Fourier Method on Sufficient Dimension Reduction in Time Series

Tharindu P. De Alwis
(Joint work with Dr. S. Y. Samadi)

School of Mathematical and Statistical Sciences,
Southern Illinois University Carbondale

mktharinu87@siu.edu

CBMS Conference
July 30, 2021

Overview

1. Background of the Problem
2. SDR in Time Series Models.
3. Simulated Example.
4. Empirical Application.
5. Key References.

Dimension Reduction (DR)

Dimension reduction is accomplished through one or both of the following approaches:

(1) Sparsity: Reduce the dimensionality by selecting a subset of the original predictors.

- Variable selection, often via penalization.

(2) Reducibility: Reduce dimensionality by (linear or nonlinear) projection of the p-dimensional vector $\mathbf{X}$ onto d-dimensional vector ($d < p$).

- (Sufficient) Dimension reduction.

Sufficient Dimension Reduction (SDR)

Basic regression setup:

- Consider response $Y \in \mathbb{R}$ and predictor $\mathbf{X} \in \mathbb{R}^p$.
- PCA is not a reliable method for predictors reduction in regression as it ignores the response variable.
- When the mean response $E[Y|\mathbf{X}]$ is of primary interest, then SDR seeks $\boldsymbol{\eta}^T \in \mathbb{R}^{p \times d}$ such that $E[Y|\mathbf{X}]$ depend on $\mathbf{X}$ only through $\boldsymbol{\eta}^T\mathbf{X}$, (Cook and Li, 2002).

$$E[Y|\mathbf{X}] = E[Y|\boldsymbol{\eta}^T\mathbf{X}]$$

SDR for Conditional moments

Central Mean Subspace (CMS)

- These current techniques, attempt to construct a small number of directions from the original predictors, i.e., $\mathbf{X} \mapsto \boldsymbol{\eta}^T\mathbf{X}$, $\boldsymbol{\eta} \in \mathbb{R}^{p \times d}$, $d < p$ such that

$$Y \perp\!\!\!\perp E[Y|\mathbf{X}]|\boldsymbol{\eta}^T\mathbf{X}$$

  Or equivalently

$$E[Y|\mathbf{X}] \stackrel{\mathcal{D}}{=} E[Y|\boldsymbol{\eta}^T\mathbf{X}], \quad \boldsymbol{\eta} \in \mathbb{R}^{p \times d}$$

- $\cap Span(\boldsymbol{\eta}) = \mathcal{S}_{E[Y|\mathbf{X}]}$- known as central mean subspace (CMS).
- If higher conditional moments are of interest, parallel definitions can be made (Yin and Cook, 2002).
- More specifically the second moment, i.e., central variance subspace (CVS), denote by $\mathcal{S}_{Var(Y|\mathbf{X})}$.

SDR and CMS- Example

For many models, $\mathcal{S}_{E[Y|\mathbf{X}]} \equiv \mathcal{S}_{Y|\mathbf{X}}$:
Let $\mathbf{X} = (X_1, \cdots, X_6)^T$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Assume $\varepsilon \perp\!\!\!\perp \mathbf{X}$.

- Consider the model

$$y = g(X_1 + 0.5X_2) + \varepsilon.$$

  We have $\mathcal{S}_{E[Y|\mathbf{X}]} = Span(\boldsymbol{\beta}_1)$, where $\boldsymbol{\beta}_1 = (1, 0.5, 0, 0, 0, 0)^T$.

- Now consider the model

$$y = g(X_1 + 0.5X_2) + X_3\varepsilon,$$

  whereas, here $\mathcal{S}_{E[Y|\mathbf{X}]} = Span(\boldsymbol{\beta}_1)$, and $\mathcal{S}_{Var(Y|\mathbf{X})} = Span(\boldsymbol{\beta}_2)$ where $\boldsymbol{\beta}_2 = (0, 0, 1, 0, 0, 0)^T$.

The Model

- Consider a general setting of univariate time series model for $\{y_t\}_{t=1}^{N}$ as
$$y_t = g_1(\boldsymbol{\eta}^T \mathbf{Y}_{t-1}) + x_t, \tag{1}$$
where $g_1(\cdot)$ is an unknown smoothing link function and $\mathbf{Y}_{t-1} = (y_{t-1}, \cdots, y_{t-p})^T$.
- x_t's are white noise error terms or x_t has a general heteroscedastic structure as
$$x_t = \sqrt{g_2(\boldsymbol{\Gamma}^T \mathbf{X}_{t-1})}\ \ \varepsilon_t, \tag{2}$$
where ε_t's are white noise terms, $g_2(\cdot)$ is an unknown smoothing link function, and $\mathbf{X}_{t-1} = (x_{t-1}^2, \cdots, x_{t-q}^2)^T$.

