Automatic trend extraction and forecasting for a family of time series

Theodore Alexandrov, Nina Golyandina

theo@pdmi.ras.ru, nina@ng1174.spb.edu

St.Petersburg State University, Russia

International Symposium on Forecasting, 12 June 2006

History – Tasks – Advantages

Origins of the “Caterpillar”-SSA approach

- Singular System Analysis *(Broomhead)* Dynamic Systems, method of delays for analysis of attractors [middle of 80’s],
- Singular Spectrum Analysis *(Vautard, Ghil, Fraedrich)*
  Geophysics/meteorology – signal/noise enhancing, signal detection in red noise (Monte Carlo SSA) [90’s],
- “Caterpillar” *(Danilov, Zhigljavsky, Solntsev, Nekrutkin, Golyandina)*
  Principal Component Analysis for time series [end of 90’s]

Tasks

- Additive components extraction and forecast (trends, harmonics, exponential modulated harmonics)
- Smoothing (self-adaptive linear filter)
- Change-point detection
- Handling of missed observations
- Multichannel

Advantages

- Non-parametric and model-free
- Handles non-stationary time series (actual constraints on time series will be described)
- Suits for short time series, robust to noise model etc

More information

- AutoSSA: http://www.pdmi.ras.ru/˜theo/autossa/
- ”Caterpillar”-SSA: http://www.gistatgroup.com/cat/

Goal

- The “Caterpillar”-SSA method works very well in different applications
- Trend extraction is one of its advantages, especially when trend is a finite dimension time series (linear combinations of exponentials, polynomials and harmonics)
- Historically, the part of work is performed manually (visually)

Trend as a slow varying deterministic additive component of a time series

Our goal is to extract a trend automatically by means of the “Caterpillar”-SSA method

"Caterpillar"-SSA basic algorithm

- Decomposes time series into sum of additive components: $F_N = F_N^{(1)} + \ldots + F_N^{(m)}$
- Provides the information about each component

Algorithm

1. Trajectory matrix construction:
   $F_N = (f_0, \ldots, f_{N-1})$, $F_N \to \mathbf{X} \in \mathbb{R}^{L \times K}$
   (L – window length, parameter)

$$\mathbf{X} = \begin{bmatrix} f_0 & f_1 & \ldots & f_{N-L} \\ f_1 & f_2 & \ldots & f_{N-L+1} \\ \vdots & \ddots & \ddots & \vdots \\ f_{L-1} & f_L & \ldots & f_{N-1} \end{bmatrix}$$

2. Singular Value Decomposition (SVD):
   $\mathbf{X} = \sum \mathbf{X}_j$

$$\mathbf{X}_j = \sqrt{\lambda_j} U_j V_j^{\mathsf{T}}$$

λ_j – eigenvalue, U_j – e.vector of $\mathbf{X}\mathbf{X}^{\mathsf{T}}$,
V_j – e.vector of $\mathbf{X}^{\mathsf{T}}\mathbf{X}$, $V_j = \mathbf{X}^{\mathsf{T}} U_j / \sqrt{\lambda_j}$

3. Grouping of SVD components:
   $\{1, \ldots, d\} = \bigoplus I_k$,

$$\mathbf{X}^{(k)} = \sum_{j \in I_k} \mathbf{X}_j$$

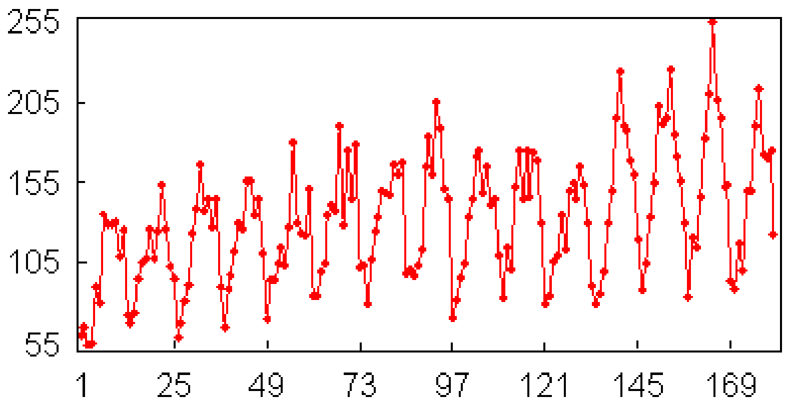
4. Reconstruction by diagonal averaging:
   $\mathbf{X}^{(k)} \to \widetilde{F}_N^{(k)}$

$$F_N = \widetilde{F}_N^{(1)} + \ldots + \widetilde{F}_N^{(m)}$$

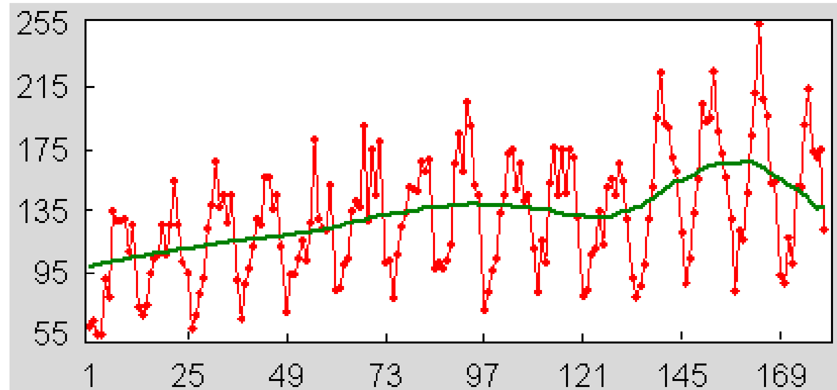
1) Does exist an SVD such that it forms sought for additive component & 2) how to group SVD components?

