump detection and test procedures (e.g. for model specification)

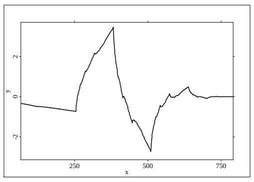


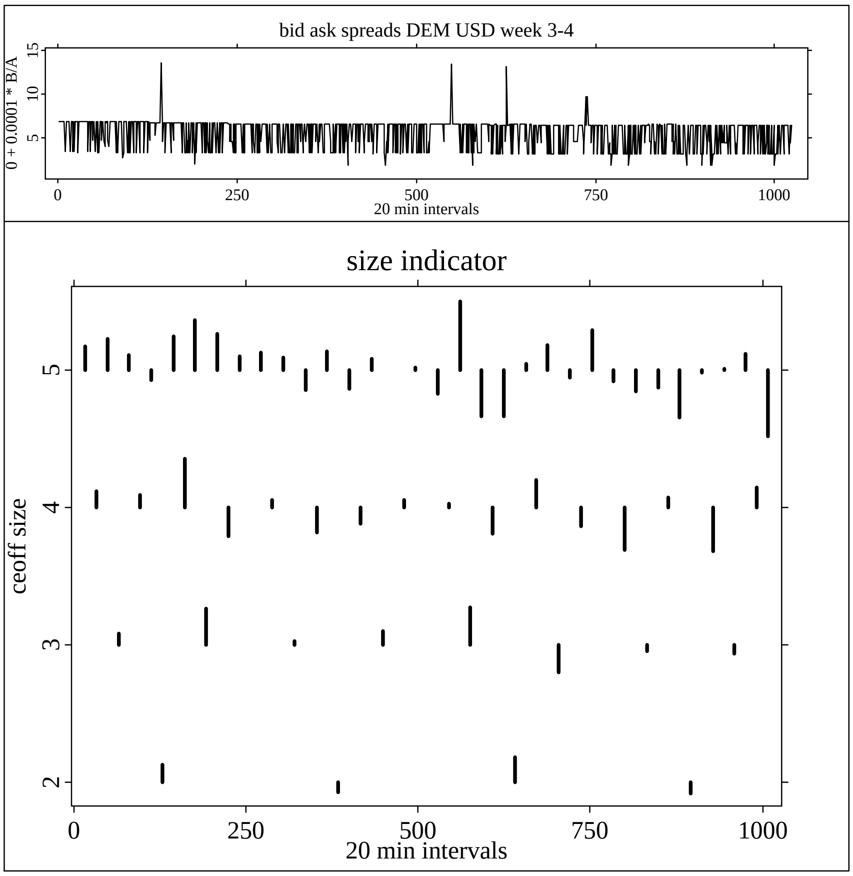
Examples:

1 location - frequency plot of
ln(ask) - ln(bid) spreads for DM / USD



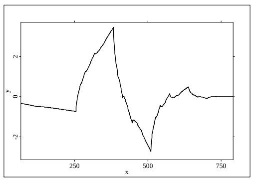
First 2 weeks of the DM/USD FX-rate.



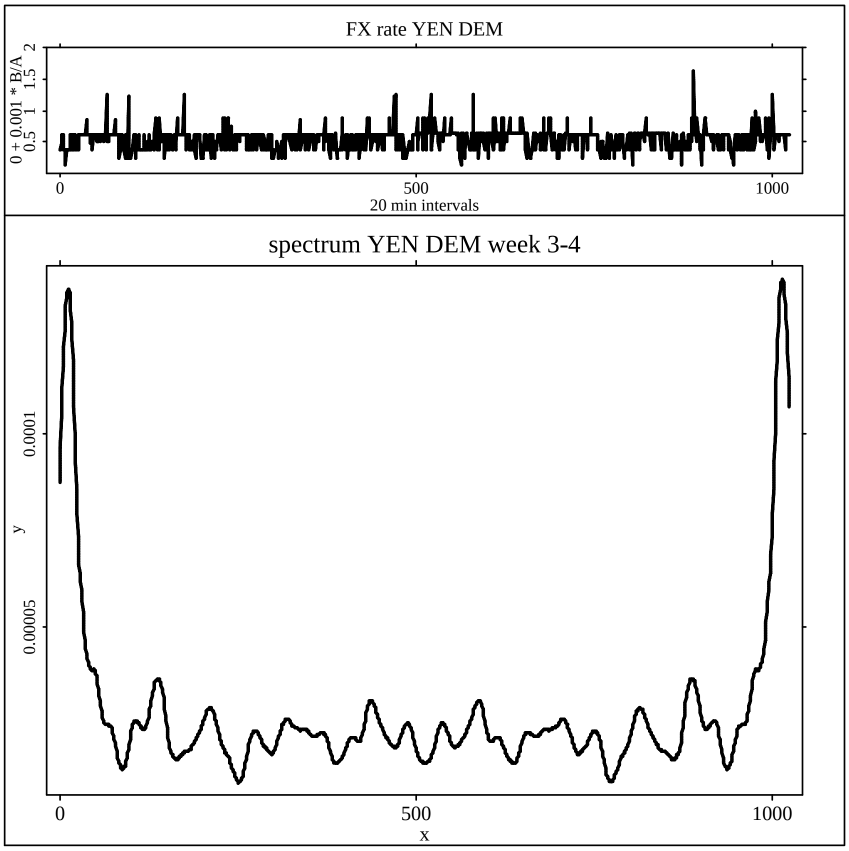


Weeks 3 - 4 of the DM/USD FX-rate.

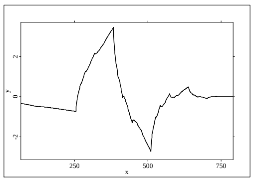
Different time intervals with different activity (compare size for different levels).



2 Periodogram Analysis

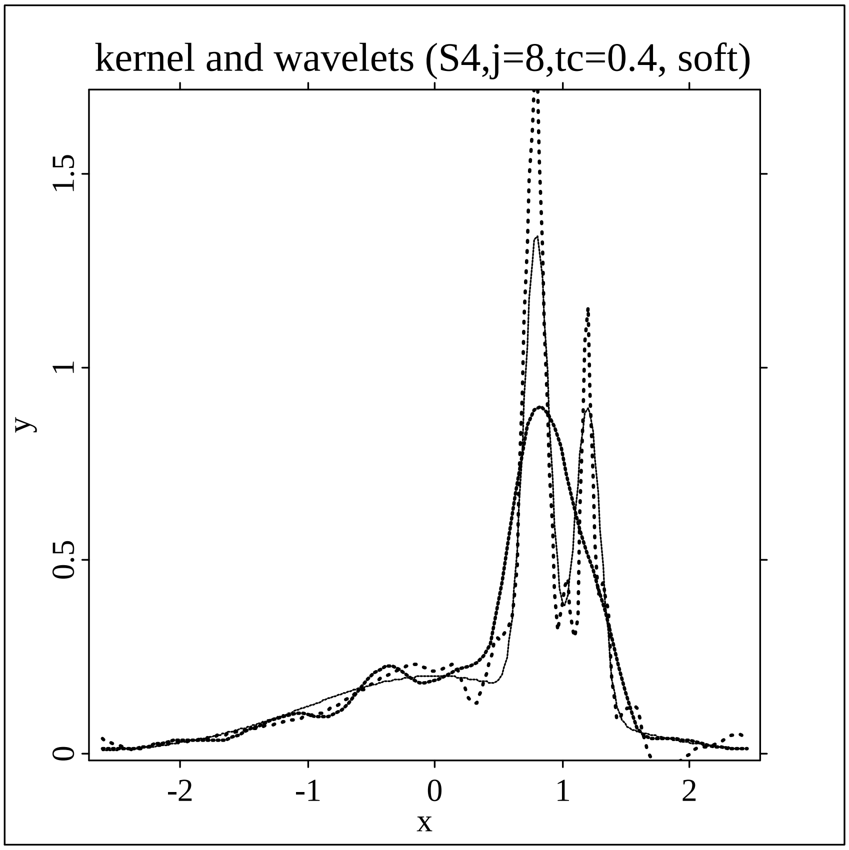


The smoothed periodogram of the YEN / DM series.

