MONASH University

MONASH BUSINESS SCHOOL

Seasonal functional autoregressive models

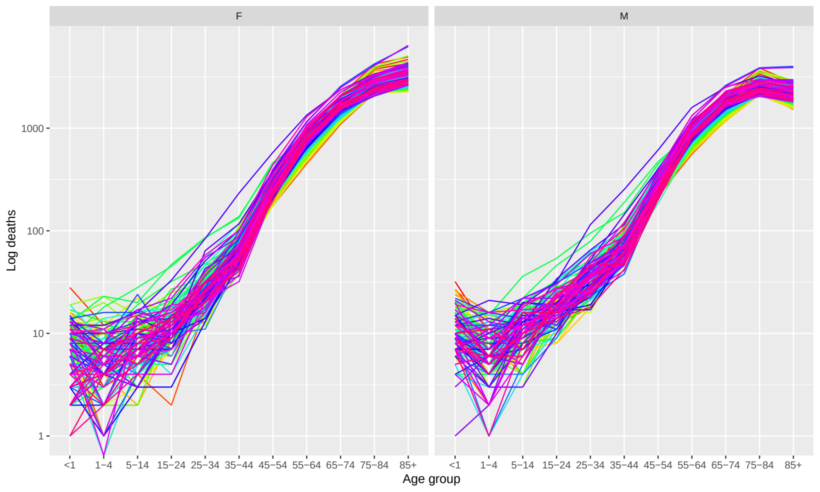
Rob J Hyndman, Maryam Hashemi, Hossein Haghbin & Atefeh Zamani

8 November 2018

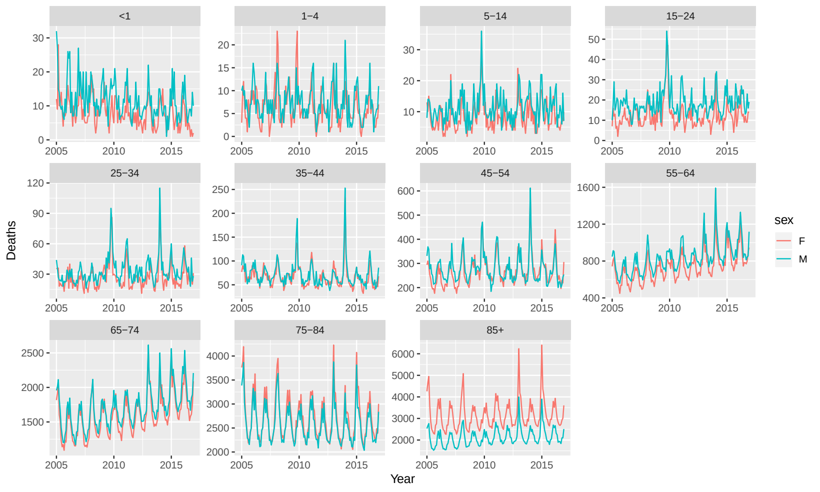
Outline

Outline

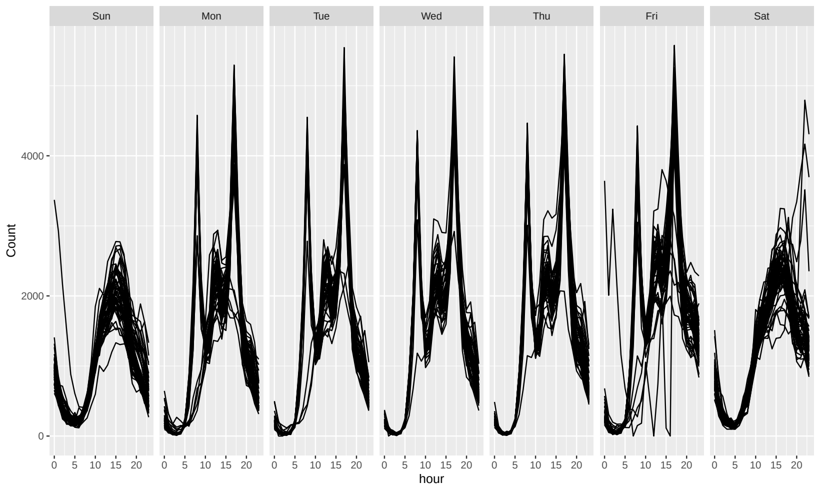
Monthly flu/respiratory mortality (US)