SDR in Time Series Models

CMS and CVS in Time Series

- As in the regression setup, we find $\boldsymbol{\eta} \in \mathbf{R}^{p\times d}$ s.t

  for CMS:

$$E[y_t|\mathbf{Y}_{t-1}] \stackrel{\mathcal{D}}{=} E[y_t|\boldsymbol{\eta}^T\mathbf{Y}_{t-1}], \qquad \text{Goal: Estimate } \mathcal{S}_{E[y_t|\mathbf{Y}_{t-1}]}$$

  and for CVS:

$$Var(y_t|\mathbf{Y}_{t-1}) \stackrel{\mathcal{D}}{=} Var(y_t|\boldsymbol{\eta}^T\mathbf{Y}_{t-1}), \qquad \text{Goal: Estimate } \mathcal{S}_{Var(y_t|\mathbf{Y}_{t-1})}$$

Fourier Method- Time Series

- Let $m(\mathbf{y}_{t-1}) = E[y_t|\mathbf{Y}_{t-1} = \mathbf{y}_{t-1}]$, and let $\mathbf{u}_{t-1} = \boldsymbol{\eta}^T\mathbf{Y}_{t-1}$.

$$\frac{\partial}{\partial \mathbf{y}_{t-1}} m(\mathbf{y}_{t-1}) = \boldsymbol{\eta} \frac{\partial}{\partial \mathbf{u}_{t-1}} g_1(\mathbf{u}_{t-1}) \in \mathcal{S}_{E[y_t|\mathbf{Y}_{t-1}]}. \tag{3}$$

- For any $\boldsymbol{\omega} \in \mathbb{R}^p$,

$$\begin{aligned} \psi_t(\boldsymbol{\omega}) &= \int \exp\{i\boldsymbol{\omega}^T\mathbf{y}_{t-1}\}\left(\frac{\partial}{\partial \mathbf{y}_{t-1}} m(\mathbf{y}_{t-1})\right) f(\mathbf{y}_{t-1})d\mathbf{y}_{t-1}, \\ &= -E_{(\mathbf{Y}_{t-1},y_t)}\left[y_t\left(G(\mathbf{Y}_{t-1}) + i\boldsymbol{\omega}\right)\exp\{i\boldsymbol{\omega}^T\mathbf{Y}_{t-1}\}\right], \end{aligned} \tag{4}$$

where $G(\mathbf{Y}_{t-1}) = \frac{\partial}{\partial \mathbf{y}_{t-1}} \log f(\mathbf{Y}_{t-1})$,

- Nice property of $\psi(\boldsymbol{\omega})$:
  Has all information about $\frac{\partial}{\partial \mathbf{y}_{t-1}} m(\mathbf{y}_{t-1})$ and it can be recover through the inverse Fourier transform.

Fourier Method- Time Series-CMS

- Let $s > t$. Often $\mathbf{Y}_{s-1}$ and $\mathbf{Y}_{t-1}$ share some observations where $\mathbf{y}_{s-1} = (y_{s-1}, \cdots, y_{s-p})^T$, and $k = |s - t|$, then define,

$$m(\mathbf{y}_{s-1}, \mathbf{y}_{t-1}) = E[y_s | (\mathbf{Y}_{s-1} = \mathbf{y}_{s-1}, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1})]$$

- Let $t = 7$, $s = 9$, and $p = 3$, then $k = 2$, $\mathbf{Y}_6 = (y_6, y_5, y_4)$, and $\mathbf{Y}_8 = (y_8, y_7, y_6)$, where $\mathbf{Y}_6$ and $\mathbf{Y}_8$ share the observation y_6.
- In general, it can be shown that if $k < p$, then $\mathbf{Y}_{t-1}$ and $\mathbf{Y}_{s-1}$ would have $p - k$ observations in common.
- $\mathbf{Y}_{s-1}$ and $\mathbf{Y}_{t-1}$ are p-dependent" random variables, for more details see Lehmann (1998).

Estimation process of FMTS-CMS Count..