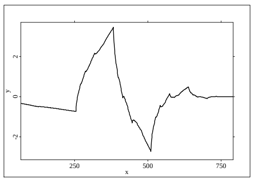


3 Local Smoothing

e.g. for density estimation, comparing wavelet transform with kernel smoothing.



Binned Belgian household data. Wavelet estimate (solid) and Kernel estimate (dotted).



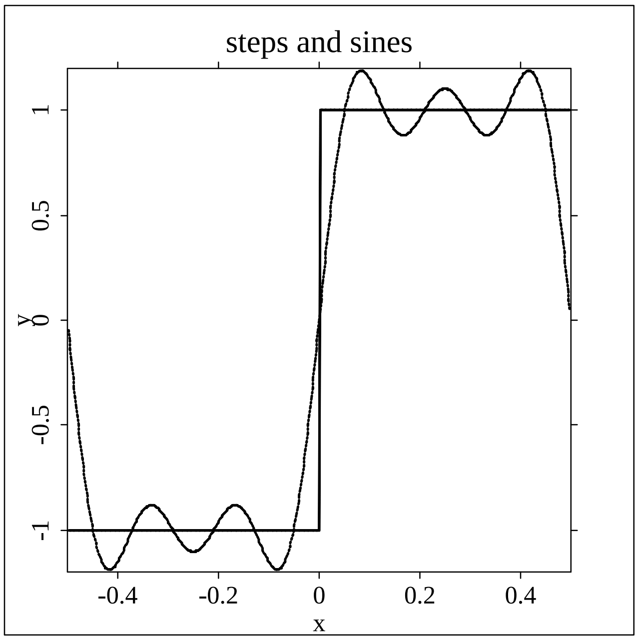
4 Data Compression

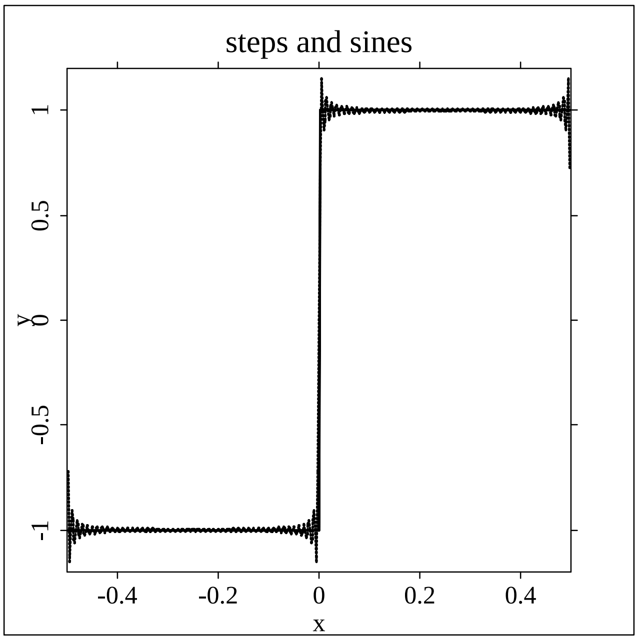
a) with **Fourier Expansion**

$$f(x) = \mathbb{1}\{x \in (0, 0.5]\} - \mathbb{1}\{x \in [-0.5, 0]\}$$

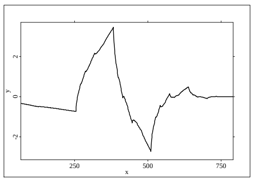
Fourier expansion of $f(x)$ has the form

$$\sum_{k=1,\ \text{odd}}^{\infty} \frac{4}{k\pi} \sin(2\pi k x)$$





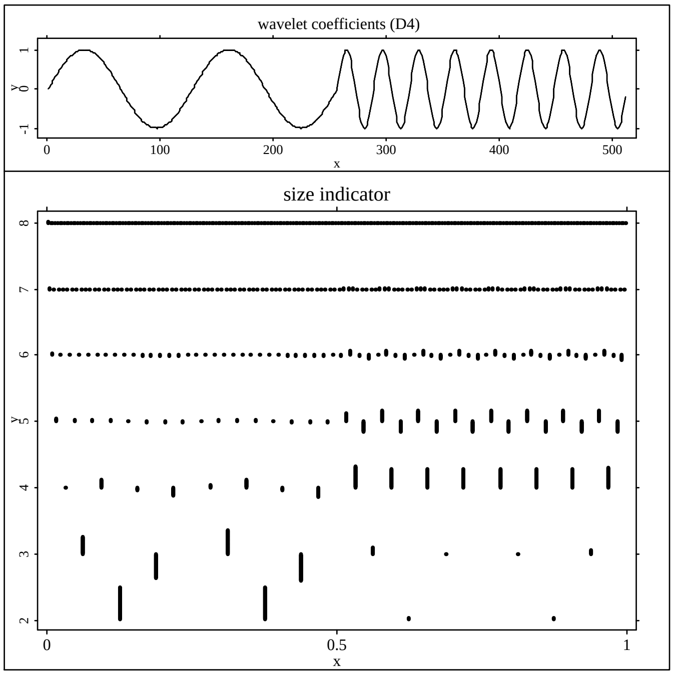
*Step function and Fourier series approximation with **5** terms (left), and with **50** terms (right).*

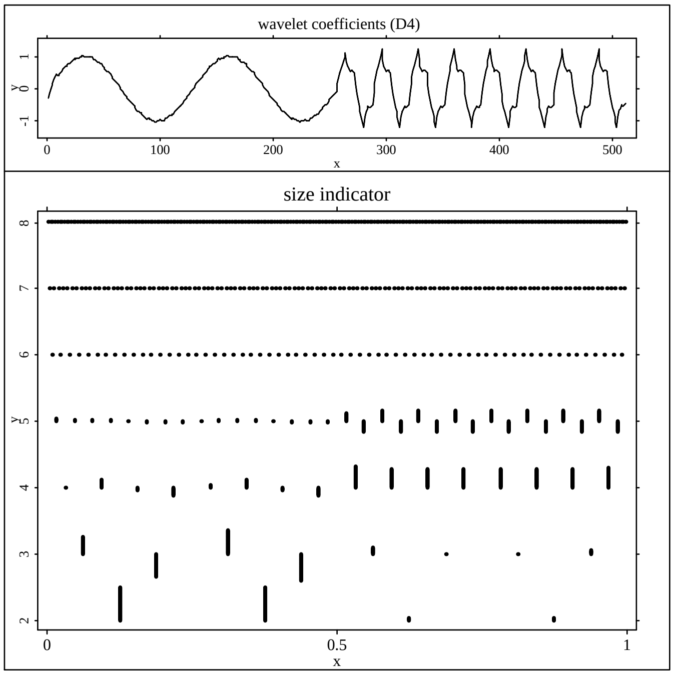


b) with **Wavelet Transform**

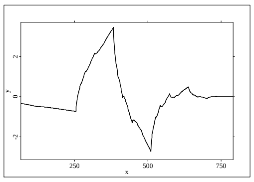
$$f(x) = \sin(8\pi x)\mathbb{1}\{x \in [0, 0.5]\} + \sin(32\pi x)\mathbb{1}\{x \in (0.5, 1]\}$$

We compare function representations by wavelets using different numbers of levels (j_1).





Location -Frequency Plot of Wavelet Transform with different number of levels.



