Central Mean Subspace in Time Series

Jin-Hong Park

Department of Mathematics
College of Charleston

July 28, 2009

Overview

Goal: Develop a new method for analyzing Time Series using Central Mean Dimension Reduction Subspace

- Time series framework for Dimension Reduction (TSCS)
- Definition of Time Series Central Mean Subspace (TSCMS)
- Estimation Methods
- Simulations, Real Data Analysis
- Discussions

Time Series Central Subspace

- Suppose x_t denotes a time series.
- Let $\mathbf{X}_{t-1} = (x_{t-1}, x_{t-2}, ..., x_{t-p})^T$ for some lag p.
- Then, conditional density of x_t given $\mathbf{X_{t-1}}$ is of interest.
- We want to find $\mathbf{\Phi}_d$, a $p \times d$ matrix such that $p(x_t|\mathbf{X}_{t-1}) = p(x_t|\mathbf{\Phi}_q^T\mathbf{X_{t-1}})$, $q < p$.
- x_t is conditionally independent of $\mathbf{X}_{t-1}$ given $\mathbf{\Phi}_q^T\mathbf{X}_{t-1}$. $(x_t \perp \mathbf{X}_{t-1}|\mathbf{\Phi}_q^T\mathbf{X}_{t-1})$.
- Here, $\mathbf{\Phi}_q = (\Phi_1, \Phi_2, ..., \Phi_q)$. Hence, we search for $\Phi_1^T\mathbf{X_{t-1}}, \Phi_2^T\mathbf{X_{t-1}}, ..., \Phi_q^T\mathbf{X_{t-1}}$ of the past such that $x_t \perp \mathbf{X}_{t-1}|\mathbf{\Phi}_q^T\mathbf{X}_{t-1}$.

Time Series Central Mean Subspace

- A subspace $S(\mathbf{\Phi}_q)$ of $\mathbb{R}^p$ for which the conditional independence, $x_t \perp E(x_t|\mathbf{X}_{t-1})|\mathbf{\Phi}_q^T\mathbf{X}_{t-1}$, holds is called a time series mean dimension reduction subspace for x_t on $\mathbf{X}_{t-1}$.
- $x_t \perp \mathbf{X}_{t-1}|\mathbf{\Phi}_q^T\mathbf{X}_{t-1}$ implies $x_t \perp E(x_t|\mathbf{X}_{t-1})|\mathbf{\Phi}_q^T\mathbf{X}_{t-1}$.
- The equivalent conditions:
  (i) $x_t \perp E(x_t|\mathbf{X}_{t-1})|\mathbf{\Phi}_q^T\mathbf{X}_{t-1}$.
  (ii) $Cov(x_t, E(x_t|\mathbf{X}_{t-1})|\mathbf{\Phi}_q^T\mathbf{X}_{t-1}) = 0$.
  (iii) $E(x_t|\mathbf{X}_{t-1})$ is a function of $\mathbf{\Phi}_q^T\mathbf{X}_{t-1}$.
- Proof: Cook and Li (2002, p.471)
- Three important issues

Estimation Methods

- Objective Function: $\mathbf{\Psi}(\mathbf{h}_d) = E(x_t - f(\mathbf{h}_d^T \mathbf{X}_{t-1}))^2$ where $f(\mathbf{h}_d^T \mathbf{X}_{t-1}) = E(x_t | \mathbf{h}_d^T \mathbf{X}_{t-1})$ and minimize $\mathbf{\Psi}$ with respect to $\mathbf{h}_d$ such that ${\mathbf{h}_d}^T \mathbf{h}_d = I_d$.
- $$\hat{f}_{\lambda_n}(\mathbf{h}_d^T \mathbf{x}) = \frac{\sum_{i=1}^n K\left(\frac{\mathbf{h}_d^T \mathbf{x} - \mathbf{h}_d^T \mathbf{X}_{i-1}}{\lambda_n}\right) x_i}{\sum_{j=1}^n K\left(\frac{\mathbf{h}_d^T \mathbf{x} - {\mathbf{h}_d}^T \mathbf{X}_{j-1}}{\lambda_n}\right)}$$
  where K is a kernel function and $\{\lambda_n\}$ is a sequence of bandwidths. - Nadaraya-Watson estimator