Therefore, we can consider following two cases on $m(\mathbf{y}_{s-1}, \mathbf{y}_{t-1})$ as

$$m(\mathbf{y}_{s-1}, \mathbf{y}_{t-1}) = \begin{cases} m(\mathbf{y}_{s-1}, \mathbf{y}_{t-1}) & k < p \\ m(\mathbf{y}_{s-1}) & k \geq p \end{cases} \tag{5}$$

Under condition $k < p$:

$$\begin{aligned} \psi_s(\boldsymbol{\omega}) &= \int \exp\{i\boldsymbol{\omega}^T \mathbf{y}_{s-1}\} \left(\frac{\partial}{\partial \mathbf{y}_{s-1}} m(\mathbf{y}_{s-1}, \mathbf{y}_{t-1}) \right) f(\mathbf{y}_{s-1} | \mathbf{y}_{t-1}) d\mathbf{y}_{s-1}, \\ &= -E_{(\mathbf{Y}_{s-1}, y_s)} \left[y_s \left(G(\mathbf{Y}_{s-1} | \mathbf{Y}_{t-1}) + i\boldsymbol{\omega} \right) \exp\{i\boldsymbol{\omega}^T \mathbf{Y}_{s-1}\} \right], \end{aligned} \tag{6}$$

where $G(\mathbf{Y}_{s-1} | \mathbf{Y}_{t-1}) = \frac{\partial}{\partial \mathbf{y}_{s-1}} \log f(\mathbf{Y}_{s-1} | \mathbf{Y}_{t-1})$,

Estimation process of FMTS-CMS Count..

- Since $\psi_t(\omega), \psi_s(\omega) \in \mathcal{S}_{E[y_t|\mathbf{Y}_{t-1}]}$, then

$$\mathcal{S}\left(\psi_t(\omega)\bar{\psi}_s(\omega)^T\right) \subseteq \mathcal{S}_{E[y_t|\mathbf{Y}_{t-1}]}. \tag{7}$$

- Define

$$\mathbf{M}_{FMTS} = Re\left(\int \psi_t(\omega)\bar{\psi}_s(\omega)^T W(\omega)d\omega\right)$$
$$= \int \left[\mathbf{a}_s(\omega)\mathbf{a}_t(\omega)^T + \mathbf{b}_s(\omega)\mathbf{b}_t(\omega)^T\right] W(\omega)d\omega$$

- $\mathbf{M}_{FMTS}$ is a real non negative definite matrix, and

$$\mathcal{S}(\mathbf{M}_{FMTS}) = \mathcal{S}_{E[y_t|\mathbf{Y}_{t-1}]}.$$

Estimation process of FMTS-CMS Count..

- Let $W(\omega) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{\frac{-||\boldsymbol{\omega}||^2}{2\sigma_w^2}\right\}$, then

$\mathbf{M}_{FMTS}$

$$= \begin{cases} E\left[y_t y_s \exp\left\{-\frac{\sigma_w^2 ||\mathbf{y}_{ts}||^2}{2}\right\}\left[\sigma_w^2 \mathbf{I}_p + \left(\mathbf{G}(\mathbf{y}_{t-1}) - \sigma_w^2 \mathbf{y}_{ts}\right)\left(\mathbf{G}(\mathbf{y}_{s-1}) + \sigma_w^2 \mathbf{y}_{ts}\right)^T\right]\right] & k \geq p \\ E\left[y_t y_s \exp\left\{-\frac{\sigma_w^2 ||\mathbf{y}_{ts}||^2}{2}\right\}\left[\sigma_w^2 \mathbf{I}_p + \left(\mathbf{G}(\mathbf{y}_{t-1}) - \sigma_w^2 \mathbf{y}_{ts}\right)\left(\mathbf{G}(\mathbf{y}_{s-1}|\mathbf{y}_{t-1}) + \sigma_w^2 \mathbf{y}_{ts}\right)^T\right]\right] & k < p \end{cases}$$

where $k = |s - t|$, $G(\mathbf{z}) = \frac{\partial}{\partial \mathbf{z}} \log f(\mathbf{z})$, and $\mathbf{y}_{ts} = \mathbf{y}_{s-1} - \mathbf{y}_{t-1}$.

Fourier Method- Time Series- CMS

- The empirical version of $\mathbf{M}_{FMTS}$ is

$$\widehat{\mathbf{M}}_{FMTS} = \frac{1}{N^2} \sum_{t=p+1}^{N} \sum_{s=p+1}^{N} \mathbf{J}_{FMTS}((\mathbf{Y}_{t-1}, y_t), (\mathbf{Y}_{s-1}, y_s)),$$