Monthly flu/respiratory mortality (US)



Hourly pedestrian count at Flinders St



Examples

Notation

$X_t(u)$ where $t = 1, \dots, T$ indexes regularly spaced time and u is a continuous variable in $\mathbb{R}$ or $\mathbb{R}^2$

1. $X_t(u)$ = mortality rate for people aged u in month t.
2. $X_t(u)$ = vegetation index at location u in month t, measured by average satellite observations.

Sometimes u may denote a second time variable.

3. $X_t(u)$ = pedestrian count observed every hour. u denotes time-of-day, t denotes day.

Seasonality

Notation

$X_t(u)$ where $t = 1, \ldots, T$ indexes regularly spaced time and u is a continuous variable in $\mathbb{R}$ or $\mathbb{R}^2$

Seasonality occurs when $X_t(u)$ is influenced by seasonal factors (e.g., the quarter of the year, the month, the day of the week, etc.).

Functional autoregression

FAR(p) processes – introduced by Bosq (2000)

$$X_t = \phi_1(X_{t-1}) + \cdots + \phi_p(X_{t-p}) + \varepsilon_t,$$

- $\{\varepsilon_t\}$ is a functional H-white noise process
- $\phi_p \neq 0$
- Some stationarity conditions on $\phi_1, \ldots, \phi_p$.
- Assume: ϕ_j are Hilbert-Schmidt operators in $\mathcal{L}(H)$

H = separable real Hilbert space of square integrable functions.

$\mathcal{L}(H)$ = space of continuous linear operators from H to H.

Seasonal univariate autoregression

AR(p) processes

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t.$$

- $\{\varepsilon_t\}$ is a real white noise process.
- $\phi_p \neq 0$
- Some stationarity conditions on $\phi_1, \ldots, \phi_p$

Seasonal univariate autoregression

AR(p) processes

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t.$$

- $\{\varepsilon_t\}$ is a real white noise process.
- $\phi_p \neq 0$
- Some stationarity conditions on $\phi_1, \ldots, \phi_p$

SAR(P)$_S$ processes

$$Y_t = \Phi_1 Y_{t-S} + \Phi_2 Y_{t-2S} + \cdots + \Phi_P Y_{t-PS} + \varepsilon_t.$$

- $\{\varepsilon_t\}$ is a real white noise process.
- $\Phi_P \neq 0$
- Some stationarity conditions on $\Phi_1, \ldots, \Phi_P$

Outline

$\text{SFAR}(1)_S$ processes

$$X_t = \Phi(X_{t-S}) + \varepsilon_t,$$

- $\{\varepsilon_t\}$ is a functional H-white noise process
- $\Phi \neq 0$
- Assume: Φ is Hilbert-Schmidt operator in $\mathcal{L}(H)$

$\text{SFAR}(1)_S$ processes

$$X_t = \Phi(X_{t-S}) + \varepsilon_t,$$

- $\{\varepsilon_t\}$ is a functional H-white noise process
- $\Phi \neq 0$
- Assume: Φ is Hilbert-Schmidt operator in $\mathcal{L}(H)$

Stationarity of a $\text{SFAR}(1)_S$ process

If there exists an integer $M \geq 1$ such that $\|\Phi^M\|_{\mathcal{L}} < 1$,

then $$X_t = \Phi(X_{t-S}) + \varepsilon_t,$$

has a unique stationary solution given by

$$X_t = \sum_{j=0}^{\infty} \Phi^j(\varepsilon_{t-jS}),$$

where the series converges in L^2_H with probability 1.