Multiresolution, Father and Mother

Consider $\varphi \in L_2(I\!R)$, so that

$$\{\varphi_{0k}, k \in Z\!\!Z\} = \{\varphi(\cdot - k), k \in Z\!\!Z\}$$

is an ONS and define

$$\varphi_{jk}(x) = 2^{j/2}\varphi(2^j x - k), j, k \in Z\!\!Z \ .$$

The sequence of spaces $\{V_j, j \in Z\!\!Z\}$, spanned by φ_{jk}, is called *multiresolution analysis* of $L_2(I\!R)$ if it satisfies

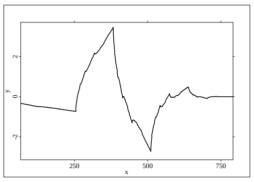
$$V_j \subset V_{j+1} \quad \text{and} \quad \bigcup_{j \geq 0} V_j \text{ is dense in } L_2(I\!R) \ .$$

Further, with a basis $\{\psi_{jk}\}$ of W_j

$$span\{\psi_{jk}\} = W_j = V_{j+1} \ominus V_j$$

we obtain

$$L_2(I\!R) = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j \, .$$



Each function $f \in L_2(I\!R)$ can be expressed by

$$f(x) = \sum_k \alpha_k \varphi_{0k}(x) + \sum_{j=0} \sum_k \beta_{jk} \psi_{jk}(x)$$

with

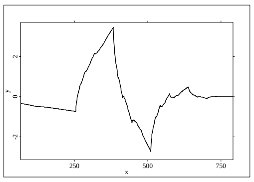
$$\alpha_k = \int f(x) \overline{\varphi_{0k}(x)} dx$$

and

$$\beta_{jk} = \int f(x) \overline{\psi_{jk}(x)} dx$$

respectively.

In the figures only coefficients β_{jk} were presented.



Example for MRA using the Haar basis

Take $\quad \varphi(x) = \mathbb{1}\{x \in (0,1]\}$ (father wavelet).

Construction of a basis:

Set $\varphi_{0k}(x) = \varphi(x-k),\ k \in \mathbb{Z}$, then

$$\varphi_{0k} \text{ is an ONB for}$$

$$V_0 = \{f \in L_2 : f \text{ is constant on } (k, k+1], k \in \mathbb{Z}\}$$

$$V_1 = \{h(x) = f(2x) : f \in V_0\}$$

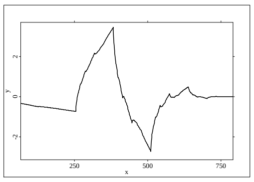
$$V_0 \subset V_1$$

$$\varphi_{1k}(x) = \sqrt{2}\varphi(2x-k)$$

for all $j \in \mathbb{N}$, set

$$V_j = \{h(x) = f(2^j x) : f \in V_0\}$$

Proposition: $\bigcup_{j \geq 0} V_j$ is dense in $L_2(\mathbb{R})$.



Further,

$$W_0 = V_1 \ominus V_0 \quad \text{, s.t.} \quad V_0 \oplus W_0 = V_1$$

$$\Rightarrow \psi(x) = \begin{cases} -1, & x \in [0, 0.5] \\ 1, & x \in (0.5, 1] \end{cases}$$

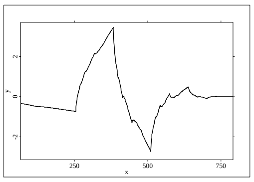
Proposition: The system $\{\varphi_{0k}\}$, where

$$\varphi_{0k}(x) = \varphi(x-k), \quad k \in \mathbb{Z}$$

is an ONB in W_0.
In other terms, W_0 is the linear subspace of $L_2(\mathbb{R})$ which is composed of the functions of the form

$$f(x) = \sum_k c_k \psi(x-k),$$

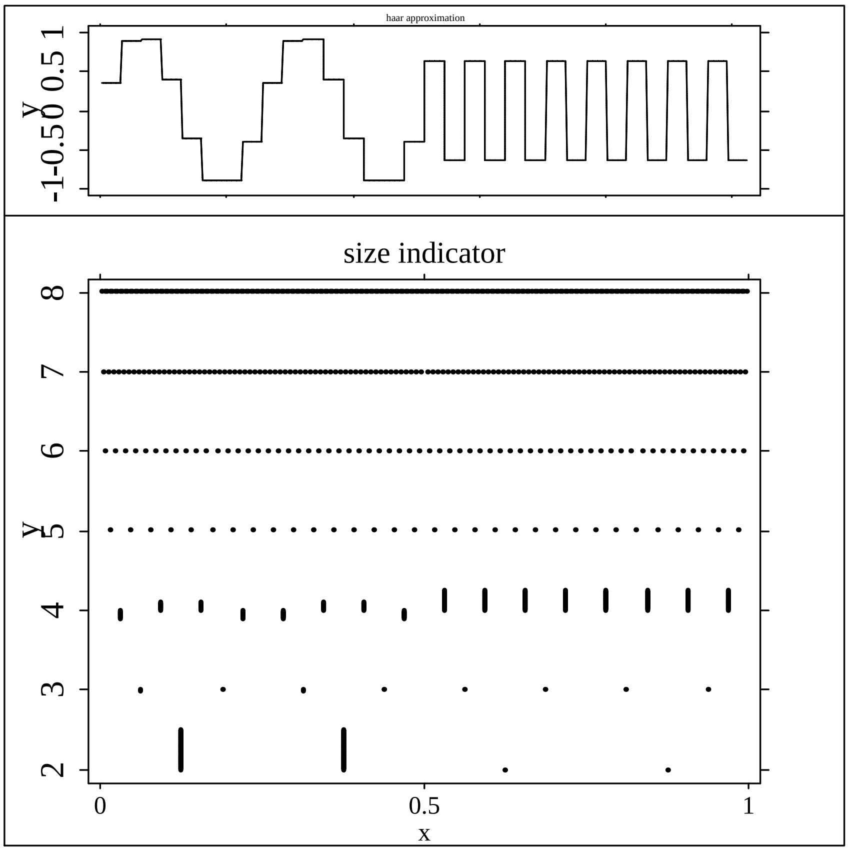
where $\sum_k |c_k|^2 < \infty$, and the series converges in $L_2(\mathbb{R})$.



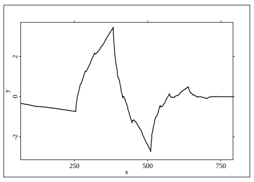
For all $j \in I\!N$ we set

$$V_{j+1} = V_j \oplus W_j$$

$$L_2(I\!R) = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j$$



The sine example with a coarse Haar approximation.



Possible Bases

The construction of φ, ψ for a MRA is rather complicated.

Well known examples are the

Haar basis, with

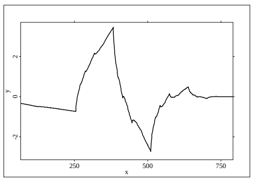
$$\varphi(x) = \mathbb{1}\{x \in (0,1]\}, \qquad \psi(x) = \begin{cases} -1, & x \in [0, 0.5] \\ 1, & x \in (0.5, 1] \end{cases}$$

and the

Shannon basis, with

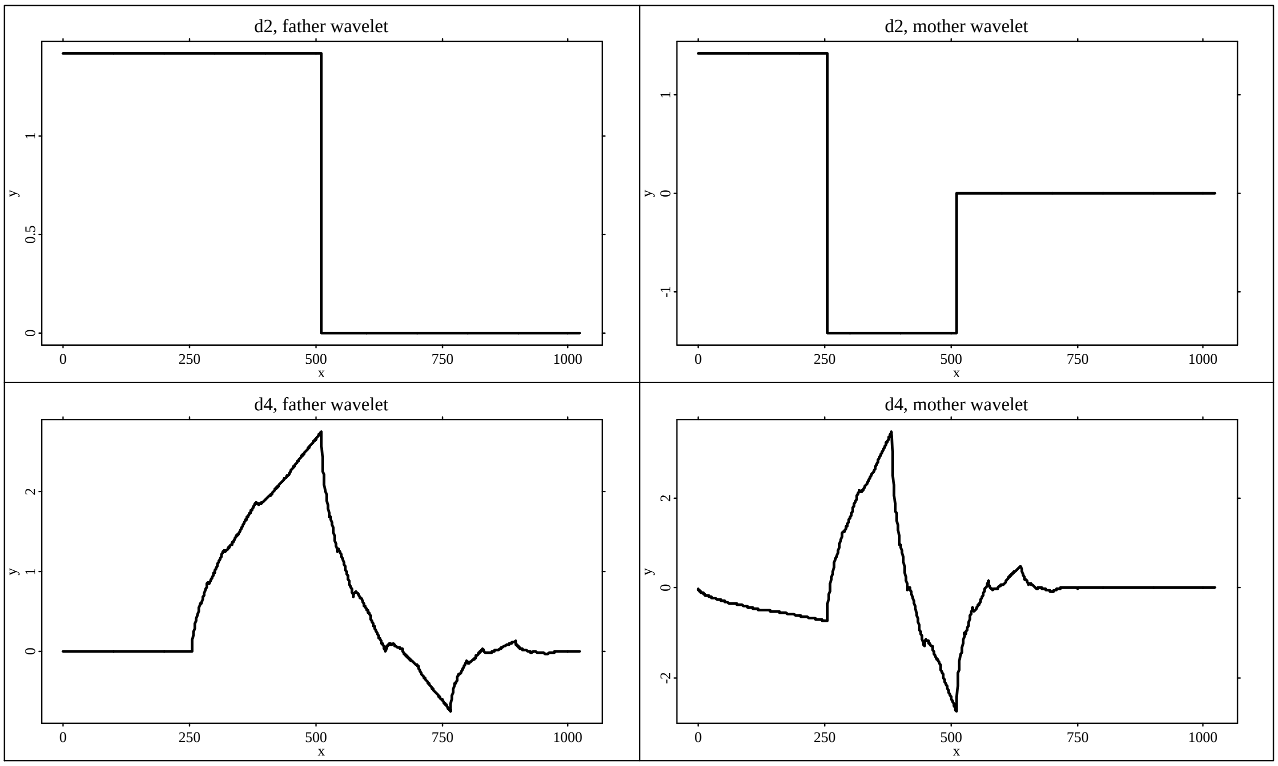
$$\varphi(x) = \sin(2\pi x)/(\pi x), \qquad \psi(x) = \frac{\sin(4\pi x) - \sin(2\pi x)}{(\pi x)}$$