with
$\mathbf{J}_{FMTS}((\mathbf{y}_{t-1}, y_t), (\mathbf{y}_{s-1}, y_s))$

$$= \begin{cases} y_t y_s \exp\left\{-\frac{\sigma_w^2 ||\mathbf{y}_{ts}||^2}{2}\right\} \left[\sigma_w^2 \mathbf{I}_p + \left(\mathbf{G}(\mathbf{y}_{t-1}) - \sigma_w^2 \mathbf{y}_{ts}\right) \left(\mathbf{G}(\mathbf{y}_{s-1}) + \sigma_w^2 \mathbf{y}_{ts}\right)^T\right] & k \geq p \\ y_t y_s \exp\left\{-\frac{\sigma_w^2 ||\mathbf{y}_{ts}||^2}{2}\right\} \left[\sigma_w^2 \mathbf{I}_p + \left(\mathbf{G}(\mathbf{y}_{t-1}) - \sigma_w^2 \mathbf{y}_{ts}\right) \left(\mathbf{G}(\mathbf{y}_{s-1}|\mathbf{y}_{t-1}) + \sigma_w^2 \mathbf{y}_{ts}\right)^T\right] & k < p \end{cases}$$

where $k = |s - t|$, $G(\mathbf{z}) = \frac{\partial}{\partial \mathbf{z}} \log f(\mathbf{z})$, and $\mathbf{y}_{ts} = \mathbf{y}_{s-1} - \mathbf{y}_{t-1}$.

Discrepancy Measure

- If $\mathbf{A}$ and $\mathbf{B}$ are two matrices, $\mathbf{P}_A = \mathbf{A}(\mathbf{A}^T\mathbf{A})^-\mathbf{A}^T$ and $\mathbf{P}_B = \mathbf{B}(\mathbf{B}^T\mathbf{B})^-\mathbf{B}^T$ are projection matrices, then

$$r = \sqrt{\frac{1}{d} tr\left(\mathbf{P}_A\mathbf{P}_B\right)}$$

  is call the trace correlation, where $d = rank(\mathbf{A}) = rank(\mathbf{B})$.

- Define

$$D(\widehat{\mathcal{S}}, \mathcal{S}) = 1 - r \tag{8}$$

- $0 \geq D(\widehat{\mathcal{S}}, \mathcal{S}) \geq 1$,
- $D(\widehat{\mathcal{S}}, \mathcal{S}) = 0$ if $\widehat{\mathcal{S}} = \mathcal{S}$ $\qquad$ $D(\widehat{\mathcal{S}}, \mathcal{S}) = 1$ if $\widehat{\mathcal{S}} \perp \mathcal{S}$

Estimating p and d

1. B bootstrap time series samples of size N are generated by a fixed block resampling procedures as $\{y_t^{(b)}\}_{t=1}^N, b = 1, \cdots, B$.
2. For each of the bootstrap time series samples $\{y_t^{(b)}\}_{t=1}^N$ and given p and d, estimate $\mathcal{S}_{E[y_t|\mathbf{Y}_{t-1}]}$, and denote it by $\widehat{\mathcal{S}}^{(b)}(p, d)$.
3. Find the distance between $\widehat{\mathcal{S}}^{(b)}(p, d)$ and $\widehat{\mathcal{S}}(p, d)$ as given in Equation (8), and call it $D^{(b)}(p, d)$.
4. Finally, calculate the mean distance over all B samples as

$$\bar{D}(p, d) = \frac{1}{B} \sum_{j=1}^{B} D^{(b)}(p, d). \tag{9}$$

- We repeat the above 4 steps for all candidates $p = \{p_1, \cdots, p_l\}$ and $d = \{1, \cdots, p_i\}$ for $i = 1, \cdots, l$.
- Therefore, for each $p_i, i = 1, \cdots, l$, we have a sequence $\{\bar{d}(p_i, d)\}_{d=1}^{p_i}$.

Estimating p and d

- For a given p_i, the estimator $\hat{d}_i$ is chosen to minimize the variability $\{\bar{D}(p_i, d)\}_{d=1}^{p_i}$ and satisfy the condition $\hat{d}_i < p_i - 1$ in order to achieve the dimension reduction.
- For each lag candidate p_i, we obtain the estimator $\hat{d}_i$. The estimation of $(\hat{p}, \hat{d})$ is the pair which gives the minimum variability among all the possible pairs of $(p_i, \hat{d}_i)$, $\hat{d}_i < p_i - 1$.
- For example, suppose candidate set $p = \{2, 3, 4, 5, 6\}$.

p	d=1	d=2	d=3	d=4	d=5
2	$\bar{D}(2,1)$	0			
3	$\bar{D}(3,1)$	$\bar{D}(3,2)$	0		
4	$\bar{D}(4,1)$	$\bar{D}(4,2)$	$\bar{D}(4,3)$	0	
5	$\bar{D}(5,1)$	$\bar{D}(5,2)$	$\bar{D}(5,3)$	$\bar{D}(5,4)$	0

- We choose the estimated pair of $(\hat{p}, \hat{d})$ which is the argument minimum value for $\bar{D}(p, q)$.