Example: trend and seasonality extraction

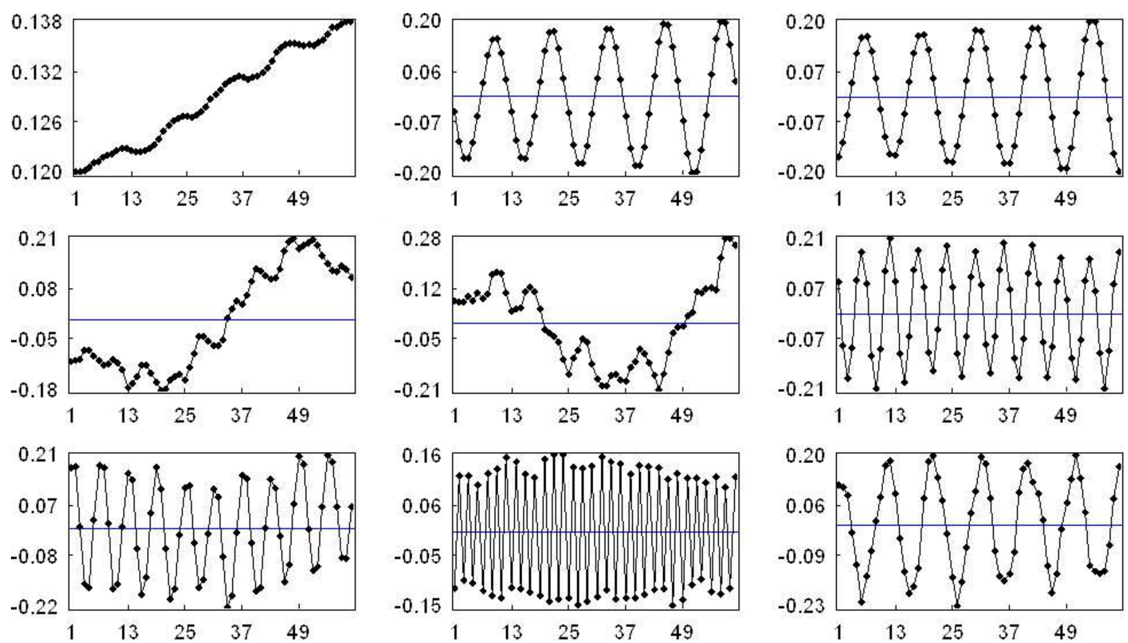
Traffic fatalities. Ontario, monthly, 1960-1974 (Abraham, Redolter. *Stat. Methods for Forecasting*, 1983)



$N = 180, L = 60$



Eigentriples group $I^{(T)} = \{1, 4, 5\}$



Sequences of elements for each of the first eigen vectors $U_i, 1 \leqslant i \leqslant 9$, $U_i = (u_1^{(i)}, \ldots, u_L^{(i)})^\mathsf{T}$

Automation of the choice of eigen triples

Known attempts of automation

- Trend and periodicity extraction: R.Vautard, P.Yiou, M.Ghil, 1992 (SSA-MTM Toolkit and KSpectraToolkit software)
- Auto-denoising in case of big SNR: F.J.Alonso, J.M.Castillo, P.Pintado, 2004 (biomechanical kinematic signals)
- Extraction of generalized cycle components: Izmailov, M.Hai, 2006 (compressors and refrigerators)

Our approach

- Methods for extraction and forecast of an additive components:
  - harmonics, exponential modulated harmonics extraction (based on the ideas of Vautard et al.)
  - trend (slow varying determenistic component)
- Criteria for setting parameters of the methods
- Technique of verification of the methods on given data

Remarks on our approach

- Based on consideration of singular vectors
- Choice of parameters and verification procedure: for a set of time series

Problem statement: processing of a set of time series

$\mathcal{F} = \{F\}$ – a data set of time series of length N

All time series are of this model (it's a general model, not a perametric one!): $F = F^{(T)} + F^{(R)}$, where $F^{(T)}$ is a trend and $F^{(R)}$ is a residual (determenistic, noise)

Problem: extraction and forecast of $F^{(T)}$ for every $F \in \mathcal{F}$

We propose

1. Method of choice of eigentriples
2. Verification of method on the data set
3. Setting of parameters of the method

The item 2) requires similarity of time series of $\mathcal{F}$. The more similar are time series, the more reliable are the results of the verification

This solution inherits the non-parametric nature from the visual “Caterpillar”-SSA method

Method of identification of trend eigentriples

Define the periodogram Π_U of a vector $U = (u_1, \ldots, u_L)^\mathsf{T}$ as

$$\Pi_U^L(k/L) = \frac{L}{2}\begin{cases} 2{c_0}^2, & k = 0, \\ {c_k}^2 + {s_k}^2, & 1 \leqslant k \leqslant \frac{L-1}{2}, \\ 2{c_{L/2}}^2, & L \text{ is even and } k = L/2, \end{cases}$$

where c_k, s_k are the coefficients of Fourier decomposition of elements of the vector U

$$u_n = c_0 + \sum_{1 \leqslant k \leqslant \frac{L-1}{2}} \big(c_k \cos(2\pi nk/L) + s_k \sin(2\pi nk/L)\big) + (-1)^n c_{L/2}$$

Periodogram value $\Pi_U(\omega)$ reflects the contribution of a harmonic with frequency ω into the Fourier decomposition of $u_1, \ldots, u_L$

Method of identification of trend eigentriples

Trend – a slow varying determenistic additive component of time series
Slow varying = harmonics with low frequencies dominate in Fourier decomposition

SVD components corresponding to a trend: their singular vectors $U_i = (u_1^{(i)}, \ldots, u_L^{(i)})^\mathsf{T}$ have slow varying sequences of elements $u_1^{(i)}, \ldots, u_L^{(i)}$ (theoretically proved fact)

