

Time-Series Analysis: Trend Extraction

Evgeny Burnaev

Skoltech, Moscow, Russia

- 1 Trend Extraction: Introduction
- 2 Linear Filtering
- 3 Nonlinear Filtering
- 4 Conclusions

1 Trend Extraction: Introduction

2 Linear Filtering

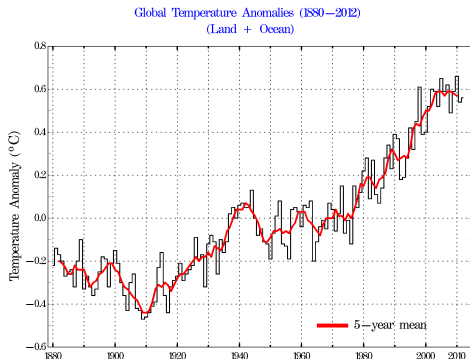
- Moving Average Filters
- Least Squares Filters
- Nonparametric Regression

3 Nonlinear Filtering

- L_1 filtering
- Spectral Methods
- Singular Spectrum Analysis

4 Conclusions

- Trend extraction (filtering) is a major task of time series analysis
- Trend of a time-series is considered to contain the global change, which contrasts with local changes due to noise
- Denoise + track the dynamics of the underlying process



Source: J.E. Hansen, R. Ruedy, M. Sato, and K. Lo
NASA Goddard Institute for Space Studies

Figure – Example of a climatological time-series with trend

- Trend-cycle decomposition = identification of the *permanent* and *transitory* (noise and/or stochastic cycle) stochastic components

$$y_t = x_t + \varepsilon_t,$$

where x_t is a trend, ε_t is a stochastic (or noise) process

- “[...] *the essential idea of trend is that it shall be smooth.*” (Kendall, 1973). In statistical terms

$$\text{Variance}(y_t - y_{t-1}) \gg \text{Variance}(x_t - x_{t-1})$$

- Finite moving average filtering for trend estimation has been used in actuarial science since the beginning of the 20th century
 - Modern theory was developed independently by Andrei Kolmogorov and Norbert Wiener in 1941
 - H. Wold, P. Whittle, R. Kalman, G. Box etc. extensively developed the field extraction methods are applied in different areas: economics, climatology, etc.
-
- We will mainly focus on filtering techniques for trend extraction
 - We will illustrate and compare different methods using S&P 500 index time-series

1 Trend Extraction: Introduction

2 Linear Filtering

- Moving Average Filters
- Least Squares Filters
- Nonparametric Regression

3 Nonlinear Filtering

- L_1 filtering
- Spectral Methods
- Singular Spectrum Analysis

4 Conclusions

- Let $y = \{y_1, \dots, y_m, \dots\}$ be observations y_t in discrete moments $t_i = i\Delta$ (for simplicity we assume that $\Delta = 1$)
- A filtering procedure consists of applying a filter \mathcal{L} to data y

$$\hat{x} = (\hat{x}_1, \dots, \hat{x}_m, \dots) = \mathcal{L}(y)$$

- We consider time invariant and causal filters

$$\hat{x}_t = \sum_{i=0}^{n-1} \mathcal{L}_i y_{t-i}$$

- The well-known Moving Average filter of length n

$$\mathcal{L}_i = \frac{1}{n} 1_{\{i < n\}},$$

i.e.

$$\hat{x}_t = \frac{1}{n} \sum_{i=1}^{n-1} y_{t-i}$$

- The drift

$$\hat{\mu}_t \approx \frac{d}{dt} \hat{x}_t \approx \sum_{i=0}^n \mathfrak{l}_i y_{t-i}$$

where

$$\mathfrak{l}_i = \begin{cases} \mathcal{L}_0 & \text{if } i = 0, \\ \mathcal{L}_i - \mathcal{L}_{i-1} & \text{if } i = 1, \dots, n-1, \\ -\mathcal{L}_{n-1} & \text{if } i = n, \end{cases}$$

- For the Moving Average

$$\mathfrak{l}_i = \frac{1}{n} (\delta_{i,0} - \delta_{i,n}),$$

where $\delta_{i,j}$ is the Kronecker delta

- **Problem:** for $\hat{\mu}_t$ only the first and the last known signal values are used for the estimation