Estimating tuning parameter σ_w^2

1. B bootstrap time series samples of size N are generated by a fixed block resampling procedure as $\{y_t^{(b)}\}_{t=1}^N, b = 1, \cdots, B$.

2. For each of the bootstrap time series $\{\mathbf{Y}_{t-1}^{(b)}, y_t^{(b)}\}$, and for given p and d, estimate $\mathcal{S}_{E[y_t|\mathbf{Y}_{t-1}]}$, and denote it by $\widehat{\mathcal{S}}^{(b)}(\sigma_{w,i}^2)$, where $i = 1, \cdots, I$.

3. Now, calculate the distance between $\hat{\mathcal{S}}^{(b)}(\sigma_{w,i}^2)$ and $\hat{\mathcal{S}}(\sigma_{w,i}^2)$ as given in Equation (8), and call it $D^{(b)}(\sigma_{w,i}^2)$.

4. Calculate the mean distance over all B bootstrap samples as

$$\bar{D}(\sigma_{w,i}^2) = \frac{1}{B} \sum_{b=1}^{B} D^{(b)}(\sigma_{w,i}^2). \tag{10}$$

- The estimator $\widehat{\sigma}_w^2$ is chosen to argument minimize value of $\bar{D}(\sigma_{w,i}^2)$.

Nadaraya-Watson (NW) Estimator

- This is the only method available in literature to estimate TS-CMS by Park et al. (2009).
- The main goal is to estimate the directions in the mean function. That is, estimate the $\left(\frac{\partial}{\partial \mathbf{y}_{t-1}} m(\mathbf{y}_{t-1})\right)$.
- This is based on optimizing a objective function over $\boldsymbol{\eta}$, thus required relatively more computer power compare to proposed Fourier transformation method.
- In the following simulation examples, we compare the accuracy and the execution time per iteration for NW method and FM method.

Simulation Study

- Model 1:

$$y_t = 0.5\{cos(1.0)y_{t-1} - sin(1.0)y_{t-2}\} + 0.4\exp\left\{-16\left[cos(1.0)y_{t-1} - sin(1.0)y_{t-2}\right]^2\right\} + 0.1\varepsilon_t,$$

- Model 2:

$$y_t = (\pi/2)(1/\sqrt(5))(y_{t-2} + 2y_{t-3})\exp(-y_{t-1}^2) + 0.2\varepsilon_t,$$

- Model 3:

$$\sqrt{y_t} = 0.5|\varepsilon_t|\sqrt{1 + \frac{1}{\sqrt{2}}(y_{t-1}^2 + y_{t-4}^2) + \cos\left\{0.1 + \frac{1}{\sqrt{2}}(y_{t-1}^2 + y_{t-4}^2)\right\}},$$

- Model 4:

$$y_t = 3 - \frac{1}{\sqrt{3}}(y_{t-2} + y_{t-4} + y_{t-6}) + x_t,\ \ x_t = \varepsilon_t\sqrt{\frac{1}{\sqrt{6}}(2 + x_{t-1}^2 + x_{t-4}^2)},$$

Simulation Study- Results Model 1

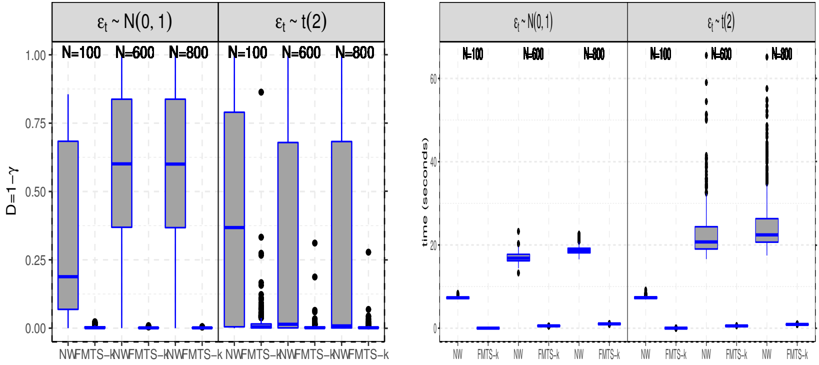


Figure: Mode 1 : $D = 1 - r$ and execution time per iteration.