Computational Algorithm

- Sample version: $\hat{\boldsymbol{\Psi}}_n(\mathbf{h}_d) = \sum_{t=1}^{n}(x_t - \hat{f}_{\lambda_n}(\mathbf{h}_d^T \mathbf{X}_{t-1}))^2$ with respect to $\mathbf{h}_d$ such that $\mathbf{h}_d{}^T \mathbf{h}_d = I_d$.
- General guidelines for choice of kernels and selection of bandwidths: Silvermann (1986) and Scott (1992) $K((\mathbf{u} - \mathbf{V})/\lambda_n) = (n \prod_{j=1}^{d} a_{nj})^{-1} \prod_{j=1}^{d} G\left(\frac{u_j - V_j}{a_{nj}}\right)$, where $\mathbf{u} = (u_1, \cdots, u_d)^T$, $\mathbf{V} = (V_1, \cdots, V_d)^T$, $a_{nj} = b_d s_j n^{-1/(4+d)}$ for $j = 1, \cdots, d$, $b_d = 4/(2d+1)^{1/d+4}$.
- $\hat{\boldsymbol{\Phi}}_{n,d} = \arg\min_{\mathbf{h}_d} \hat{\boldsymbol{\Psi}}_n(\mathbf{h}_d)$ such that $\|\mathbf{h}\| = I$: MATLAB code

Estimation of dimension and lag

In time series, there may not be any prior information available on the dimension d and number of lags p.
$\rightarrow$ data-dependent method to determine d and p.
For a fixed $p(\geq 2)$, we determine $\hat{d}_p$ using the following the MSBC (Modified Schwarz Bayesian information Criterion).

- $\hat{d}_p = \arg\min_d \{n \log(\hat{\boldsymbol{\Psi}}_n(\hat{\boldsymbol{\Phi}}_{p,d})/n) + pdn^{0.375}\}$

Also, for a fixed d, we estimate $\hat{p}_d$.

- $\hat{p}_d = \arg\min_p \{n \log(\hat{\boldsymbol{\Psi}}_n(\hat{\boldsymbol{\Phi}}_{p,d})/n) + pdn^{0.375}\}$

Simulation - Two Measures

To evaluate accuracy of TSCMS estimates, use distance between $\mathcal{S}_{E(x_t|\mathbf{X}_{t-1})}(\hat{\mathbf{\Phi}}_d)$ and $\mathcal{S}_{E(x_t|\mathbf{X}_{t-1})}(\mathbf{\Phi}_d)$.

- Ye and Weiss (2003) - vector correlation coefficient
  $\rho = \sqrt{|\hat{\mathbf{\Phi}}_d^T \mathbf{\Phi}_d \mathbf{\Phi}_d^T \hat{\mathbf{\Phi}}_d|}$ $(0 \leq \rho \leq 1)$
  $\rightarrow$ the bigger, the more accurate
- Xia, Tong, Li, and Zhu (2002)
  $m^2 = \left\|(I - \mathbf{\Phi}_d {\mathbf{\Phi}_d}^T)\hat{\mathbf{\Phi}}_q\right\|^2$, if $q < d$
  $m^2 = \left\|(I - \hat{\mathbf{\Phi}}_q \hat{\mathbf{\Phi}}_q^T)\mathbf{\Phi}_d\right\|^2$, if $q \geq d$
  $\rightarrow$ the smaller, the more accurate

Model 1

- Model 1: $x_t = 0.5\{\cos(1.0)x_{t-1} - \sin(1.0)x_{t-2}\} + 0.4\exp[-16\{\cos(1.0)x_{t-1} - \sin(1.0)x_{t-2}\}^2] + 0.2\varepsilon_t$

n	ρ	m^2	MSBC ($p=2$)	MSBC ($d=1$)
100	0.9313	0.0817	$f_{(d=1)} = 100$ $f_{(d=2+)} = 0$	$f_{(p=1)} = 0$ $f_{(p=2)} = 94$ $f_{(p=3)} = 6$
200	0.9443	0.0622	$f_{(d=1)} = 100$ $f_{(d=2+)} = 0$	$f_{(p=1)} = 0$ $f_{(p=2)} = 93$ $f_{(p=3)} = 7$
300	0.9492	0.0600	$f_{(d=1)} = 100$ $f_{(d=2+)} = 0$	$f_{(p=1)} = 0$ $f_{(p=2)} = 93$ $f_{(p=3)} = 7$

Table: Model 1: All results are based on 100 Monte Carlo replications. Average values of ρ and m^2 are reported for true $p=2$ and $d=1$. Frequency of estimated dimension (with true lag $p=2$) and estimated lag (with true $d=1$) are reported using MSBC.

Model 2

- Model 2: $x_t = (\pi/2)(1/\sqrt{5})(x_{t-2} + 2x_{t-3})\exp(-x_{t-1}^2) + 0.2\varepsilon_t$

n	ρ	m^2	MSBC ($p = 3$)	MSBC ($d = 2$)
100	0.9484	0.0302	$f_{(d=1)} = 41$	$f_{(p=2)} = 0$
		0.0449	$f_{(d=2)} = 59$	$f_{(p=3)} = 87$
			$f_{(d=3+)} = 0$	$f_{(p=4)} = 13$
200	0.9408	0.0342	$f_{(d=1)} = 0$	$f_{(p=2)} = 0$
		0.0420	$f_{(d=2)} = 100$	$f_{(p=3)} = 92$
			$f_{(d=3+)} = 0$	$f_{(p=4)} = 8$
300	0.9688	0.0204	$f_{(d=1)} = 0$	$f_{(p=2)} = 0$
		0.0258	$f_{(d=2)} = 100$	$f_{(p=3)} = 90$
			$f_{(d=3+)} = 0$	$f_{(p=4)} = 10$

Table: Model 2: All results are based on 100 Monte Carlo replications. Average values of ρ and m^2 are reported for true $p = 3$ and $d = 2$. Frequency of estimated dimension (with true lag $p = 3$) and estimated lag (with true $d = 2$) are reported using MSBC.