- Improvement of the Uniform Moving Average Filter

$$\mathfrak{l}_i = \frac{4}{n^2} \operatorname{sgn} \left(\frac{n}{2} - i \right) \Leftrightarrow \mathcal{L}_i = \frac{4}{n^2} \left(\frac{n}{2} - \left| i - \frac{n}{2} \right| \right)$$

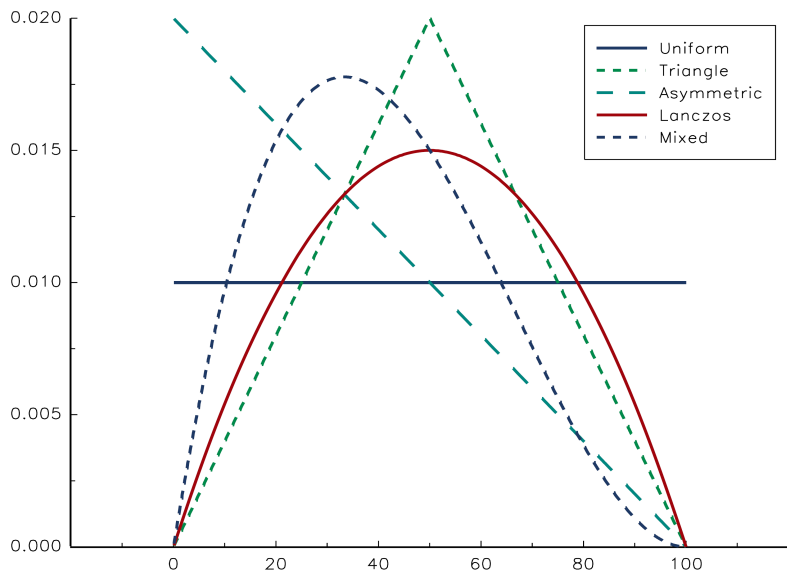
- Assymmetric window function with a triangular form

$$\mathfrak{l}_i = \frac{2}{n} (\delta_i - 1\{i < n\}) \Leftrightarrow \mathcal{L}_i = \frac{2}{n^2} (n - i) 1\{i < n\}$$

- The Lanczos derivative $\frac{d^L}{dt} f(t) = \lim_{h \rightarrow 0} \frac{\sum_{k=-n}^n k f(x+kh)}{2 \sum_{k=1}^n k^2 n}$, so estimating the derivative of the trend at the point $t - n/2$ we get

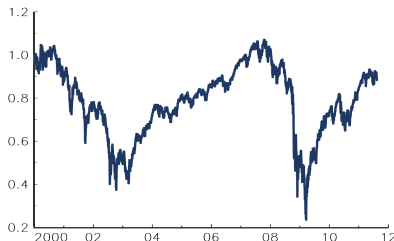
$$\mathfrak{l}_i = \frac{12}{n^3} \left(\frac{n}{2} - i \right) 1\{0 \leq i \leq n\} \Leftrightarrow \mathcal{L}_i = \frac{6}{n^3} (n - i) 1\{0 \leq i \leq n\}$$

Window function \mathcal{L}_i ($n = 100$)