The idea of identification is to find all singular vectors with slow varying sequences of elements

For each U we calculate the contribution of harmonics with low frequencies into its F. decomposition:

$$\mathcal{C}(U) = \frac{\sum_{0 \leqslant \omega \leqslant \omega_0} \mathbf{\Pi_U}(\omega)}{\sum_{0 \leqslant \omega \leqslant 0.5} \mathbf{\Pi_U}(\omega)}, \quad \omega \in k/L, k \in \mathbb{Z}.$$

Trend eigentriples $I^{(T)} = \{i : \mathcal{C}(U_i) \geqslant \mathcal{C}_0\}$ for the given $\mathcal{C}_0$

Parameters

- ω_0 – prescribe low frequencies interval $[0, \omega_0]$, harmonics with frequencies from $[0, \omega_0]$ are considered to be slow varying
- $\mathcal{C}_0$ – a threshold, $0 < \mathcal{C}_0 < 1$

Trend extraction for all $F \in \mathcal{F}$ (with verification)

Problem

1. Necessary conditions for using the “Caterpillar”-SSA: finite dimension of trend, moderate SNR, ...
2. Are they fulfilled for all $F \in \mathcal{F}$? Does the procedure extract trends with acceptable quality?

We can verify if the method handles $\mathcal{F}$ by taking (at random) a test subset $\mathcal{T} \subset \mathcal{F}$

Verification

For every time series from the test subset $G \in \mathcal{T}$

1. Manually (visually) extract trend $\widehat{G}^{(T)}$ (we suppose that $\widehat{G}^{(T)} \approxeq G^{(T)}$)
2. Define the trend extracted using the procedure with threshold $\mathcal{C}_0$ as $\widetilde{G}^{(T)}(\mathcal{C}_0)$
   Calculate $\mathcal{C}_0^{\text{opt}} = \arg\min_{\mathcal{C}_0} \big\| \widehat{G}^{(T)} - \widetilde{G}^{(T)}(\mathcal{C}_0) \big\|_{l_2}$, it extracts the trend which is the closest to the manually extracted one.

Estimate quality (on average) of operation of the procedure on $\mathcal{F}$: $\dfrac{1}{\sharp\mathcal{T}} \sum_{G \in \mathcal{T}} \big\| \widehat{G}^{(T)} - \widetilde{G}^{(T)}(\mathcal{C}_0^{\text{opt}}) \big\|_{l_2}$

If it’s small enough we apply the procedure the all $F \in \mathcal{F} \setminus \mathcal{T}$

Trend extraction for all $F \in \mathcal{F}$ (with verification)

Size of the test set $\mathcal{T}$

Size of the test set $\mathcal{T}$ depends on the level of similarity between all F from $\mathcal{F}$

It can be controlled in such a way:

- estimate the width of sampling confidence interval for $\left\| \widehat{G}^{(T)} - \widetilde{G}^{(T)}(\mathcal{C}_0) \right\|_{l_2}$, $G \in \mathcal{T}$,
- if it is small enough then $\mathcal{T}$ is sufficiently large.

Choice of parameters for approximation

Low freq. interval boundary ω_0

Based on our understanding of **low** frequencies (which harmonic is to be considered as slow varying)

- Examining the periodogram of F
- There is a periodicity with period T (besides a trend) $\Rightarrow \omega_0 < 1/T$

Threshold $\mathcal{C}_0$

$\forall F \in \mathcal{F} \quad \mathcal{C}_0^{\text{opt}} = \arg\min_{\mathcal{C}_0} \left\| F^{(T)} - \widetilde{F}^{(T)}(\mathcal{C}_0) \right\|$ but $F^{(T)}$ is unknown

We propose: $exp(\mathcal{R}(\mathcal{C}_0))$ has the same behavior as $\left\| F^{(T)} - \widetilde{F}^{(T)}(\mathcal{C}_0) \right\|$, where

$$\mathcal{R}(\mathcal{C}_0) = \frac{\mathcal{C}\big(F - \widetilde{F}^{(T)}(\mathcal{C}_0)\big)}{\mathcal{C}(F)},$$

$\mathcal{C}(F)$ is the contribution of harmonics with low freq. into Fourier decomposition of F

Because of similar behavior of $\mathcal{R}(\mathcal{C}_0)$ and $\left\| F^{(T)} - \widetilde{F}^{(T)}(\mathcal{C}_0) \right\|$ we can estimate $\mathcal{C}_0^{\text{opt}}$ for F from the $\mathcal{R}(\mathcal{C}_0)$

Forecast

$F = F^{(T)} + F^{(R)}$, $\widetilde{F}^{(T)}$ is the extracted approximation of $F^{(T)}$, $\widetilde{F}^{(T)} \leftrightarrow I^{(T)}$,
($I^{(T)}$ is a group of (trend) eigentriples)

$F^{(T)}$ is *separable* from $F^{(R)}$ (e.g. it holds asymptotically when $F^{(R)}$ is periodical, stochastic noise of arbitrary structure) $\Rightarrow$

$F^{(T)}$ is governed by the linear recurrent formula (LRF) $f_n^{(T)} = \sum_{k=1}^{L-1} a_k f_{n-k}$,

LRF coefficients

Define for $U = (u_1, u_2, \ldots, u_{L-1}, u_L)^\mathsf{T}$: $U^\nabla := (u_1, u_2, \ldots, u_{L-1})^\mathsf{T}$, $\pi_i := u_L$

Then $(a_{L-1}, a_{L-2}, \ldots, a_2, a_1)^\mathsf{T} = \frac{1}{1-\nu^2} \sum_{i \in I^{(T)}} \pi_i U_i^\nabla$, where $\nu^2 := \sum_{i \in I^{(T)}} \pi_i^2$