$SFAR(1)_S$ processes

$$X_t = \Phi(X_{t-S}) + \varepsilon_t,$$

- $\{\varepsilon_t\}$ is a functional H-white noise process
- $\Phi \neq 0$
- Assume: Φ is Hilbert-Schmidt operator in $\mathcal{L}(H)$

Link to $SAR(1)_S$ processes

Let

$$\Phi = \sum_{j=1} \alpha_j e_j \otimes e_j$$

be a symmetric compact operator on H.

Then, X_t is a $SFAR(1)_S$ process if and only if $\langle X_t, e_k \rangle$ is a $SAR(1)_S$ process.

SFAR(1)$_S$ processes

Let $\mathbf{Y}_t = (X_t, \ldots, X_{t-S+1})'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_t, 0, \ldots, 0)'$.

Define operator $\boldsymbol{\rho}$ on H^S:

$$\boldsymbol{\rho} = \begin{bmatrix} 0 & 0 & \ldots & 0 & \Phi \\ I & 0 & \ldots & 0 & 0 \\ 0 & I & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \ldots & I & 0 \end{bmatrix}, \qquad \text{where } I \text{ is identity operator.}$$

Lemma

If $\boldsymbol{X}$ is SFAR(1)$_S$ process, associated with (ε, Φ), then $\boldsymbol{Y}$ is FAR(1) with values in product Hilbert space H^S associated with $(\boldsymbol{\varepsilon}, \boldsymbol{\rho})$, i.e., $\boldsymbol{Y}_n = \boldsymbol{\rho}\boldsymbol{Y}_{n-1} + \boldsymbol{\varepsilon}_n$.

SFAR(1)$_S$ processes

Theorem

Let $\mathcal{L}(H^S)$ be space of bounded linear operators on H^S equipped with norm $\|\cdot\|_{\mathcal{L}^S}$. If

$$\|\boldsymbol{\rho}^M\|_{\mathcal{L}^S} < 1, \quad \text{for some } M \geq 1,$$

then SFAR(1)$_S$ has unique stationary solution given by

$$X_t = \sum_{j=0}^{\infty} (\boldsymbol{\pi\rho})(\boldsymbol{\varepsilon}_{t-j}),$$

where the series converges in $L^2_{H^S}$ with probability 1 and $\boldsymbol{\pi}$ is the projector of H^S onto H, defined as $\boldsymbol{\pi}(x_1, \ldots, x_S) = x_1, \quad (x_1, \ldots, x_S) \in H^S$.

Limit Theorems for SFAR(1)$_S$ processes

Theorem: Law of large numbers for X

If $\boldsymbol{X}$ is a standard SFAR(1)$_S$ then, as $T \to \infty$,

$$\frac{T^{0.25}}{(\log T)^{\beta}} \frac{(X_1 + X_2 + \cdots + X_T)}{T} \to 0, \qquad \text{for } \beta > 0.5.$$

Central Limit Theorem

Let $\boldsymbol{X}$ be a standard SFAR(1)$_S$ associated with a strong white noise ε and such that $I - \Phi$ is invertible. Then

$$\frac{(X_1 + X_2 + \cdots + X_T)}{\sqrt{T}} \to \mathcal{N}(0, \Gamma),$$

where $\Gamma = (I - \Phi)^{-1} C_{\varepsilon} (I - \Phi^{*})^{-1}$.