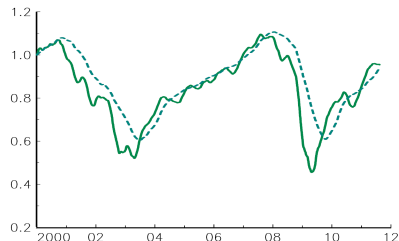


Trend estimate for the *S&P 500* index

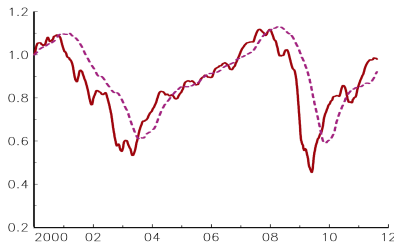
Signal



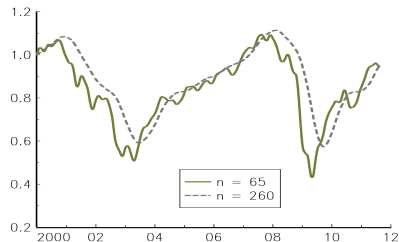
Uniform



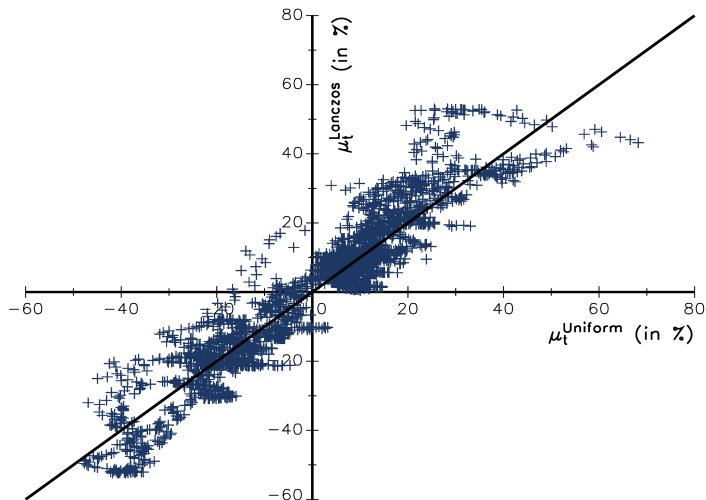
Asymmetric



Lanczos



Correlation between uniform and Lanczos derivatives



1 Trend Extraction: Introduction

2 Linear Filtering

- Moving Average Filters
- Least Squares Filters
- Nonparametric Regression

3 Nonlinear Filtering

- L_1 filtering
- Spectral Methods
- Singular Spectrum Analysis

4 Conclusions

- Least squares methods are often used to define trend estimators:

$$\{\hat{x}_1, \dots, \hat{x}_n\} = \arg \min_{\{\hat{x}_t\}_{t=1}^n} \frac{1}{2} \sum_{t=1}^n (y_t - \hat{x}_t)^2.$$

- The problem is not well-defined \Rightarrow impose restrictions on the underlying process y_t or on the filtered trend \hat{x}_t , e.g.:
 - deterministic constant model $x_t = x_{t-1} + \mu \Leftrightarrow y_t = \mu t + \varepsilon_t$
 - smooth trend condition

$$\frac{1}{2} \sum_{t=1}^n (y_t - \hat{x}_t)^2 + \lambda \sum_{t=2}^{n-1} (\hat{x}_{t+1} - 2\hat{x}_t + \hat{x}_{t-1})^2 \rightarrow \min_{\{\hat{x}_t\}_{t=1}^n}$$

- The objective function is equivalent to

$$\frac{1}{2} \|y - \hat{x}\|_2^2 + \lambda \|D\hat{x}\|_2^2,$$

where $y = (y_1, \dots, y_n)$, $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)$, $D \in \mathbb{R}^{(n-2) \times n}$,

$$D = \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & & \ddots & & \\ & & & & 1 & -2 & 1 \\ & & & & & 1 & -2 & 1 \end{pmatrix}.$$

- The estimator (Hodrick-Prescott filter) is then given by the following solution

$$\hat{x} = (I + 2\lambda D^T D)^{-1} y.$$

- The trend μ_t is a hidden process which follows a given dynamic

$$\begin{cases} y_t = x_t + \sigma_\xi \xi_t, \\ x_t = x_{t-1} + \sigma_\eta \eta_t \end{cases}$$

- We define $\hat{x}_{t|t-1} = \mathbb{E}_{t-1} x_t$, $P_{t|t-1} = \mathbb{E}_{t-1} (\hat{x}_{t|t-1} - x_t)^2$.
- Optimal in L_2 sense estimate is

$$\hat{x}_{t+1|t} = (1 - K_t) \hat{x}_{t|t-1} + K_t y_t,$$

where the Kalman gain

$$K_t = \frac{P_{t|t-1}}{P_{t|t-1} + \sigma_\xi^2},$$

the estimation error is determined by Riccati's equation

$$P_{t+1|t} = P_{t|t-1} + \sigma_\eta^2 - P_{t|t-1} K_t$$

- Riccati's equation gives us the stationary solution

$$P^* = \frac{\sigma_\eta}{2} \left(\sigma_\eta + \sqrt{\sigma_\eta^2 + 4\sigma_\xi^2} \right).$$