which is the corresponding to the Haar basis in the frequecy domain.

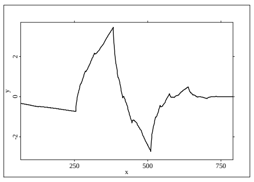


Further bases:

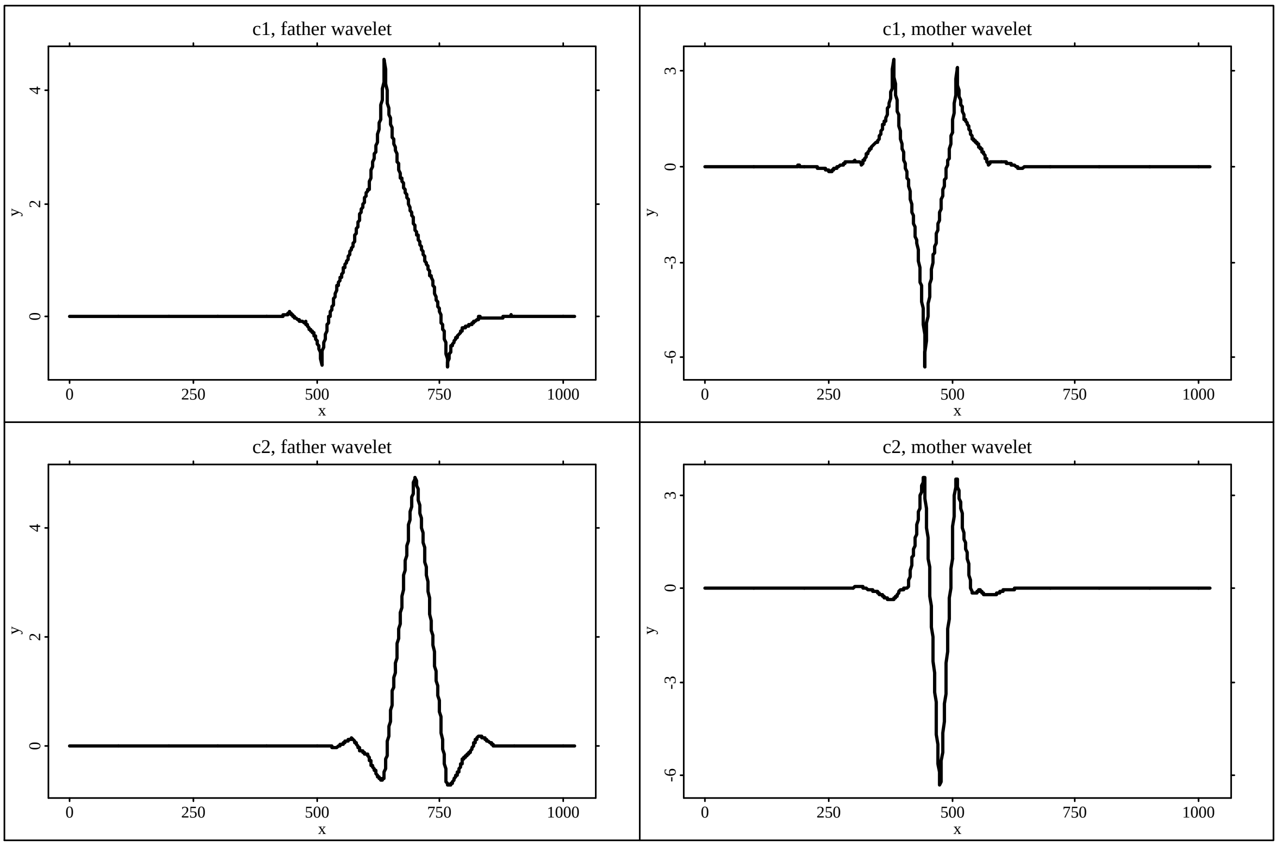
Functional forms of the
Daubechies, father and mother wavelets:



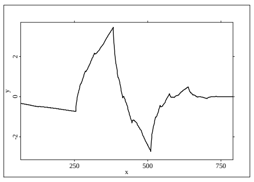
Daubichies' wavelets D2 (upper), D4 (lower), father (left) and mother (right).



Functional forms of the
Coiflets, father and mother wavelets:

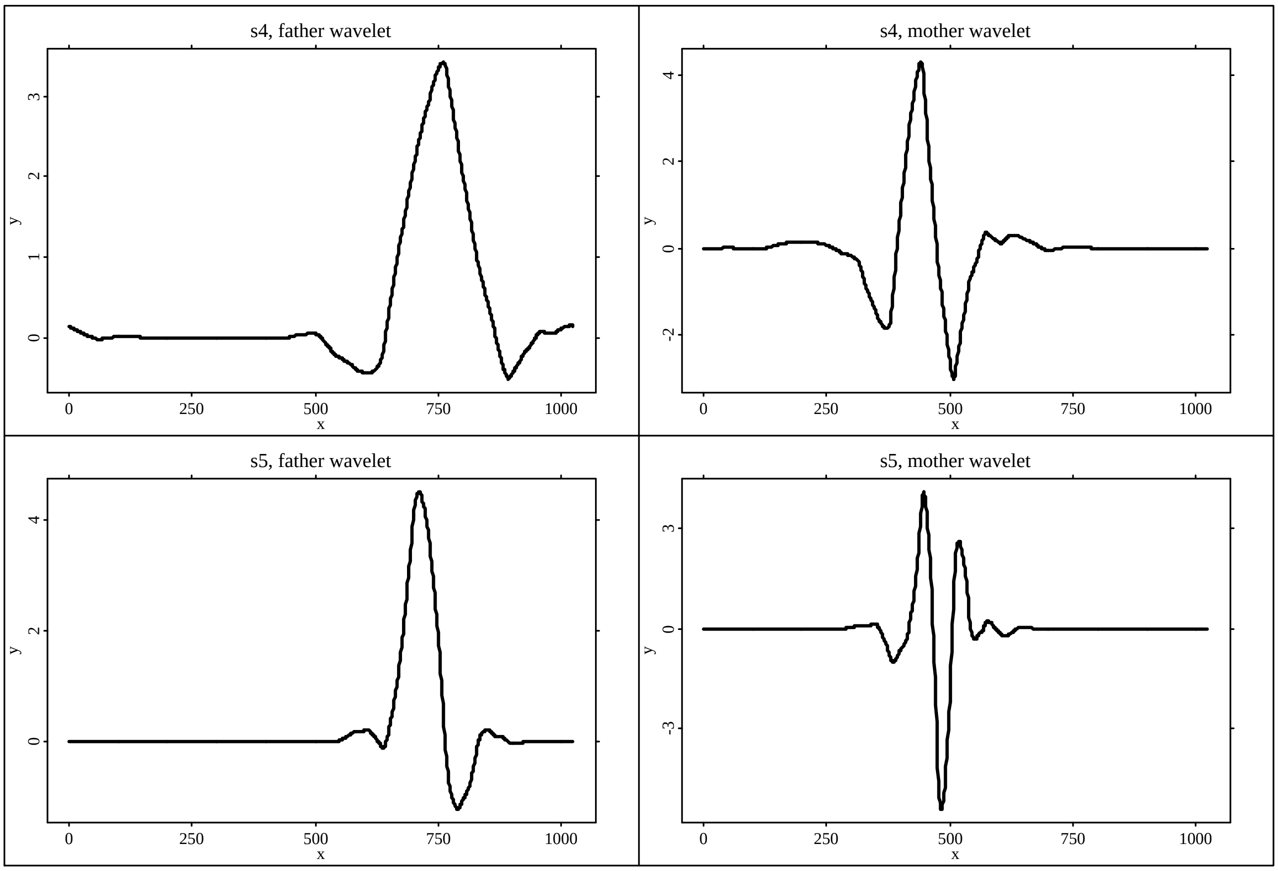


Coiflets, C1 (upper), C2 (lower), father (left) and mother (right).



Functional forms of the

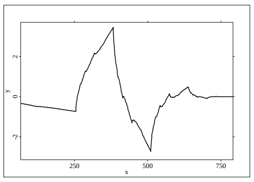
Symmlets, father and mother wavelets:



Symmlet wavelets, S4 (upper), S5 (lower), father (left) and mother (right).

Problem in practice:

proper choice of suitable basis.



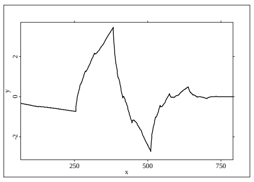
Linear Wavelet Transformation and Thresholding

Parameters which affect the presentation and smoothness are:

- choice of basis
- choice of j_0 and # of coefficients
- choice of *threshold* (cutting of coeffs.)

For choice of basis, compare with earlier slides.

Thresholding and Linear Wavelet Estimation works as follows:



a) Thresholding

is mainly used for

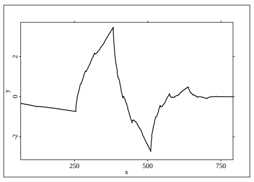
smoothing, denoising and data compression

by suppressing especially small and irregular signals (noise).

First, calculate or determine a threshold t, e.g. $\sigma_x \sqrt{2 \ln(n^2)}$, then

- hard : $\beta = \beta \, \mathbb{1}\{|\beta| > t\}$
- soft : $\beta = sign(\beta) \, (|\beta| - t) \, \mathbb{1}\{|\beta| > t\}$

for all resolution levels.



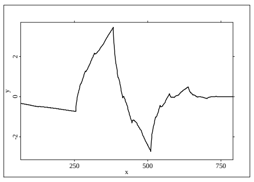
A more data adaptive and level-wise method is to use *Stein*'s unbiased risk estimator (SURE):

- define a risk function depending on t
- minimize SURE with respect to $t \Rightarrow t_{opt}$
- level wise thresholding with t_{opt}

In general (in practice), set the finest level to zero.

Remark:

In kernel estimation, data adaptive smoothing (bandwidth choice) is more complicated. To obtain the same quality there, local adaptive bandwidths have to be calculated.



b) Linear Wavelet Estimation

We consider the choice of j_1,
setting $t = \infty$ (i.e. no thresholding).

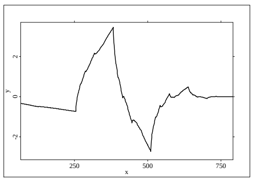
j_1 can be chosen using SURE:
" minimize SURE with respect to j_1 "

Illustrating Example:

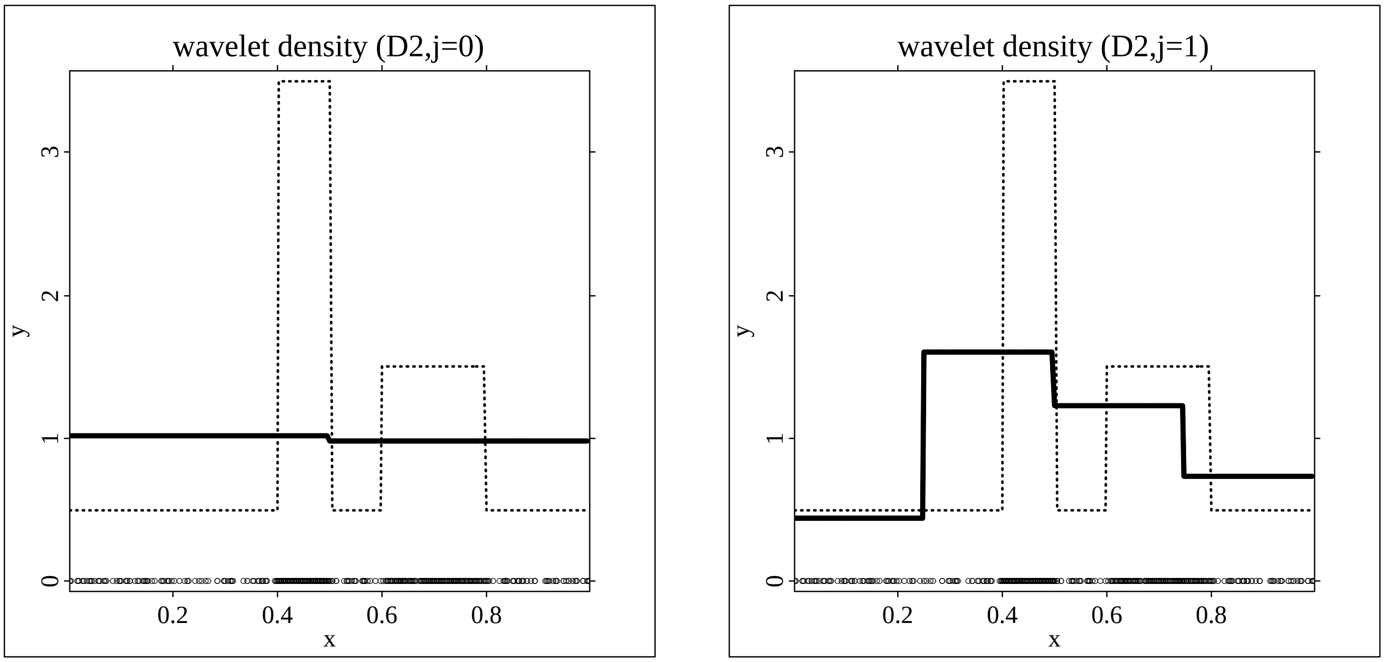
With

$$f(x) = 0.5 \ \mathbb{1}\{x \in [0,1]\} +$$

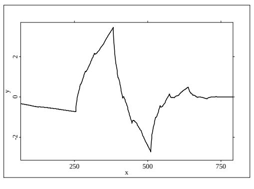
$$+0.3 \ \mathbb{1}\{x \in [0.4, 0.5]\} + 0.2 \ \mathbb{1}\{x \in [0.6, 0.8]\}$$