Forecast of a trend

1. Extract a trend $\widetilde{F}^{(T)}$
2. Find coefficients of the LRF governing $\widetilde{F}^{(T)}$
3. Prolong $\widetilde{F}^{(T)}$ in future recurrently

Thus the problem of trend forecast is reduced to the problem of trend extraction

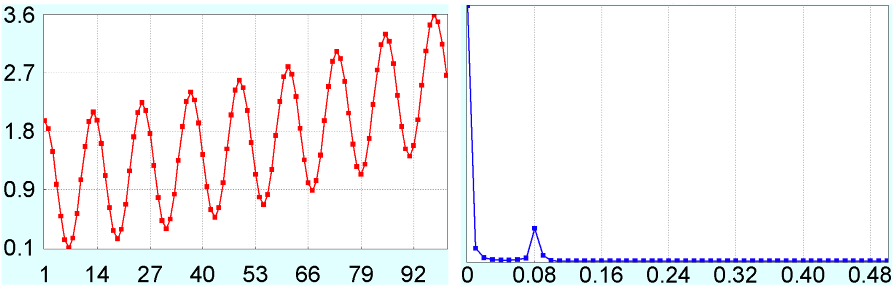
Choice of parameters for forecast

Monte-Carlo simulation of time series of different models (e.g. polynomial trend+gaussian white noise) showed that

the values of $\omega_0, \mathcal{C}_0$ which lead to the best forecast results,
are close to those which give the best approximation $\| F^{(T)} - \widetilde{F}^{(T)}(\mathcal{C}_0) \|$

Thus we can use the described approach based on $\mathcal{R}$

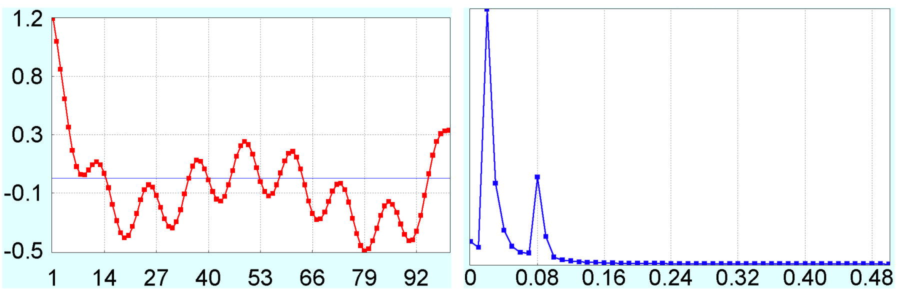
Examples: choice of ω_0



Exp+cos and its periodogram,

$f_n = e^{0.01*n} + \cos(2*\pi*n/12)$

$0.3 < \omega_0 < 0.8$

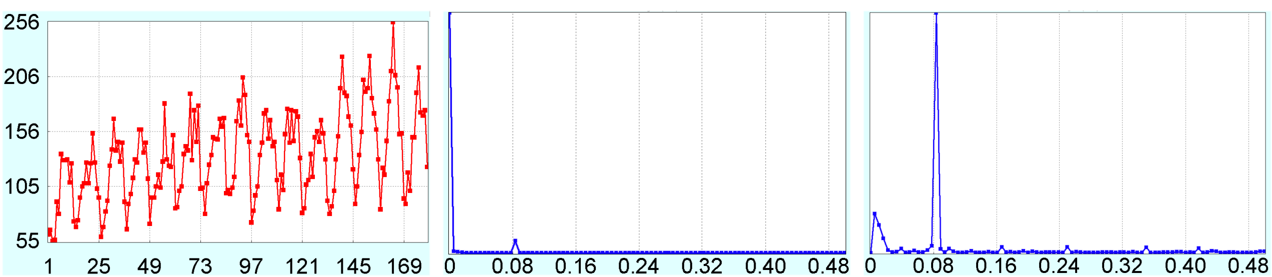


Pn+cos and its periodogram,

$f_n = (x-10)(x-40)(x-60)(x-95)$

$\cos(2*\pi*n/12)$

$\omega_0 \approxeq 0.7 < 0.8$

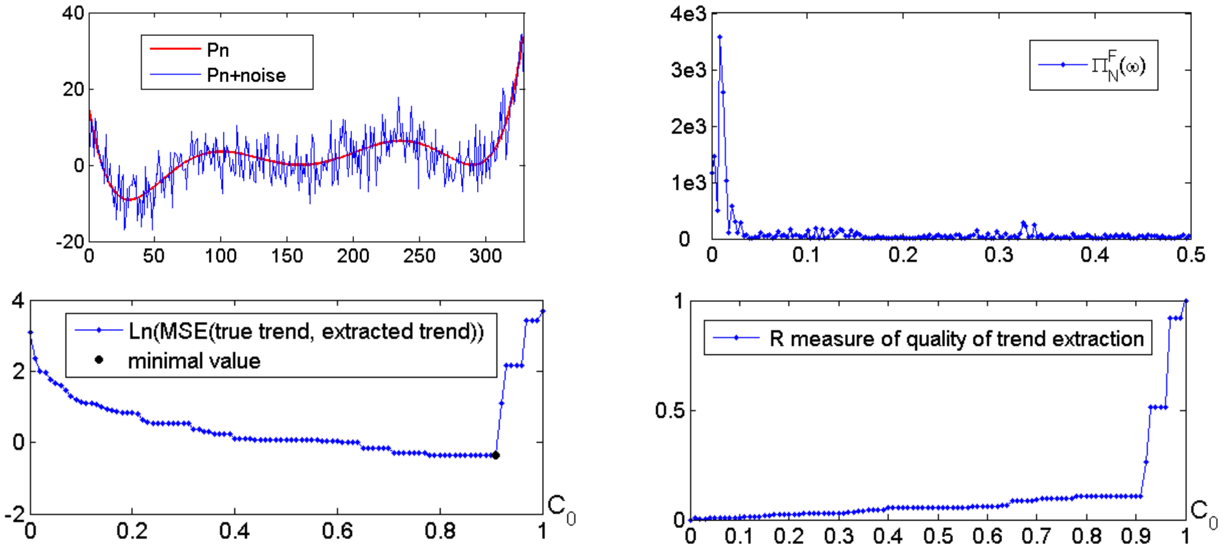


Traffat (left),
its periodogram (center) and
periodogram of normalized
time series (right)