- In the long run the filter equation becomes

$$\hat{x}_{t+1|t} = (1 - k)\hat{x}_{t|t-1} + ky_t, \quad k = \frac{2\sigma_\eta}{\sigma_\eta + \sqrt{\sigma_\eta^2 + 4\sigma_\xi^2}}.$$

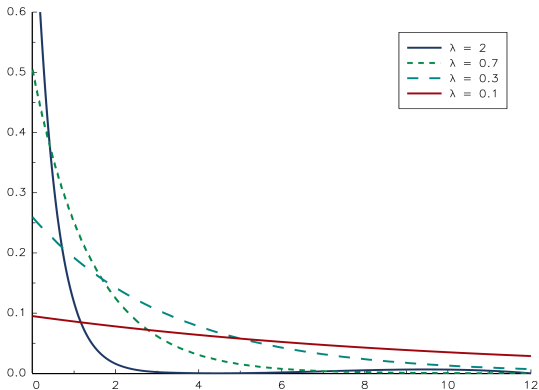
- This Kalman filter can be approximated by an exponential moving average filter with parameter $\lambda = -\log(1 - k)$

$$\hat{x}_t = (1 - e^{-\lambda}) \sum_{i=0}^{\infty} e^{-\lambda i} y_{t-i}$$

and the drift coefficient

$$\hat{\mu}_t = (1 - e^{-\lambda})y_t - (1 - e^{-\lambda})(e^\lambda - 1) \sum_{i=1}^{\infty} e^{-\lambda i} y_{t-i}.$$

- The half-life of this filter is approximately equal to $\lceil (\lambda^{-1} - 2^{-1}) \log 2 \rceil$. For example, the half-life for $\lambda = 5\%$ is 14 days
- Window function \mathcal{L}_i

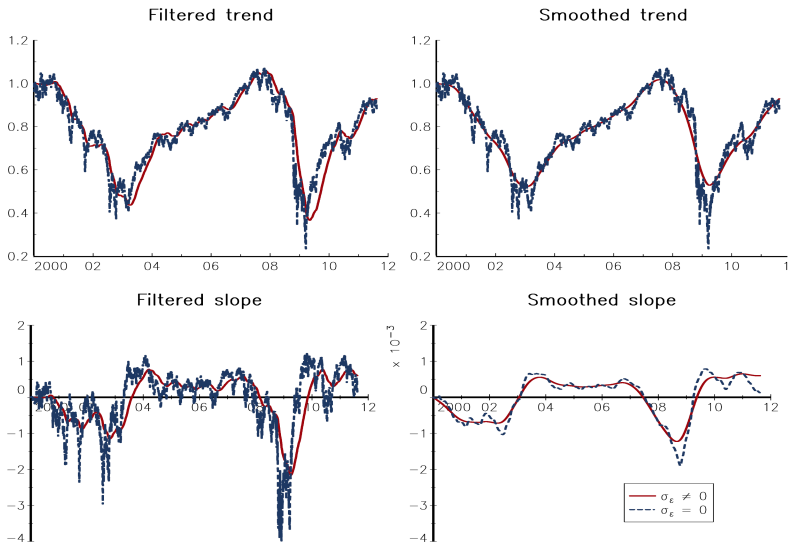


- More general (local linear) trend model (the slope of the trend is stochastic)

$$\begin{cases} y_t = x_t + \sigma_\varepsilon \varepsilon_t, \\ x_t = x_{t-1} + \mu_{t-1} + \sigma_\xi \xi_t, \\ \mu_t = \mu_{t-1} + \sigma_\eta \eta_t. \end{cases}$$

- Remarks
 - The Kalman filter is optimal in the case of the linear Gaussian model
 - Efficient computational solution of the least squares method
 - Can be applied to more sophisticated models
 - Kalman smoother improves the estimate of \hat{x}_{t-i} by using all the information between $t-i$ and t

- The filtered and smoothed components x_t and μ_t
- The Kalman smoother: more information \Rightarrow less noise
- $\sigma_\varepsilon = 0 \Rightarrow$ bigger variance of the trend and slope estimators