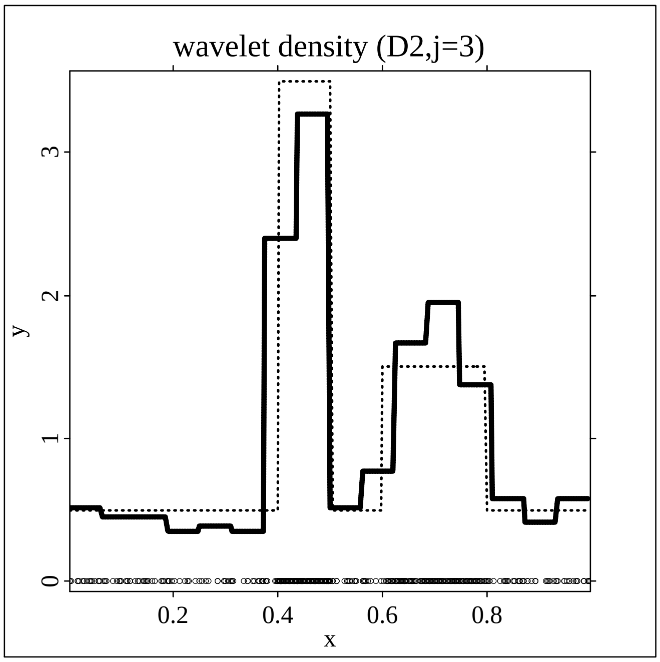


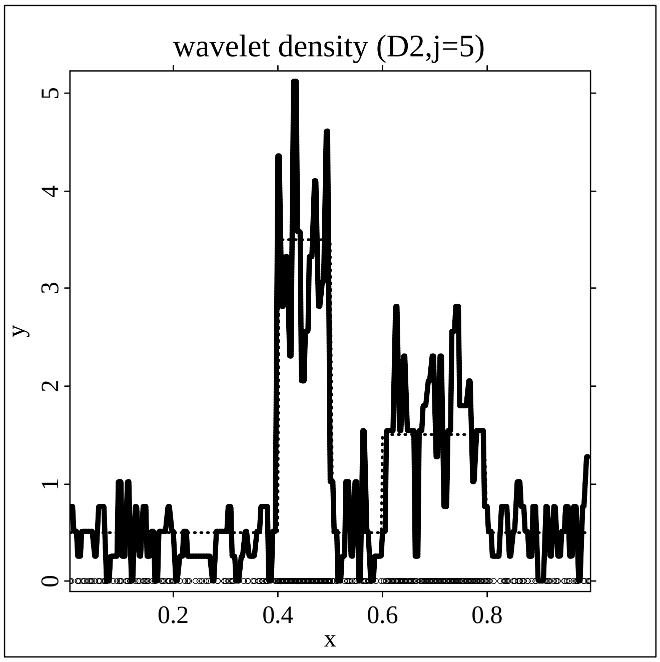
Simulate $n = 500$ observations
via *binning* reduce them to $m = 256$.



Haar wavelet estimates with
$j_1 = 0$ *(left),* $j_1 = 1$ *(right).*







Haar wavelet estimates with
$j_1 = 3$ *(left)*, $j_1 = 5$ *(right).*

As can be seen clearly, we have the typical bias / variance trade–of problem.

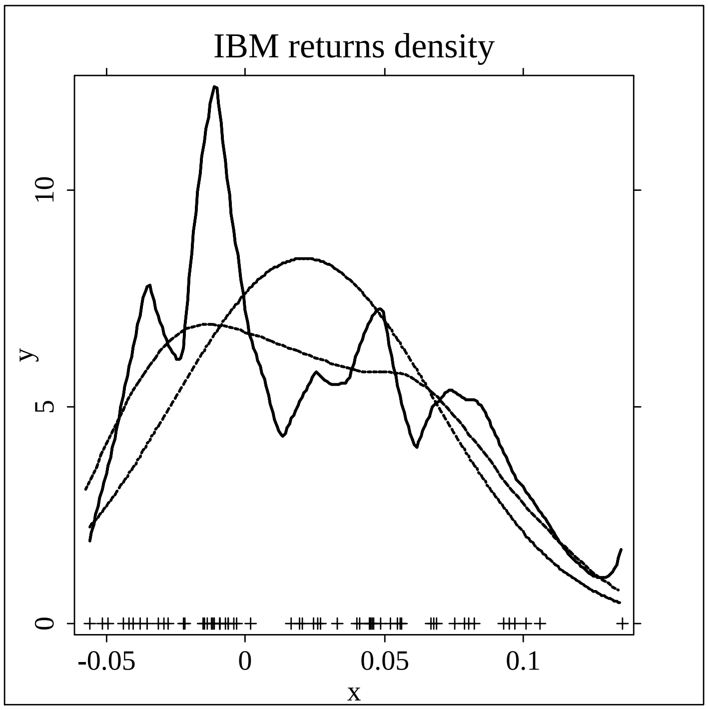
" Optimal " choice of j_1 is around 4.

Example 1:

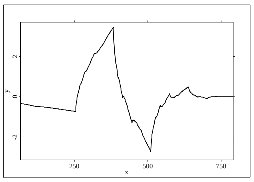
For **density estimation** we have

$$\hat{\alpha}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \varphi_{jk}(X_i), \quad \hat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(X_i).$$

Returns of IBM stocks July 1963 – June 1968.



Density est. of IBM returns. **a)** *normal*, **b)** *kernel*, **c)** *Symmlet4, Soft with* $t = 0.6 \max_{j,k} |\hat{\beta}_{jk}|$



Example 2:

Linear Wavelet Est. vs. Thresholding

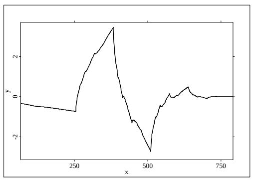
For **regression estimation** we have

$$\hat{\alpha}_{jk} = \frac{1}{n} \sum_{i=1}^{n} Y_i \ \varphi_{jk}(X_i), \qquad \hat{\beta}_{jk} = \frac{1}{n} \sum_{i=1}^{n} Y_i \ \psi_{jk}(X_i).$$

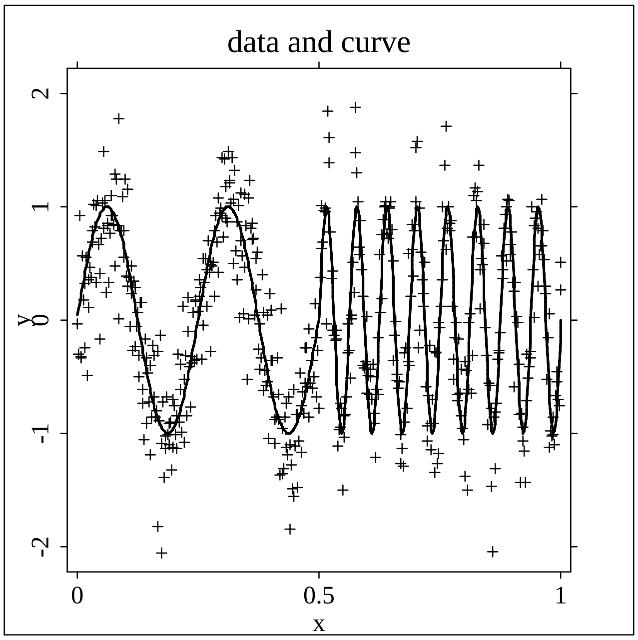
Consider again the *sine–example*:

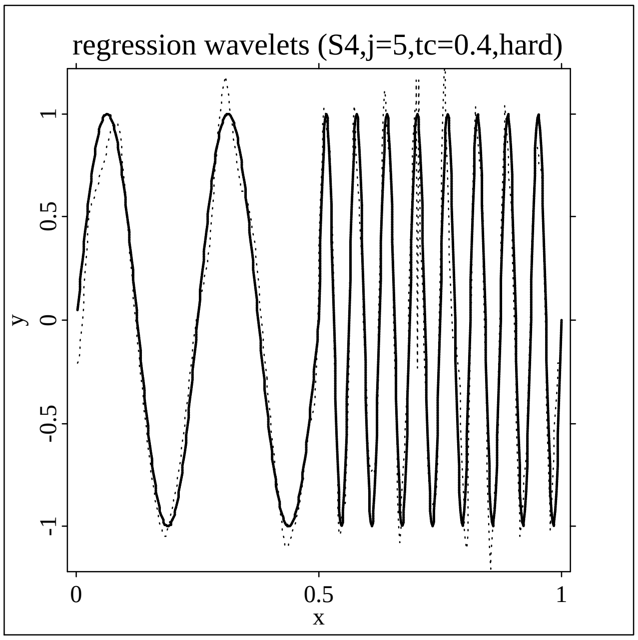
$$f(x) = \sin(8\pi x)\mathbb{1}\{x \in [0, 0.5]\} + \sin(32\pi x)\mathbb{1}\{x \in (0.5, 1]\}$$

Simulate $n = 512$ observations
plus noise $\varepsilon \sim N(0, 0.4^2)$.