Outline

Method of Moments

Covariance operator

$$C_k^X = \mathrm{E}(X_t \otimes X_{t-k}) \qquad \hat{C}_k^X = \frac{1}{T}\sum_{t=k+1}^{T} X_t \otimes X_{t-k}$$

Eigendecomposition

For any $x \in H$,

$$\Phi(x) = \sum_{j=1}^{\infty} \langle x, \nu_j \rangle \, \Phi(\nu_j) = \sum_{j=1}^{\infty} \frac{C_S^X(\nu_j)}{\lambda_j} \langle x, \nu_j \rangle$$

- Estimate λ_j and ν_j from $\hat{C}_0^X$.
- $\hat{\Phi}(x) = \sum_{j=1}^{J} \frac{\hat{C}_S^X(\hat{\nu}_j)}{\hat{\lambda}_j} \langle x, \hat{\nu}_j \rangle$

Unconditional Least Squares

Let $X_{tk} = \langle X_t, \nu_k \rangle$ be projection of the tth observation onto the kth largest FPC.

$$X_{tk} = \sum_{j=1}^{p} \Phi_{kj} X_{t-S,j} + \delta_{tk}, \qquad k = 1, \ldots, p$$

where $\Phi_{kj} = \langle \Phi(\nu_j), \nu_k \rangle$. Note that the δ_{tk} are not iid.

- Replace ν_k by $\hat{\nu}_k$ to get $\hat{X}_{tk}$

Unconditional Least Squares

Set $\boldsymbol{X}_t = (\hat{X}_{t1}, \ldots, \hat{X}_{tp})'$, $\boldsymbol{\delta}_t = (\delta_{t1}, \ldots, \delta_{tp})'$,

$\boldsymbol{\Phi} = (\Phi_{11}, \ldots, \Phi_{1p}, \Phi_{21}, \ldots, \Phi_{2p}, \ldots, \Phi_{p1}, \ldots, \Phi_{pp})'$.

$$\boldsymbol{Z}_t = \begin{bmatrix} \boldsymbol{X}_t' & \boldsymbol{0}_p' & \ldots & \boldsymbol{0}_p' \\ \boldsymbol{0}_p' & \boldsymbol{X}_t' & \ldots & \boldsymbol{0}_p' \\ \vdots & \vdots & \ldots & \vdots \\ \boldsymbol{0}_p' & \boldsymbol{0}_p' & \ldots & \boldsymbol{X}_t' \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix}, \quad \boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_N \end{bmatrix}, \quad \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_{1-S} \\ \boldsymbol{Z}_{2-S} \\ \vdots \\ \boldsymbol{Z}_{N-S} \end{bmatrix},$$

$$\hat{\boldsymbol{\Phi}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$$

The Kargin-Onatski Method

Find A, approximating Φ, minimizing $\mathrm{E}\|X_t - A(X_{t-S})\|^2$.

Let $\hat{C}_\alpha = \hat{C}_0 + \alpha I, \quad \alpha > 0$

Let $\{\upsilon_{\alpha,i}\}$ be eigenfunctions of $\hat{C}_\alpha^{-1/2}\hat{C}_S'\hat{C}_S\hat{C}_\alpha^{-1/2}$, corresponding to eigenvalues $\{\hat{u}_{\alpha,i}\}, \quad \hat{u}_{\alpha,j} > \hat{u}_{\alpha,j+1}$.

$$\hat{\Phi}_{\alpha,k_T} = \sum_{i=1}^{k_T} \hat{C}_\alpha^{-1/2}\upsilon_{\alpha,i} \otimes \hat{C}_S\hat{C}_\alpha^{-1/2}\upsilon_{\alpha,i}.$$

$\hat{\Phi}_{\alpha,k_T}$ is a consistent estimator of Φ if $\{k_T\}$ is sequence of positive integers such that $KT^{-1/4} \le k_T \le T$, for some $K > 0$ and $\alpha \sim T^{-1/6}$.