1 Trend Extraction: Introduction

2 Linear Filtering

- Moving Average Filters
- Least Squares Filters
- Nonparametric Regression

3 Nonlinear Filtering

- L_1 filtering
- Spectral Methods
- Singular Spectrum Analysis

4 Conclusions

- Parametric model $x_t = \mu t \rightarrow$ Nonparametric model $x_t = f(t)$
- Local polynomial regression

$$y_t = f(t) + \varepsilon_t = \beta_0(t_0) + \sum_{j=1}^p \beta_j(t_0)(t_0 - t)^j + \varepsilon_t$$

- For a given t_0 , a kernel function $K(t)$ and a kernel width h parameters $\beta_j(t_0)$ are estimated as

$$\sum_{t=1}^n \left(y_t - \beta_0(t_0) - \sum_{j=1}^p \beta_j(t_0)(t_0 - t)^j \right)^2 \omega_t \rightarrow \min_{\{\beta_j(t_0)\}_{j=0}^p},$$
$$\omega_t = K\left(\frac{t - t_0}{h}\right)$$

- The trend estimate $\hat{x}_t = \mathbb{E}[y_t | t_0 = t] = \hat{\beta}_0(t)$

- Improvement to the kernel regression: two-stage procedure
- First, fit a local polynomial regression to estimate the residuals $\hat{\varepsilon}_t$
- Second, compute $\delta_t = (1 - u_t^2)1\{|u_t| \leq 1\}$ with

$$u_t = \frac{\hat{\varepsilon}_t}{6 \cdot \text{median}(|\hat{\varepsilon}|)}$$

and run a local polynomial regression with weights $\delta_t \omega_t$

- Spline = C^2 function $S(t)$ which corresponds to a cubic polynomial function on each interval $[t, t + 1)$.
- Let \mathcal{SP} be the set of spline functions
- Spline regression

$$(1 - h) \sum_{t=0}^n \omega_t (y_t - S(t))^2 + h \int_0^T \omega_\tau [S''(\tau)]^2 d\tau,$$

where $T = n \cdot \Delta$

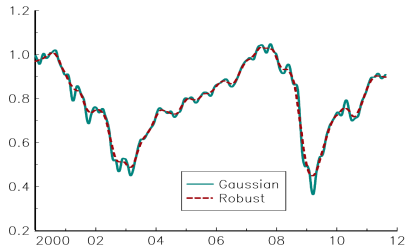
- $h = 0 \Leftrightarrow$ interpolation and $h = 1 \Leftrightarrow$ linear regression

Kernel, Loess and Spline Regression

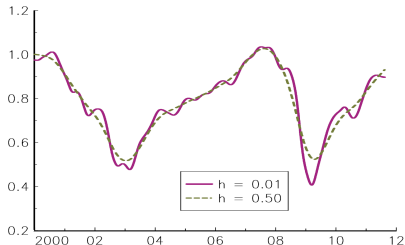
Kernel regression



Loess regression



Spline regression



1 Trend Extraction: Introduction

2 Linear Filtering

- Moving Average Filters
- Least Squares Filters
- Nonparametric Regression

3 Nonlinear Filtering

- L_1 filtering
- Spectral Methods
- Singular Spectrum Analysis

4 Conclusions

- Lasso penalty

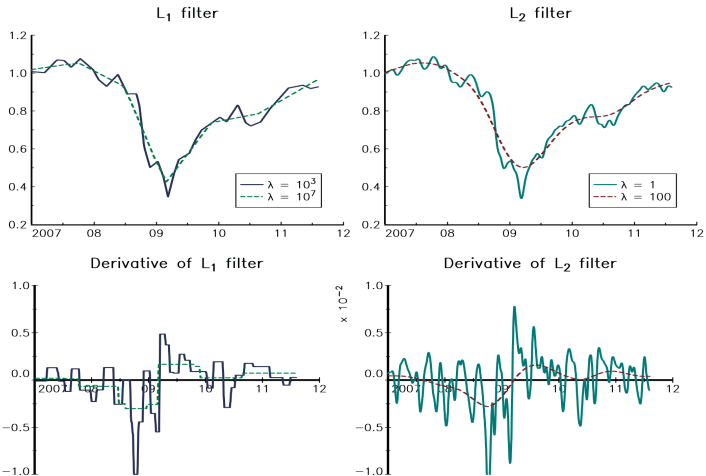
$$\frac{1}{2} \sum_{t=1}^n (y_t - \hat{x}_t)^2 + \lambda \sum_{t=2}^{n-1} |\hat{x}_{t+1} - 2\hat{x}_t + \hat{x}_{t-1}| \rightarrow \min_{\{\hat{x}_t\}_{t=1}^n}$$