Model 3

- Model 3: $x_t = -1 + (0.4)(1/\sqrt{5})(x_{t-1} + 2x_{t-4}) - \cos((\pi/2)(1/\sqrt{5})(x_{t-3} + 2x_{t-6})) + \exp(-(1/\sqrt{15})^2(-2x_{t-1} + 2x_{t-2} - 2x_{t-3} + x_{t-4} - x_{t-5} + x_{t-6})^2) + 0.2\varepsilon_t$

n	ρ	m^2	MSBC ($p = 6$)	MSBC ($d = 3$)
100	0.7859	0.1088	$f_{(d=1)} = 4$	$f_{(p=4)} = 0$
		0.1476	$f_{(d=2)} = 83$	$f_{(p=5)} = 0$
		0.1260	$f_{(d=3)} = 13$	$f_{(p=6)} = 100$
			$f_{(d=4+)} = 0$	$f_{(p=7)} = 0$
200	0.9465	0.0363	$f_{(d=1)} = 0$	$f_{(p=4)} = 0$
		0.0356	$f_{(d=2)} = 37$	$f_{(p=5)} = 0$
		0.0316	$f_{(d=3)} = 63$	$f_{(p=6)} = 100$
			$f_{(d=4+)} = 0$	$f_{(p=7)} = 0$
300	0.9684	0.0204	$f_{(d=1)} = 0$	$f_{(p=4)} = 0$
		0.0201	$f_{(d=2)} = 9$	$f_{(p=5)} = 0$
		0.0219	$f_{(d=3)} = 91$	$f_{(p=6)} = 100$
			$f_{(d=4+)} = 0$	$f_{(p=7)} = 0$

Canadian Lynx Data: 1821 - 1934

- A benchmark series to test new statistical methodologies: The periodic fluctuation of this series has profoundly influenced ecologists to examine concepts such as “balance of nature”, predator (lynx) and prey (hare) interaction, food web dynamics etc. (see Stenseth et al. 1999)

Canadian Lynx Data

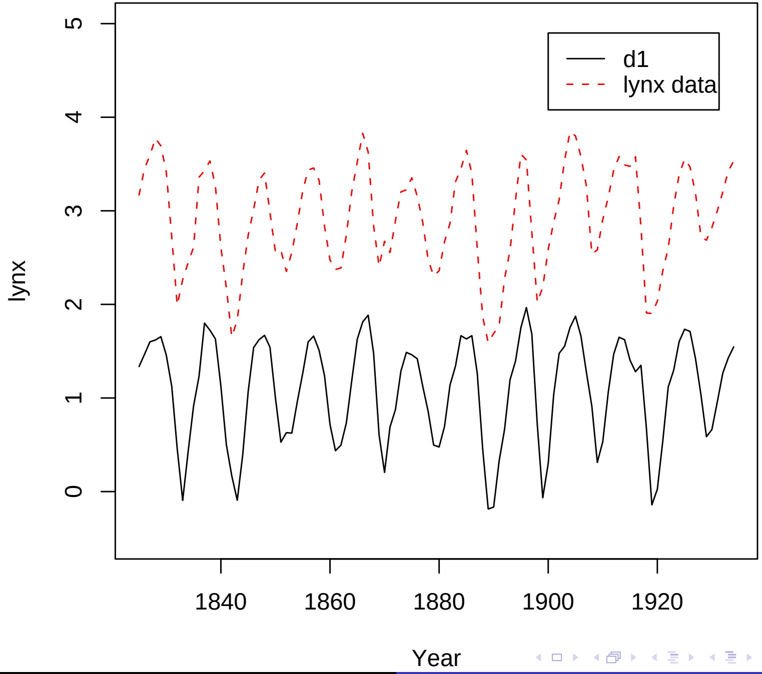
- Use MSBC and detect $\hat{d} = 1$ and $\hat{p} = 4$.

p	d=1	d=2	d=3	d=4
2	-2.0577	-2.0486		
3	-2.1497	-2.1125	-2.0781	
4	-2.1719*	-2.0919	-2.0350	-1.8795
5	-2.1189	-1.9972	-1.8950	-1.6924
6	-2.0692	-1.8987	-1.7434	-1.7034

Table: *Canadian lynx data*: The table gives modified SBC values for the pairs (p, d). Here, * denotes ($\hat{d}$, $\hat{p}$) = (1,4) using MSBC, respectively.