Simulation Study- Results Model 2

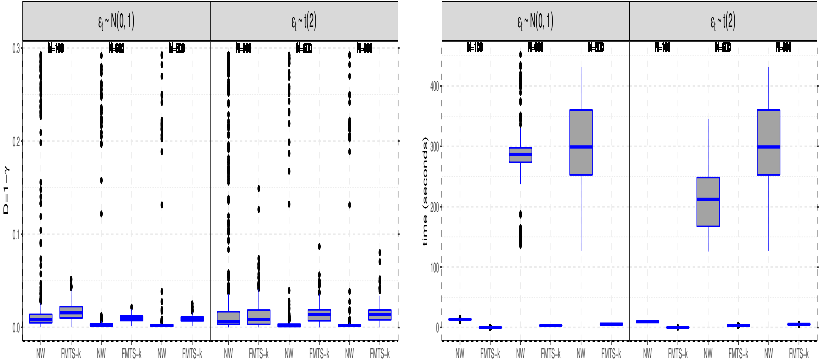


Figure: Mode 2 : $D = 1 - r$ and execution time per iteration.

Simulation Study- Results Model 3

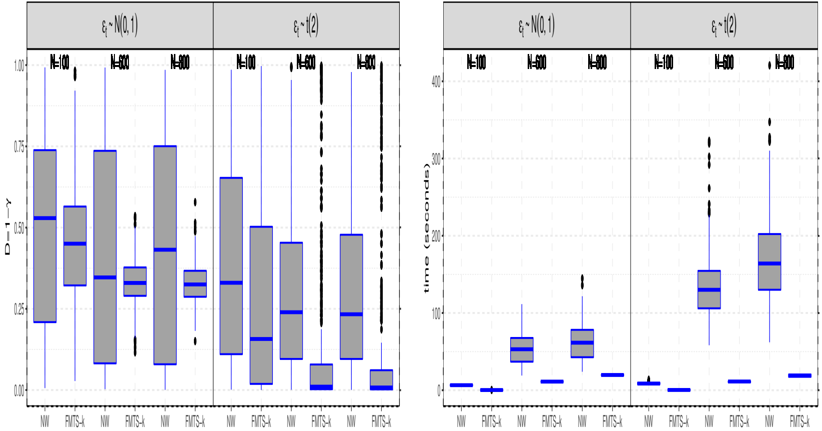


Figure: Mode 3 : $D = 1 - r$ and execution time per iteration.

Simulation Study- Results Model 4

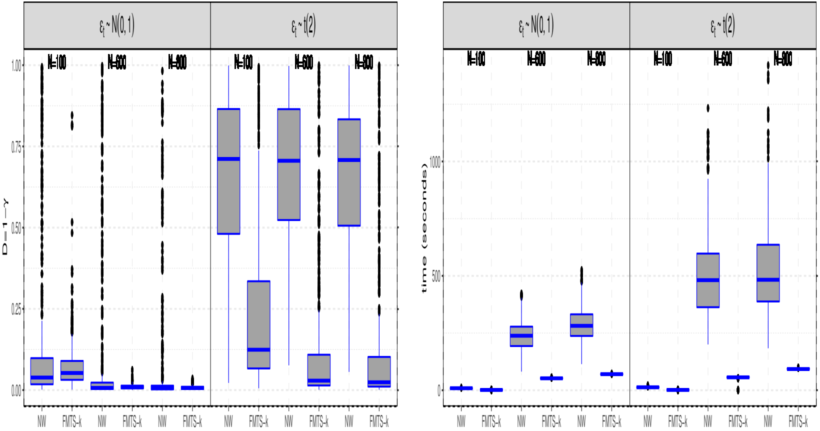


Figure: Mode 4-CMS : $D = 1 - r$ and execution time per iteration.

Simulation Study- Results Model 4

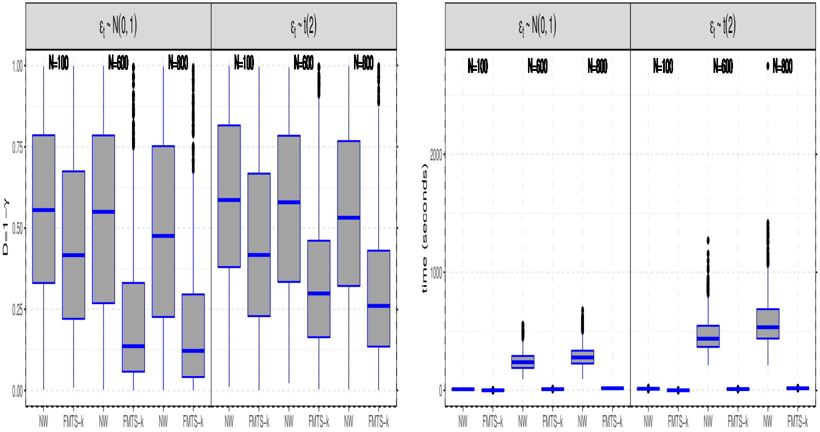


Figure: Mode 4-CVS : $D = 1 - r$ and execution time per iteration, with $N = 600$.