- The objective function is equivalent to

$$\frac{1}{2} \|y - \hat{x}\|_2^2 + \lambda \|D\hat{x}\|_1 \rightarrow \min_{\hat{x}}$$

and optimization can be done using primal-dual interior point method

- Filtered signal comprises a set of straight trends and breaks
- Smoothing parameter λ influence on the number of breaks
- It is easy to estimate the slope $\hat{\mu}$



1 Trend Extraction: Introduction

2 Linear Filtering

- Moving Average Filters
- Least Squares Filters
- Nonparametric Regression

3 Nonlinear Filtering

- L_1 filtering
- Spectral Methods
- Singular Spectrum Analysis

4 Conclusions

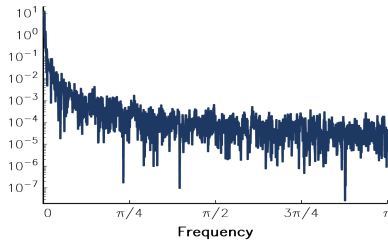
- The trend is “located” in low frequencies
- Fourier transform

$$y(\omega) = \mathcal{F}(y) = \sum_{t=1}^n y_t e^{-i\omega t}$$

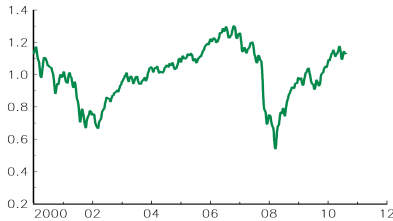
- Remove high frequencies (thresholding) and estimate the trend
 $\hat{x} = \mathcal{F}^{-1}(y(\omega))$
- **Problem:** bad time location for low frequency signals and bad frequency location for the high frequency signals \Rightarrow difficult to localize when the trend reverses (nonstationary/transient process)

- Denoising based on Fourier transform

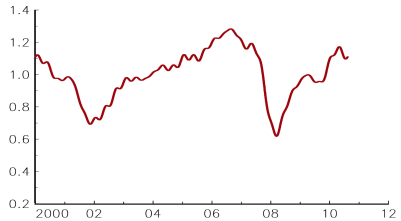
Spectrum of the signal



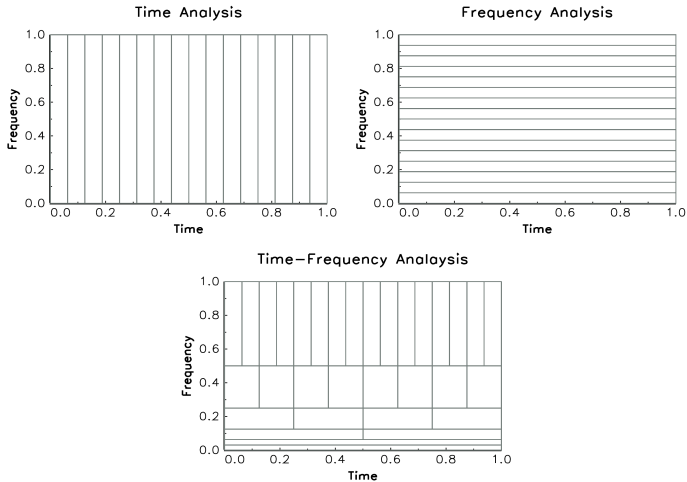
Filtered trend
95% thresholding



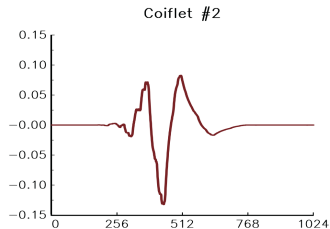
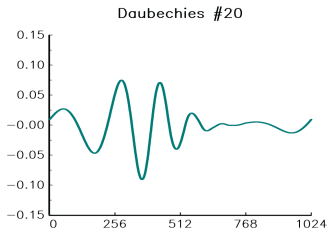
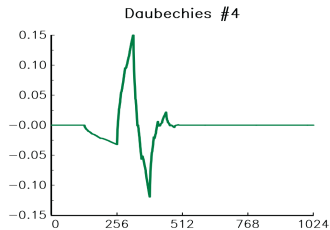
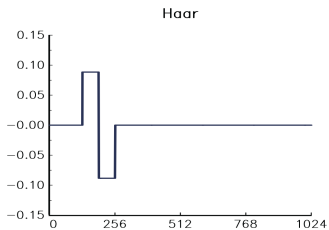
Filtered trend
99% thresholding