Simulations

Let $\{X_t\}$ follow a SFAR$(1)_S$ model,

$$X_t(u) = \Phi X_{t-S}(u) + \varepsilon_t(u), \quad t = 1, \cdots, T,$$

where Φ is an integral operator with *parabolic* kernel

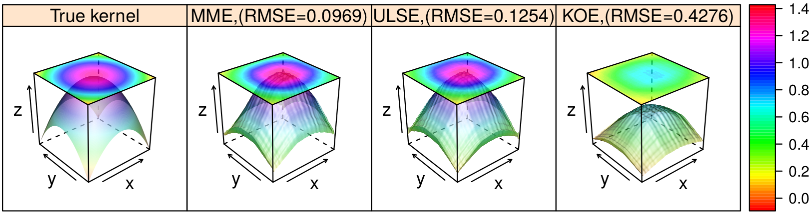
$$k_\Phi(u, v) = \gamma_0 \left(2 - (2u - 1)^2 - (2v - 1)^2\right),$$

and γ_0 is such that $\|\Phi\|_{\mathcal{S}}^2 = \int_0^1 \int_0^1 |k_\Phi(u, v)|^2 \, dudv = 0.9$.

- White noise terms $\varepsilon_t(u)$ are independent standard BM on $[0, 1]$ with variance 0.05.
- B = 1000 trajectories simulated
- Φ estimated using MME, ULSE and KOE.
- RMSE = $\sqrt{\frac{1}{B} \sum_{i=1}^{B} \|\hat{\Phi}_i - \Phi\|_{\mathcal{S}}^2}$

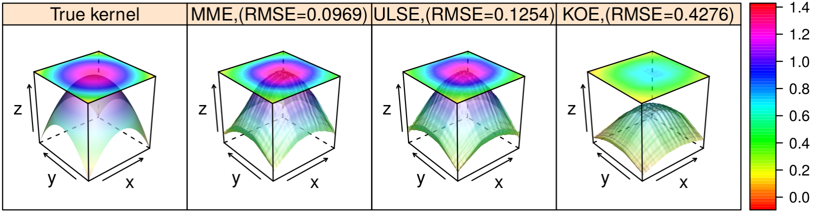
Simulations

$T = 200$, $\|\Phi\|_{\mathcal{S}} = 0.9$ and $k_T = 1$.

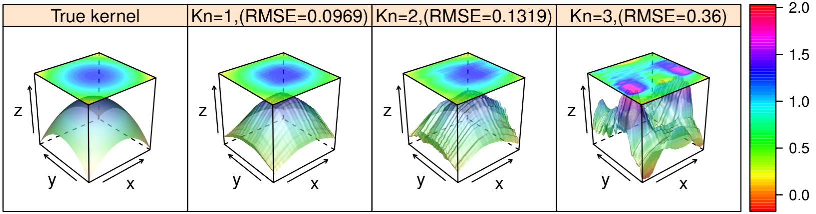


Simulations

$T = 200$, $\|\Phi\|_{\mathcal{S}} = 0.9$ and $k_T = 1$.



$T = 200$, $\|\Phi\|_{\mathcal{S}} = 0.9$ and $k_T = 1, 2, 3$. MME only



Simulations: RMSE

T	k_T	$\|\Phi\|_{\mathcal{S}} = 0.1$			$\|\Phi\|_{\mathcal{S}} = 0.5$			$\|\Phi\|_{\mathcal{S}} = 0.9$		
		MME	ULSE	KOE	MME	ULSE	KOE	MME	ULSE	KOE
50	1	0.1750	0.1645	0.0951	0.2403	0.2838	0.3716	0.1986	0.2323	0.5096
	2	0.5484	0.5189	0.0959	0.4931	0.7000	0.3720	0.4387	1.0381	0.5099
	3	1.0239	0.9657	0.0961	0.9988	1.1478	0.3721	1.0435	1.7282	0.5099
	4	1.5573	1.4934	0.0962	1.5340	1.6513	0.3721	1.6382	2.4725	0.5099
100	1	0.1222	0.1183	0.0861	0.2050	0.2579	0.3539	0.1387	0.1709	0.4134
	2	0.3662	0.3598	0.0866	0.3325	0.6087	0.3541	0.2743	0.9728	0.4136
	3	0.6830	0.6798	0.0868	0.6661	0.8723	0.3541	0.6694	1.3925	0.4136
	4	1.0645	1.0243	0.0868	1.0245	1.1973	0.3542	1.0193	1.9377	0.4136
150	1	0.1033	0.1027	0.0825	0.1946	0.2505	0.3460	0.1205	0.1517	0.3735
	2	0.2903	0.2900	0.0830	0.2666	0.5704	0.3462	0.2149	0.9478	0.3737
	3	0.5533	0.5449	0.0831	0.5387	0.7601	0.3462	0.5272	1.2493	0.3736
	4	0.8560	0.8237	0.0831	0.8256	1.0040	0.3462	0.8106	1.6683	0.3736
200	1	0.0917	0.0935	0.0798	0.1879	0.2457	0.3393	0.1114	0.1419	0.3496
	2	0.2490	0.2610	0.0803	0.2285	0.5568	0.3394	0.1818	0.9411	0.3497
	3	0.4790	0.4745	0.0804	0.4684	0.7047	0.3394	0.4542	1.1896	0.3497
	4	0.7438	0.7127	0.0804	0.7199	0.9042	0.3394	0.7040	1.5134	0.3497