$0.3 < \omega_0 < 0.8$

Examples of $\mathcal{C}_{\text{opt}}$ estimation

Model example, Pn+noise



$f_n = (n-10)(n-70)(n-160)^2 \cdot (n-290)^2/1e11 + N(0,25)$,

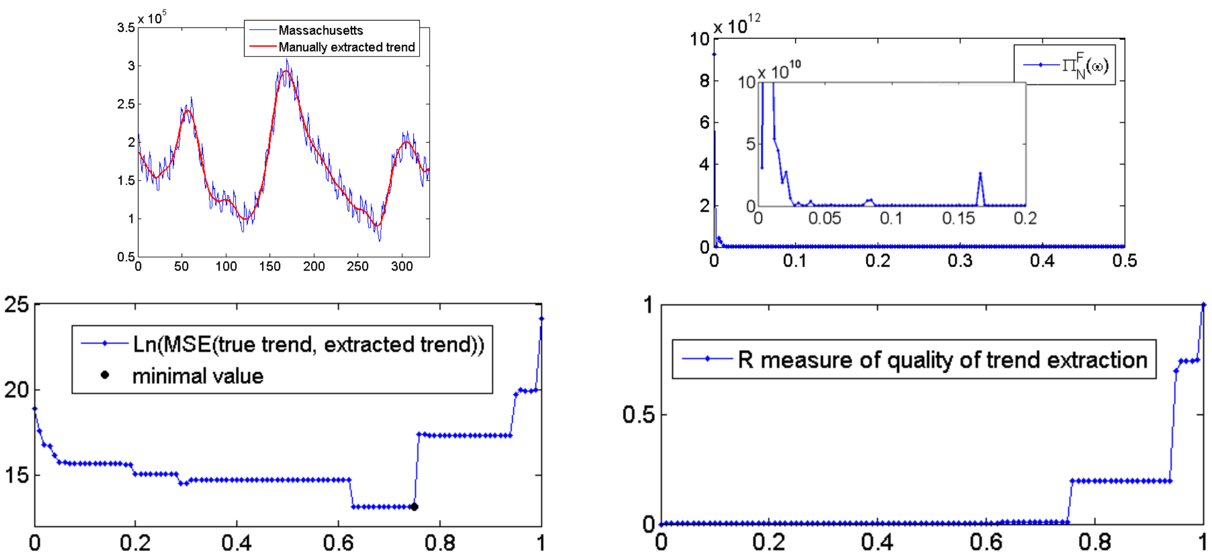
$N = 329$, $L = N/2 = 160$,

$\omega_0 = 0.07$

$\mathcal{C}_0 = 1 \ldots 0.9$: graphics reflect stepwise identification of trend SVD components $\Rightarrow$ considerable changes of MSE$(F^{(T)}, \widetilde{F}_0^{(T)})$

$\mathcal{C}_{\text{opt}} < 0.9 (\approx 0.9)$

Real-life example, Massachusetts unemployment



Massachusetts unemployment (thousands, monthly), from economagic.com

$N = 331$, $L = N/2 = 156$,

$\omega_0 = 0.05 < 1/12 = 0.08(3)$

$\mathcal{C}_0 = 1 \ldots 0.75$: graphics reflect stepwise identification of trend SVD components $\Rightarrow$ considerable changes of MSE$(F^{(T)}, \widetilde{F}_0^{(T)})$