- Spectral analysis both in time and frequency



- Localized basis functions



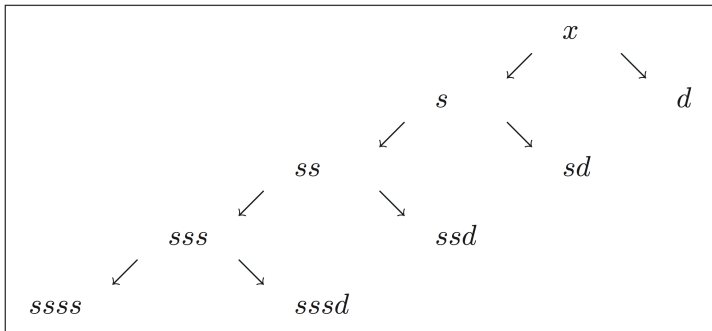
- Compute the wavelet transform

$\omega = (s_0, d_{j,k}, j = \overline{0, J-1}, k = \overline{0, 2^j-1}) = \mathcal{W}(y)$. This corresponds to the representation

$$x(t) = s_0 \phi(t) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(t),$$

$$\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k)$$

- Multiscale representation



- Modify the wavelet coefficients according to the thresholding rule

$$\omega^* = D(\omega),$$

where e.g.

- Hard thresholding

$$\omega_i^* = \omega_i \cdot 1 \{ |\omega_i| > \omega^+ \},$$

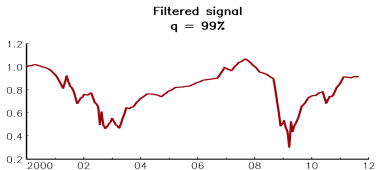
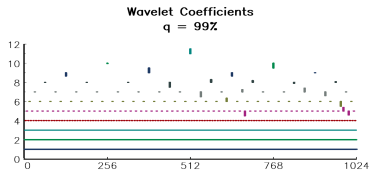
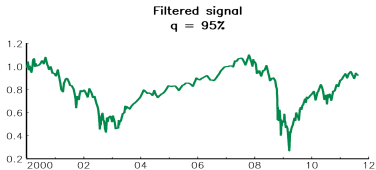
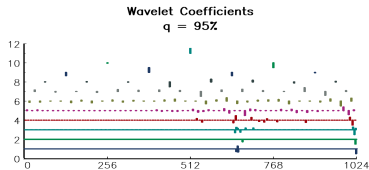
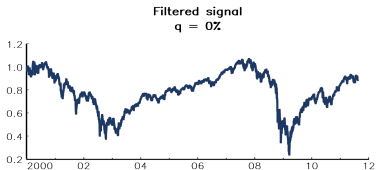
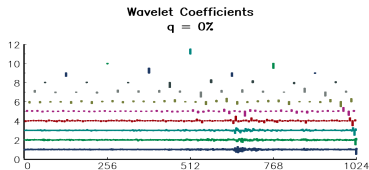
- Soft thresholding

$$\omega_i^* = \text{sign}(\omega_i) \cdot \max \{ |\omega_i| - \omega^+, 0 \}$$

- Apply the inverse wavelet transform

$$\hat{x} = \mathcal{W}^{-1}(\omega^*)$$

- Low-pass and high-pass filters Daubechies 6
- We remove 95% and 99% wavelet coefficients



- Singular Spectrum Analysis
- Empirical Mode Decomposition
- etc.