Canadian Lynx Data

- Estimate 4×1 basis vector $\mathit{\Phi_1}$.
  $\hat{\Phi}_1^T = (0.9317, -0.0761, -0.1777, -0.3074)$.
- x_t vs $d_{1,t} (= \Phi_1 \mathbf{X}_{t-1})$
- Using trial and error approach and graphic tools, detect both linear and nonlinear pattern with cycles.
  $x_t = \beta_0 + \beta_1 d_{1,t} + \beta_2 cos_t + \eta_t$ where
  $cos_t = \cos(3.87 d_{1,t} - 3.44)$.
- After autocorrelation check,
  $x_t = 2.08 + 0.75 d_{1,t} - 0.13 cos_t + \eta_t$ where $\eta_t = 0.52 \eta_{t-1} + \varepsilon_t$.
- Final fitted Model: $\hat{x}_t =$
  $0.99 + 0.52 x_{t-1} + 0.75 d_{1,t} - 0.39 d_{1,t-1} - 0.13 \cos_t + 0.07 cos_{1,t-1}$
  where $d_{1,t-1}$ and $cos_{1,t-1}$ denote the lag-1 of the corresponding series.



d1
lynx data
lynx
Year
0
1
2
3
4
5
1840
1860
1880
1900
1920

Canadian Lynx Data

- Compare performance of our model with Tong's models, SETAR (2;2,2) and SETAR(2;7,2), and Tsay's SETAR(3;1,7,2) model.
  Two Mean Prediction Errors:
  1. Mean Absolute Percentage Error (MAPE) = $n^{-1}\sum_{t=1}^{n}\{|x_t - \hat{x}_t|/x_t\}$
  2. Mean Square Relative Error (MSRE) = $n^{-1}\sum_{t=1}^{n}\{(x_t - \hat{x}_t)^2/x_t\}$

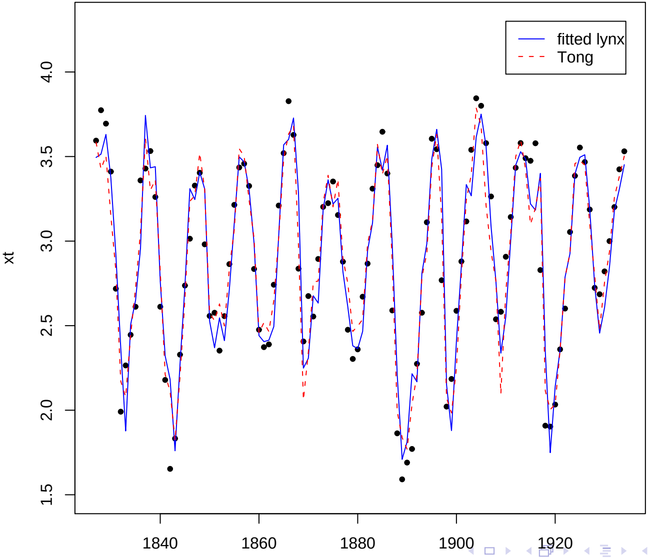
  Table and Figure: Our model is comparative and simple.
- Finally, our TSCMS approach provides a viable alternative to the existing techniques such as TAR for modeling nonlinear time series.

Canadian Lynx Data

Model	MAPE	MSRE	n
Our model	0.0618	0.0177	108
Tong's model 1	0.0593	0.0160	108
Tong's model 2	0.0561	0.0136	108
Tsay's model	0.0557	0.0133	108

Table: *Canadian lynx data*: Mean Average Percentage Error (MAPE) and Mean Square Relative Error (MSRE) values for our model, Tong's models and Tsay's model.

Canadian Lynx Data



Canadian Lynx Data

- Tong's best model - SETAR(2;7,2) (1990, p. 386):

$$
\begin{aligned}
x_t &= \{0.546 + 1.032x_{t-1} - 0.173x_{t-2} + 0.171x_{t-3} - 0.431x_{t-4} \\
&+ 0.332x_{t-5} - 0.284x_{t-6} + 0.210x_{t-7} + \varepsilon_t^{(1)}\} I(x_{t-2} \leq 3.116) \\
&+ \{2.632 + 1.492x_{t-1} - 1.324x_{t-2} + \varepsilon_t^{(2)}\} I(x_{t-2} > 3.116).
\end{aligned}
$$

Discussion

- Develop TSCMS: DR in TS focused only on the mean function of series
- Estimate minimum dimension and lag using a unified criterion (MSBC)
- The encouraging results presented here seem to suggest that our method has great potential for providing a viable alternative to traditional time series analysis.