Canadian Lynx data

- The *Canadian Lynx Data* is the annual number of the Canadian lynx "trapped" in the Mackenzie River district of the North-West Canada for the period 1821–1934.
- The performances of the models are measured by mean absolute relative error (MARE) and mean square relative error (MSRE) given as

$$
\begin{aligned}
MARE &= \frac{1}{N-p}\sum_{t=p+1}^{N}\{|y_t - \hat{y}_t|/y_t\}, \\
MSRE &= \frac{1}{N-p}\sum_{t=p+1}^{N}\{(y_t - \hat{y}_t)^2/y_t\},
\end{aligned}
\tag{11}
$$

where N is the sample size and p is the number of lags. Smaller values of MARE and MSRE indicate a better fit.

Canadian Lynx data-Previous models

- Park, et al. (2009) fitted the following model to the $\log_{10}$ of the *Canadian Lynx Data*,

$$y_t = 0.99 + 0.52y_{t-1} + 0.75d_{1,t} - 0.39d_{1,t-1} - 0.13cos_t + 0.07cos_{1,t-1} \tag{12}$$

where $cos_t = \cos(3.87d_{1,t} - 3.44)$ and $d_{1,t} = \widehat{\boldsymbol{\eta}}_1^T \mathbf{Y}_{t-1}$ such that $\widehat{\boldsymbol{\eta}}_1 = (0.9317, -0.0761, -0.1777, -0.3074)^T$.

- There are three self-exciting threshold autoregressive (SETAR) models already fitted for *Canadian Lynx data.*

- Tong (1990) fitted SETAR(2;2,2) to the $\log_{10}$ of this data

$$\begin{aligned} y_t =& \{0.62 + 1.25y_{t-1} - 0.43y_{t-2} + \epsilon_t^{(1)}\}\mathbf{I}(y_{t-2} \leq 3.25) \\ &+ \{2.25 + 1.52y_{t-1} - 1.24y_{t-2} + \epsilon_t^{(2)}\}\mathbf{I}(y_{t-2} > 3.25), \end{aligned} \tag{13}$$

where $\{\varepsilon_t^{(1)}\}$, and $\{\varepsilon_t^{(2)}\}$ be a white noise sequences independent of $(y_{t-1}, y_{t-1})^T$ and they are independent from each other.

Canadian Lynx data-Previous models Count...

- Tong (1990) fitted SETAR(2;7,2) to the $\log_{10}$ of if as

$$y_t = \{0.546 + 1.032y_{t-1} - 0.173y_{t-2} + 0.171y_{t-3} - 0.43y_{t-4} + 0.332y_{t-5} - 0.284y_{t-6} + 0.210y_{t-7} + \epsilon_t^{(1)}\}I(y_{t-2} \leq +3.116) + \{2.632 + 1.492y_{t-1} - 1.324y_{t-2} + \epsilon_t^{(2)}\}I(y_{t-2} > 3.116). \quad (14)$$

- Tsay (1988) developed SETAR(3;1,7,2) with three thresholds as

$$y_t = \{0.083 + 1.096y_{t-1} + \epsilon_t^{(1)}\}I(y_{t-2} \leq 2.373) + \{0.63 + 0.96y_{t-1} - 0.11y_{t-2} + 0.23y_{t-3} - 0.61y_{t-4} + 0.48y_{t-5} - 0.39y_{t-6} + 0.28y_{y-7} + \epsilon_t^{(2)}\}I(2.373 < y_{t-2} \leq 3.154) + \{2.323 + 1.530y_{t-1} - 1.266y_{t-2} + \epsilon_t^{(3)}\}I(y_{t-2} > 3.154), \quad (15)$$

Canadian Lynx data- Estimating p and d

Table: Mean distance obtained for each pair (p,d). Here * indicates the optimal value of (p, d), i.e., $(\hat{p}, \hat{d})$=(2,1)

p	d=1	d=2	d=3	d=4
2	0.0001*	0		
3	0.0040	0.0032	0	
4	0.0020	0.0148	0.0115	0
5	0.0043	0.1634	0.0677	0.0100

- The FMTS fitted model is,

$$\widehat{y}_{t-1} = 0.4942 + 1.1706u_{1,t} + 0.0877sin(u_{1,t}) + 0.4495y_{t-10} + 0.2707y_{t-20} \\ -0.5262u_{1,t-10} - 0.0394\sin(u_{1,t-10}) - 0.3169u_{1,t-20} - 0.0237\sin(u_{1,t-20}) \quad (16)$$