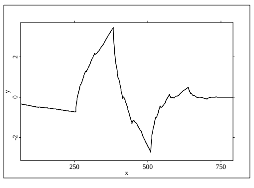


We compare different smoothing procedures:

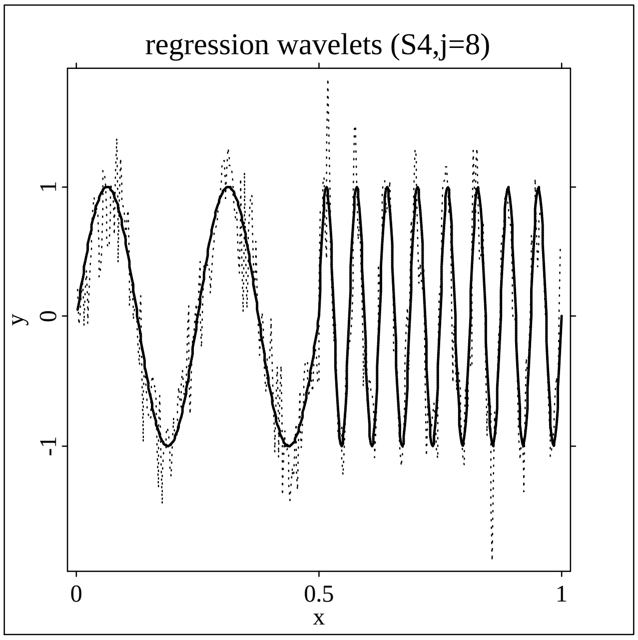


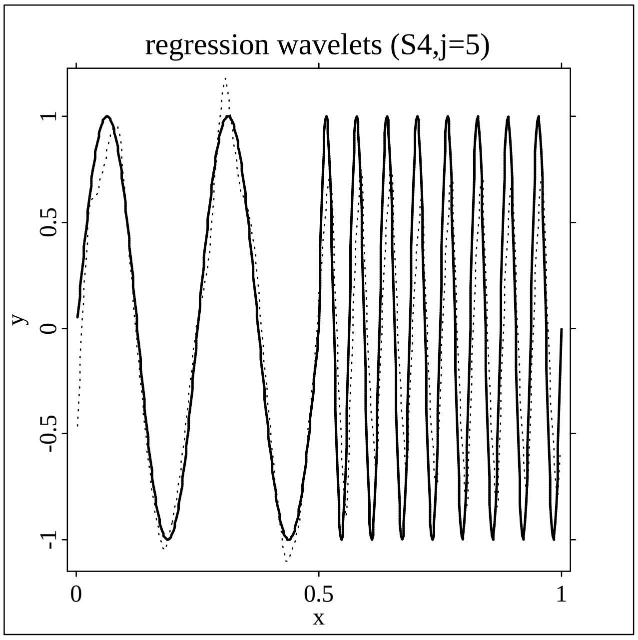


Left: Data and true function
Right: Wavelet smoother
with soft threshold $0.2 \max_{jk} |\hat{\beta}_{jk}|$.

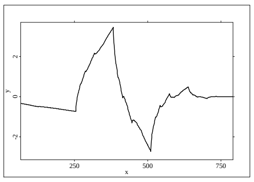


Linear Wavelet estimation ($t = \infty$):





Linear Wavelet Estimator
(using S4) with
$j_1 = 8$ *(left)*, $j_1 = 5$ *(right)*.



Testing Model Specification

(Component Analysis in Additive Models)

Consider regression problem

$$E[Y|X=x] = c + \sum_{j=1}^{d} g_j(x_j), \qquad X \in \mathbb{R}^d,$$

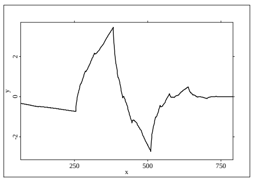
where $c = E[Y]$ is a constant and
$g_k(x_k)$ is the (relative) influence of x_k;
an arbitrary, unknown function for $k = 1, \ldots, d$.

Test problem for input X_1 :

$$\mathrm{H}_0\colon\ g_1(u) = \alpha u\ .$$

Reference:

Härdle, W., Sperlich, S. and V.Spokoiny (1997)
Component Analysis for Additive Models.
DP.52, SFB 373, Humboldt-Universität Berlin,



Idea of the test:

instead of looking at g_1 directly,
consider those wavelet coefficients $\beta_{jk,1}$,
for which under H_0: $\beta_{jk,1} \sim N(0, \rho^2)$
and check for large $\hat{\beta}_{jk,1}$

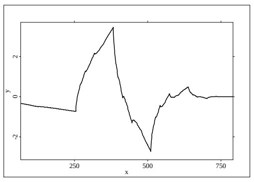
Coefficient estimation under H_0:

1 take the Haar basis and add the linear function for x_1

2 renormalize the new basis

3 for $l = 2, \ldots, d$ take $j_1 = [\log_2(n/d) - 1]$; for $l = 1$ we start with $j_1 = 0$

4 estimate all coefficients $\beta_{jk,l}$, $l = 1, \ldots, d$



We introduce **two methods** to test H_0.

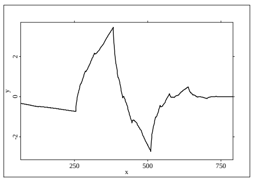
Test **1** looks for local deviation from H_0:

Next steps, for test 1 are:

5 test 1: $\mathbb{1}\{\max_k |\hat{\beta}_{j_1 k,1}|/s_{j_1 k,1} > \lambda_n\}$, where $s_{j_1 k,1}$ is the st.dev. of $\hat{\beta}_{j_1 k,1}$ and $\lambda_n \approx 1 + 2a\sqrt{\ln n}$.

6 if we do not reject and $j_1 < [\log_2(n/d) - 1]$, set $j_1 = j_1 + 1$ and repeat the procedure beginning with step 3.

The scalar a is for adjusting the rejection boundary (the test niveau, respectively).



Test **2** is looking for a deviation from H_0 in L_2 norm:

With $\gamma_{j_1 k,1} = \hat{\beta}_{j_1 k,1}/s_{j_1 k,1}$ and $\Gamma_{j_1} = \sum_k \gamma^2_{j_1 k,1}$ we have a test statistic

$$T_{j_1} = \frac{\Gamma_{j_1} - E(\Gamma_{j_1})}{\sqrt{Var(\Gamma_{j_1})}} = (\Gamma_{j_1} - 2^{j_1})/\sqrt{2 \cdot 2^{j_1}} \ .$$

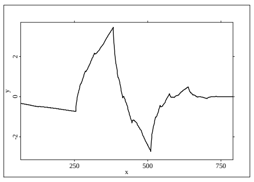
So in step 5 we take test 2: $1\!\!1\{|T_{j_1}| > \lambda_n\}$.

Combining test 1 and 2 we decide as follows:

" reject if test 1 or 2 does so "

Main advantages of the test procedure are

- very weak conditions on the design X
- test has optimal rate up to a *log log* factor, the price for adaptation.



Remarks:

1. Analogous it works for an arbitrary polynomial structure instead of linearity
$\Rightarrow$ can be used to test for significant variables!

2. For small samples λ_n should be approximated level–wise and more exactly by

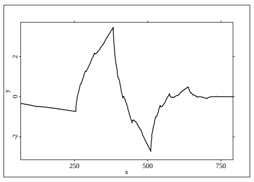
$$\lambda_1(j) = 1 + \sqrt{2\log(2^j) + a_1 \log\log(n)},$$

for test 1, and

$$\lambda_2(j) = 2^{-1/2} + \sqrt{2^j/2n} + \sqrt{a_2 \log\log(n)}$$

for test 2.