1 Trend Extraction: Introduction

2 Linear Filtering

- Moving Average Filters
- Least Squares Filters
- Nonparametric Regression

3 Nonlinear Filtering

- L_1 filtering
- Spectral Methods
- Singular Spectrum Analysis

4 Conclusions

- $y = (y_1, \dots, y_t)$ is transformed into Hankel matrix H of the m concatenated lag vectors of y , where the window length $n = t - m + 1$, $m < t/2$

$$H = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_m \\ y_2 & y_3 & y_4 & \cdots & y_{m+1} \\ y_3 & y_4 & y_5 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & y_{t-1} \\ y_n & y_{n+1} & y_{n+2} & \cdots & y_t \end{pmatrix}$$

- The time series can be recovered from the matrix H as

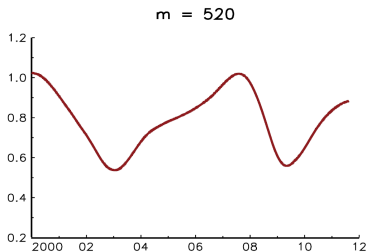
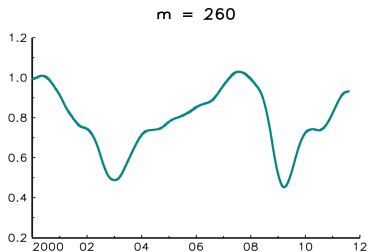
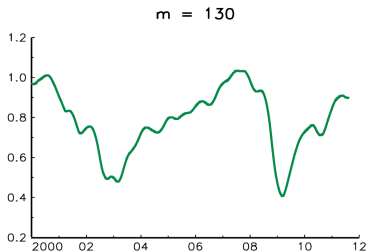
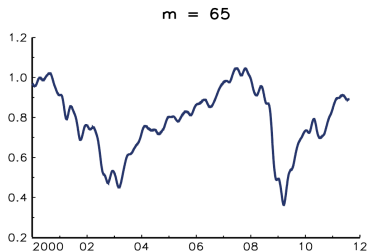
$$y_p = \frac{1}{\alpha_p} \sum_{j=1}^m H^{(i,j)},$$

where $i = p - j + 1$, $0 < i < n + 1$ and

$$\alpha_p = \begin{cases} p & \text{if } p < m, \\ t - p + 1 & \text{if } p > t - m + 1, \\ m & \text{otherwise .} \end{cases}$$

- Let $C = H^T H$ be the covariance matrix of H
- We recover \hat{H} using $k < m$ singular vectors and values
- We remove noise and obtain \hat{x} by recovering it not from H but from \hat{H}

- SSA: Only the first eigenvector is used



1 Trend Extraction: Introduction

2 Linear Filtering

- Moving Average Filters
- Least Squares Filters
- Nonparametric Regression

3 Nonlinear Filtering

- L_1 filtering
- Spectral Methods
- Singular Spectrum Analysis

4 Conclusions

- Very often trend and residual models are defined implicitly by the computational procedure used for a trend extraction
- Selection of the particular model/method is determined by the subsequent usage of the extracted trend:
 - Filtering for ex-post analysis and separation of positive/negative trends
 - Prediction of the future signal values

- **Local Polynomial Regression.** Trend is globally smooth and locally approximated by a polynomial
- **Hodrick-Prescott.** No model
- **SSA.** Large window length m : deterministic (finite rank time series); short m : no model
- **Wavelets.** Semi-parametric, specified by the wavelet

- **Local Polynomial Regression.** A stationary and invertible ARMA process, usually $NID(0, \sigma^2)$
- **Hodrick-Prescott.** No model
- **SSA.** Large window length m : typically a combination of cycle and seasonal components with varying amplitudes plus irregular component; short m : irregular component
- **Wavelets.** Very general, could include a combination of cycles and irregular components

- **Local Polynomial Regression.** Polynomial degree, kernel type, kernel width
- **Hodrick-Prescott.** Regularization coefficient
- **SSA.** Window length m , the number of SVD components
- **Wavelets.** Wavelet basis, levels used for trend reconstruction

- **Local Polynomial Regression.** Pros: fast, simple, a few prespecifications is required. Cons: a residual of a complex structure is not allowed, only seasonally adjusted data
- **Hodrick-Prescott.** Pros: the same as above. Cons: the same as above
- **SSA.** Pros: a few prespecifications, can separate a trend from a complex residual, good for time-seris with a large noise. Cons: few theoretical studies of trend estimators, computational complexity of SVD, small $m \rightarrow$ seasonally adjusted data
- **Wavelets.** Pros: efficients algorithms, many available wavelet bases, good smoothing properties. Cons: subjective choice of levels used for trend reconstruction, boundary effects