$\mathcal{C}_{\text{opt}} < 0.75 (\approx 0.75)$

Approximation and forecast of one time series

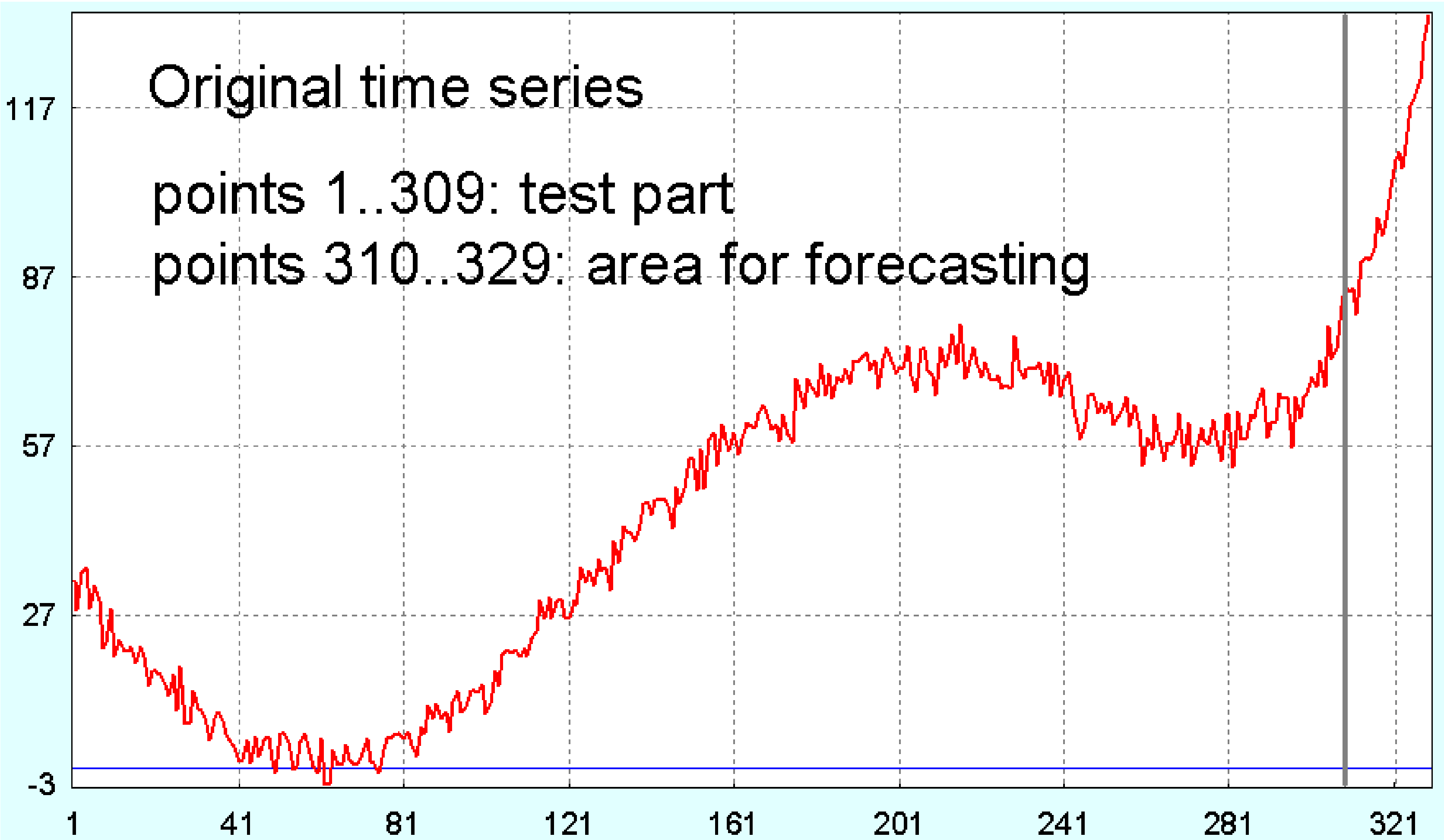
- Simulate a time series $F = f_1, \ldots, f_N, \quad N = 329$
- Take its first part $G = f_1, \ldots, f_M, \quad M = 309$
- Extract a trend $G^{(T)}$ of this first part
- Make 20 points ahead forecast

  Thus we received trend forecast for points $310, \ldots, 329$

- Compare the forecasted trend with the original one on points $310, \ldots, 329$

Approximation and forecast of one time series

$f_n = 10^{-9}(n+100)(n-30)(n-110)(n-230)(n-350) + 7e^{0.01n} + N(0, 3^2)$, noise is i.i.d.



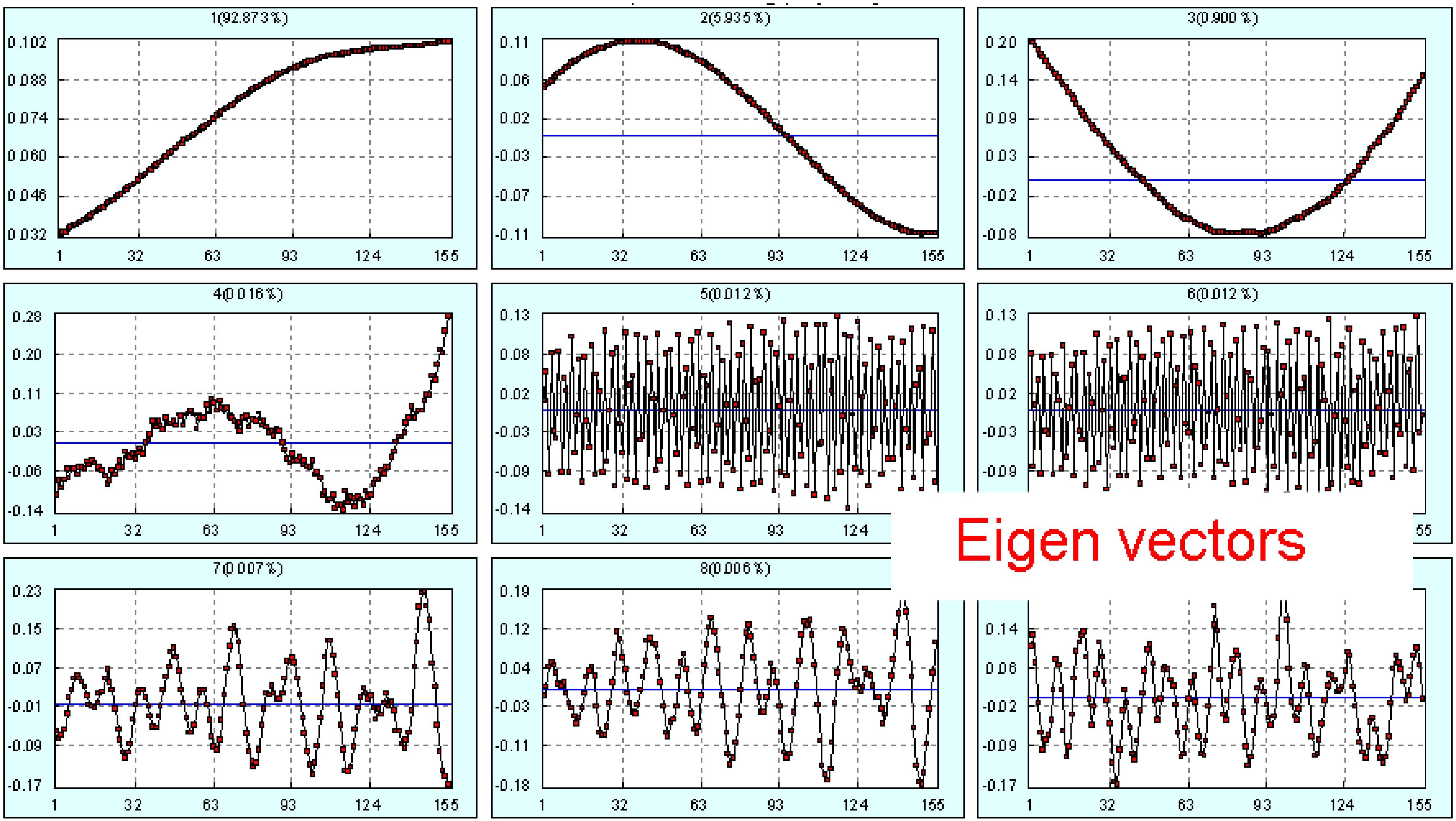
$N = 309$, 20 points ahead forecast, $L = 155$,