3. We could also apply bootstrapping for these test procedures.



Application to Female Labor Supply (in East Germany 1992)

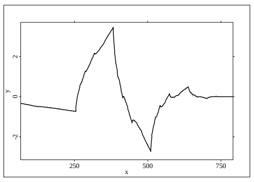
$n = 607$, G-SOEP data

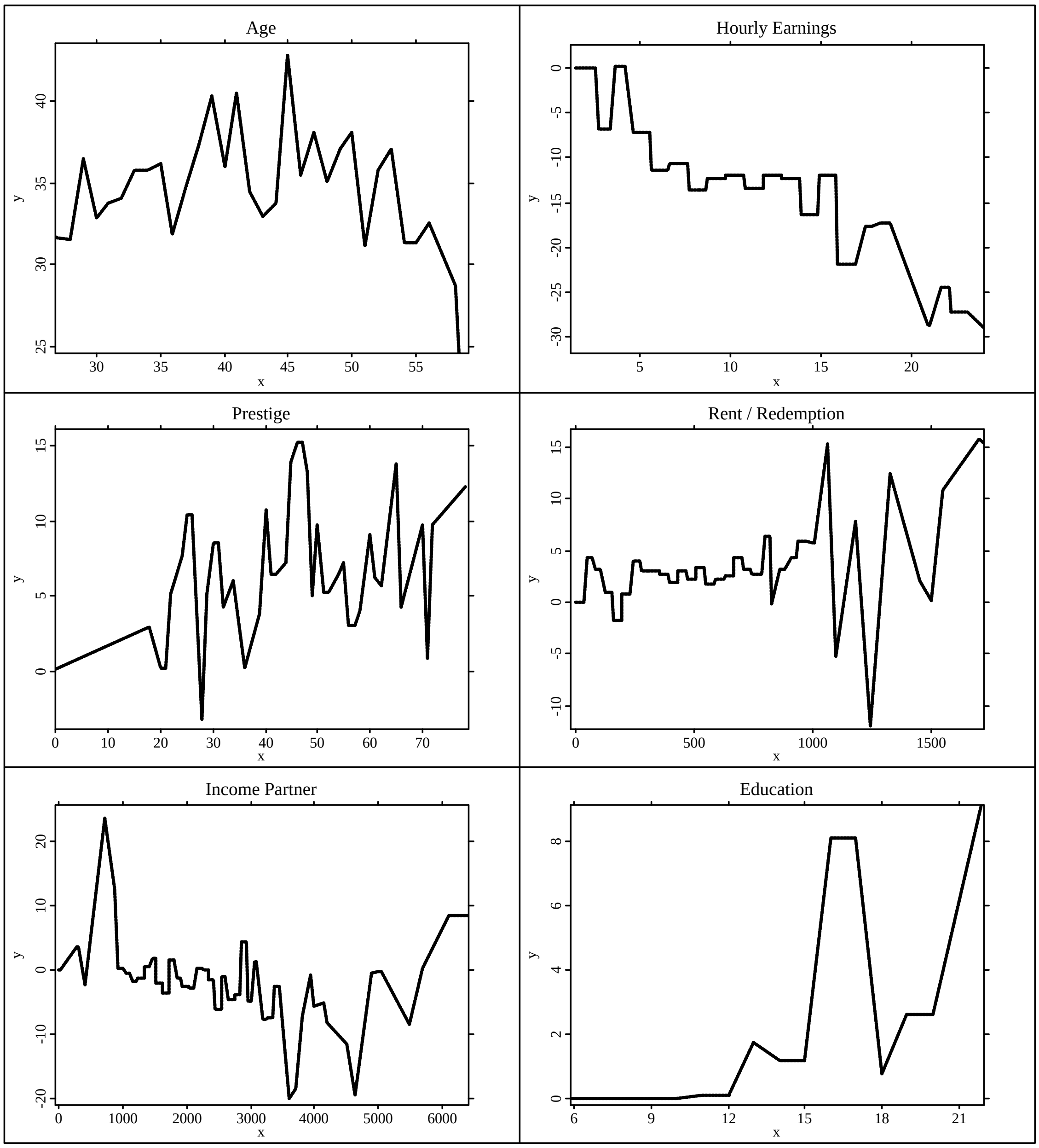
only females with job and partner

Y = weekly number of working hours

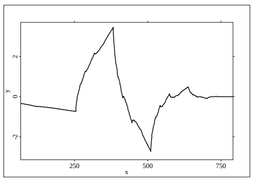
X_1 age of the woman
X_2 earnings per hour
X_3 prestige index of her profession
X_4 rent or redemption for their flat or house
X_5 monthly net income of her partner
X_6 education, measured in years
X_7 the unemployment rate (of country)
X_8 number of children younger than 16 years

(X_7 and X_8 have only 5 realizations).





The estimated functions for the response variables, except for X_7 and X_8.



Model assumed:

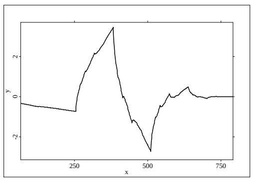
$$Y = c + \sum_{k=1}^{8} g_k(X_k) + \varepsilon$$

We used
$j_1 = 5$ for X_1, X_2, X_3, X_4, X_5
$j_1 = 2$ for X_6, X_7, and
$j_1 = 0$ or 1 for X_8.

Testing $g_1, \ldots, g_6$ for linearity,
setting $a_1 = a_2 = 2$,
the hypothesis has been rejected for

X_1 (age), X_2 (earnings per hour), and
X_3 (prestige index)

(compare also with figures)



Jump Detection

is a further important application,
e.g. for analysing stock returns

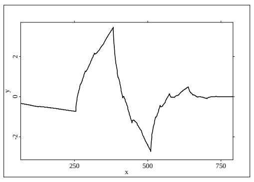
Motivation for our procedure:

Taking the mother Haar wavelet, we estimate (pointwise) the derivative of the underlying function.
Imagine we do this for different bandwidths.
This is to make sure that

  a) every jump gets detected and
  b) the noise does not get overevaluated.

When the derivative is huge, we say *there is a jump*.

The procedure is as follows:



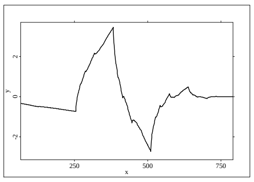
1 *Mirroring* the data, so we avoid problems at the boundaries.

2 Perform the *time invariant wavelet transform*, i.e. we calculate the wavelet transform for all shifted data.
This is very fast $= O(n \ \ln n)$.

3 We have to threshold to suppress the noise,
and further, we are only interested in huge coefficients β_{jk}.
Set $t_j = \sigma_j \sqrt{2 \ln n}$ (level-wise),
and $\sigma_j = \frac{1}{n} \sum_{k=1}^{n} |\beta_{jk}| \sqrt{\pi/2}$.

4 With t_j we set for all k and j

$$g_{jk} = \text{mov.ave.}(\mathbb{1}\{|\beta_{jk}| > t_j\}).$$



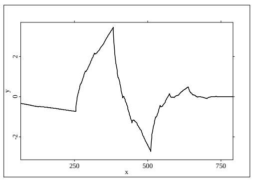
5 We determine clusters (bundles of huge coefficients in a level) by

$$g_{jk} = \begin{cases} 0, & g_{jk} > 0.5 \\ 1, & else \end{cases}$$

Such a cluster indicates a jump.

6 Calculate the number of clusters for each level,
and take the level with the maximal number of clusters.

7 Determine the jump points for each cluster in that level.
This can be done by taking the largest coefficient in a cluster and localize it in time domain.



Library: Generalized Additive Models

See also: intest intestpl gintest gintestpl backfit

Macro: **wavetest**

Description: Additive component analysis in additive separable models using wavelet estimation. An additive component can be tested against a given polynomial form with degree p, e.g. when p is set to zero we test for significant influence of that component. For details see Haerdle, Sperlich, Spokoiny (1997).

Usage: **`erg = wavetest(x,y,p{,dis{,levels{,data{,adjust}}}})`**

Input:

x	n x p matrix , the observed explanatory variable where the 'dis' last columr are expected to be dummy variables
y	n x 1 matrix , the observed response variable
p	scalar , the polynomial degree of the Hypothesis
dis	scalar , the number of dummy variables
levels	(p−dis) x 1 vector, wavelet levels for the components
data	string , if output of the estimates is wished, the name of the output file
adjust	2 x 1 vector , multiplicators for a part of the criticalvalues (0.5 , 1) give abou 5% first kind error, (1, 2.5) give about 1%, default is (1, 1)

Output:

erg	objekt of string, the table of results
data	data file , containing the wavelet estimates