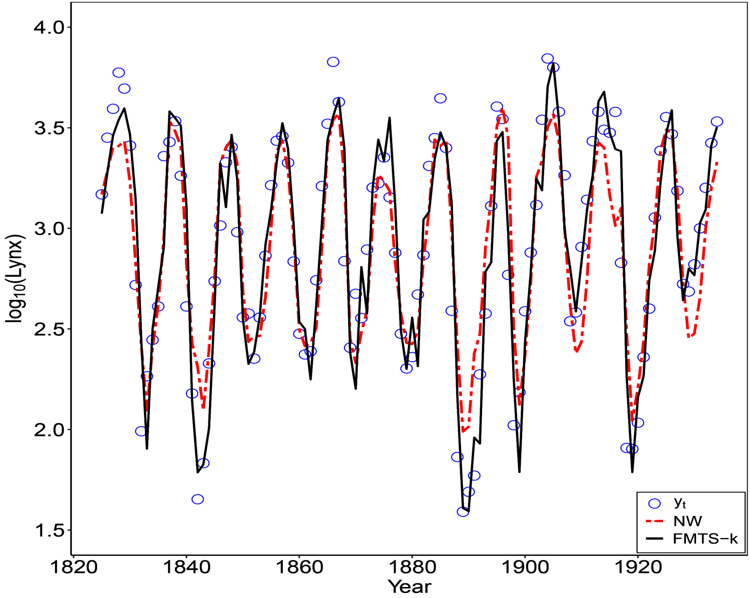
where $u_{1,t} = \hat{\eta}_1^T \mathbf{Y}_{t-1}$ and $\widehat{\eta}_1 = (0.9576 \quad -0.2880)^T$.

Canadian Lynx data- Model comparison.

Table: Comparison of the performance of different time series models for the Canadian Lynx data

Model Equation	MARE	MSPE	n	no. of parameters
FMTS - (16)	0.05826	0.01581	110	8
MSBC - (12)	0.0743	0.0254	110	10
Tong's - (13)	0.0594	0.0162	110	10
Tong's - (14)	0.0564	0.0137	110	13
Tsay's - (15)	0.0557	0.0130	110	17

Canadian Lynx data-True vs Fitted values



Acknowledgement

- This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC). Moreover, I get the allocation for PSC by Y.S. Samadi on the charge number TG-DMS200007.

- Meanwhile this work also used the Bigdawg cluster at Souther Illinois University Carbondale.

References

Cook R. D., (2018).
Principal Components, Sufficient Dimension Reduction, and Envelopes.
Annual Review of Statistics and Its Applications. 5, 533–559.

Cook R. D., and Li, B., (2002).
Dimension Reduction for Conditional Mean in Regression.
The Annals of Statistics. 30, 455-474.

Dehling H., (2006).
Limit Theorems for Dependent U-statistics. Lecture Notes in Statitics.
Vol 187, Springer, New York.

Folland G. B., (1992).
Fourier Analysis and Its Applications.
Wadsworth and Brooks/Cole Advance Books and Software.

Lehmann E. L., (1998).
Elements of Large Sample Theory.
Springer texts in statistics.

Park J. H., and Samadi S. Yaser, (2014).
Heteroscedastic modelling via the autoregressive conditional variance subspace

References Count...

Park J. H., and Samadi S. Yaser, (2020).
Dimension Reduction for the conditional mean and varaince functions in time series.
Scandinavian Journal of Statistics, 47, 134-155.

Park J. H., Sriram T. N., and Yin X., (2009).
Central Mean subspace in Time Series.
Journal of American Statistical Association, 18, 717-730.

Park J. H., Sriram T. N., and Yin X.,(2010).
Dimension Reduction in Time Series.
Statistica Sinica, 20, 747-770.

Wolfgang H., Joel H., and Kreiss J.P., (2003).
Bootstrap Methods for Time Series.
International Statistical Institute (ISI), Vol. 71, No. 2 (Aug., 2003), pp. 435-459.

References Count...

Zeng P., and Zhu Y., (2010).
An Integral Transform Method for Estimating the Central Mean and Central Subspaces.
Journal of Multivariate Analysis. 101, 1, 271-290.

Zhu Y., and Zeng P., (2006).
Fourier Methods for Estimating the Central Subspace and Central Mean Subspace in Regression.
Journal of the American Statistical Association. 101, 476, 1638-1651.

Questions?

Thank you