$\omega_0 = 0.02$: calculated $\mathcal{C}_0 = 0.73$, identified components $I^{(T)} = \{1, 2, 3, 4\}$

$\omega_0 = 0.04$: calculated $\mathcal{C}_0 = 0.79$, identified components $I^{(T)} = \{1, 2, 3, 4, 34\}$

Approximation and forecast of one time series

$f_n = 10^{-9}(n+100)(n-30)(n-110)(n-230)(n-350) + 7e^{0.01n} + N(0,3^2)$, noise is i.i.d.



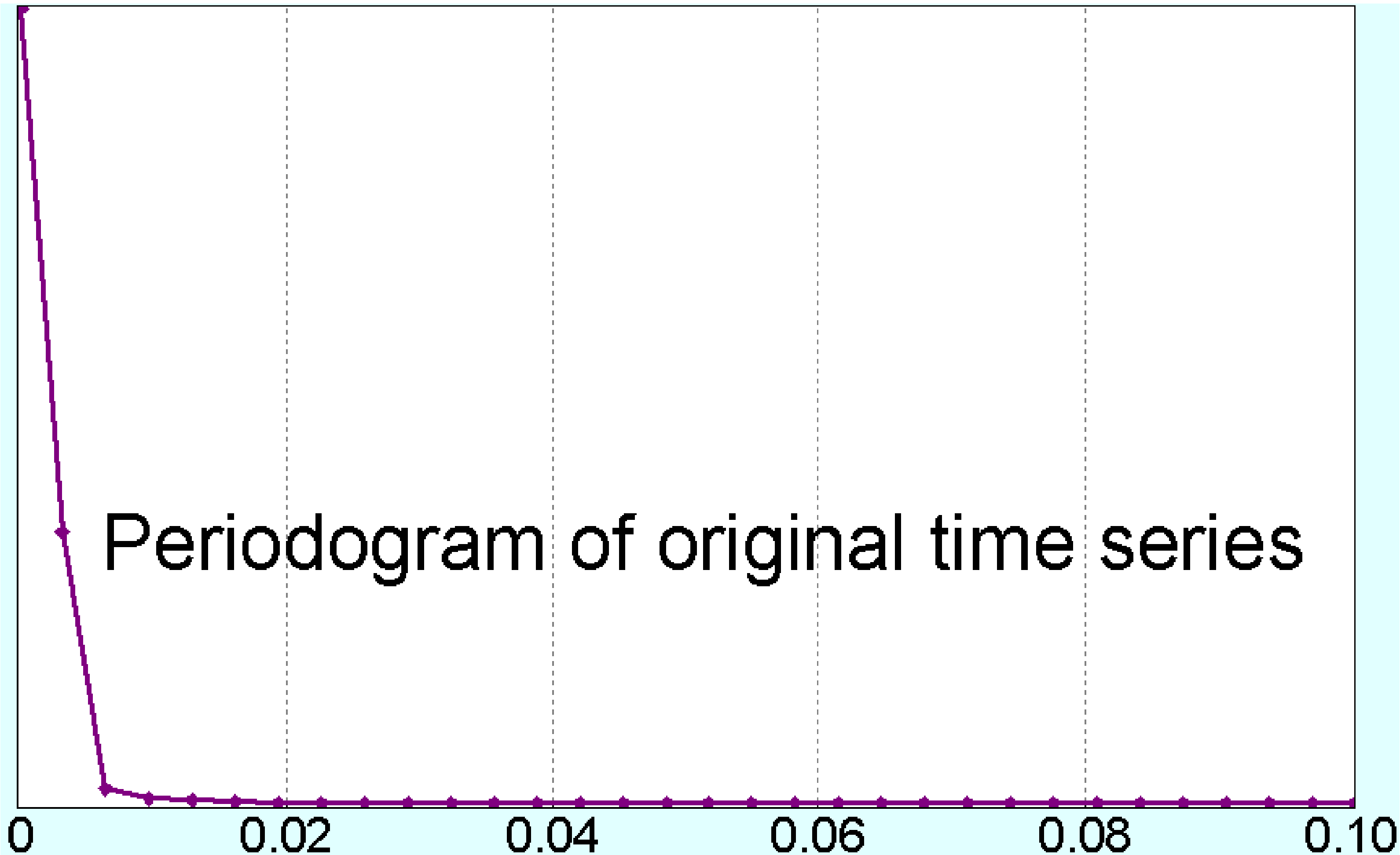
$N = 309$, 20 points ahead forecast, $L = 155$,

$\omega_0 = 0.02$: calculated $\mathcal{C}_0 = 0.73$, identified components $I^{(T)} = \{1,2,3,4\}$

$\omega_0 = 0.04$: calculated $\mathcal{C}_0 = 0.79$, identified components $I^{(T)} = \{1,2,3,4,34\}$

Approximation and forecast of one time series

$f_n = 10^{-9}(n+100)(n-30)(n-110)(n-230)(n-350) + 7e^{0.01n} + N(0, 3^2)$, noise is i.i.d.



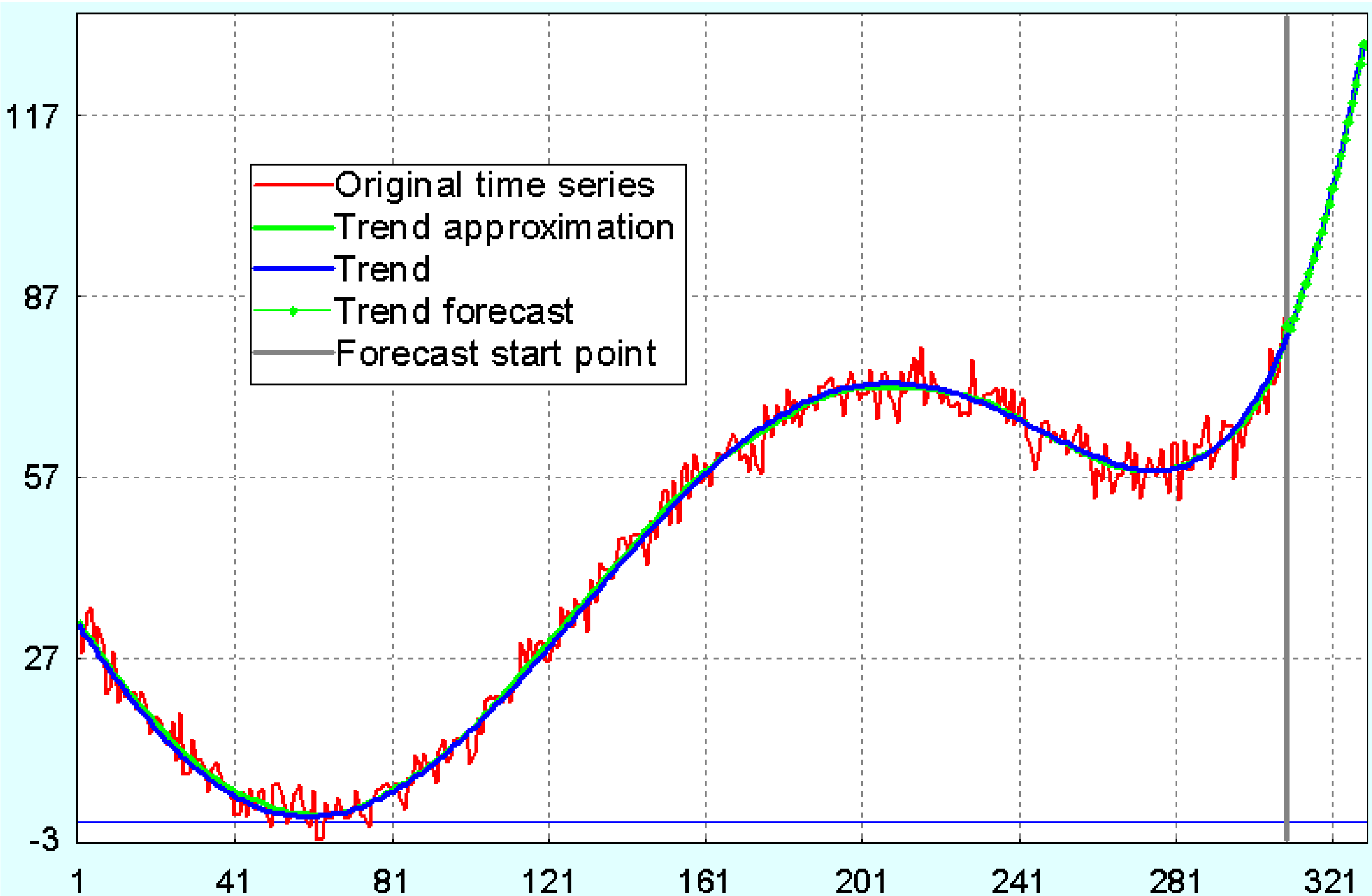
$N = 309$, 20 points ahead forecast, $L = 155$,

$\omega_0 = 0.02$: calculated $\mathcal{C}_0 = 0.73$, identified components $I^{(T)} = \{1, 2, 3, 4\}$

$\omega_0 = 0.04$: calculated $\mathcal{C}_0 = 0.79$, identified components $I^{(T)} = \{1, 2, 3, 4, 34\}$

Approximation and forecast of one time series

$f_n = 10^{-9}(n+100)(n-30)(n-110)(n-230)(n-350) + 7e^{0.01n} + N(0, 3^2)$, noise is i.i.d.



$N = 309$, 20 points ahead forecast, $L = 155$,

$\omega_0 = 0.02$: calculated $\mathcal{C}_0 = 0.73$, identified components $I^{(T)} = \{1, 2, 3, 4\}$

$\omega_0 = 0.04$: calculated $\mathcal{C}_0 = 0.79$, identified components $I^{(T)} = \{1, 2, 3, 4, 34\}$

Final slide

Information about “Caterpillar”-SSA

- Golyandina N., Nekrutkin V., Zhigljavsky A.*Analysis of Time Series Structure: SSA and Related Techniques*, Chapman&Hall/CRC (2001), 305 p.
- ”Caterpillar”-SSA: http://www.gistatgroup.com/cat/

Information about AutoSSA

- Th. Alexandrov, N. Golyandina, *Automatic extraction and forecast of time series cyclic components within the framework of SSA*, Proceedings of the 5th Workshop on Simulation (2005), pp.45-50
- Th. Alexandrov, *Batch extraction of additive components of time series by means of the Caterpillar-SSA method*, Vestnik St. Petersburg Univ. Math. (2006, in print)
- AutoSSA: http://www.pdmi.ras.ru/˜theo/autossa/