Outline

Forecasting

Let $\boldsymbol{X}_T = (X_1, X_2, \ldots, X_T)'$.

Let G be closure of $\left\{\ell_0 \boldsymbol{X}_T;\ \ell_0 \in \mathcal{L}\left(H^T, H\right)\right\}$.

Best linear h-step predictor of X_{T+h} is projection of X_{T+h} on G, i.e., $\hat{X}_{T+h} = P_G X_{T+h}$.

Forecasting

Let $\boldsymbol{X}_T = (X_1, X_2, \ldots, X_T)'$.

Let G be closure of $\left\{\ell_0 \boldsymbol{X}_T;\ \ell_0 \in \mathcal{L}\left(H^T, H\right)\right\}$.

Best linear h-step predictor of X_{T+h} is projection of X_{T+h} on G, i.e., $\hat{X}_{T+h} = P_G X_{T+h}$.

Proposition (based on Bosq 2014)

For $h \in \mathbb{N}$ the following statements are equivalent:

1. There exists $\ell_0 \in \mathcal{L}\left(H^T, H\right)$ such that $C_{\boldsymbol{X}_T, X_{T+h}} = \ell_0 C_{\boldsymbol{X}_T}$.
2. $P_G X_{T+h} = \ell_0 \boldsymbol{X}_T$ for some $\ell_0 \in \mathcal{L}\left(H^T, H\right)$.

Forecasting

Let $\boldsymbol{X}_T = (X_1, X_2, \dots, X_T)'$.

Let G be closure of $\left\{\ell_0 \boldsymbol{X}_T;\ \ell_0 \in \mathcal{L}\left(H^T, H\right)\right\}$.

Best linear h-step predictor of X_{T+h} is projection of X_{T+h} on G, i.e., $\hat{X}_{T+h} = P_G X_{T+h}$.

Proposition (based on Bosq 2014)

For $h \in \mathbb{N}$ the following statements are equivalent:

1. There exists $\ell_0 \in \mathcal{L}\left(H^T, H\right)$ such that $C_{\boldsymbol{X}_T, X_{T+h}} = \ell_0 C_{\boldsymbol{X}_T}$.
2. $P_G X_{T+h} = \ell_0 \boldsymbol{X}_T$ for some $\ell_0 \in \mathcal{L}\left(H^T, H\right)$.

How to find $\ell_0 \in \mathcal{L}\left(H^T, H\right)$ such that $C_{\boldsymbol{X}_T, X_{T+h}} = \ell_0 C_{\boldsymbol{X}_T}$?

Forecasting

Forecast horizon $h = aS + c$, $a \geq 0$ and $0 \leq c < S$.

$$C_{\boldsymbol{X}_T, X_{T+h}}(\boldsymbol{x}) = \mathrm{E}\left(\langle \boldsymbol{X}_T, \boldsymbol{x} \rangle_{H^T} X_{T+h}\right) = \Phi^{a+1}_{T-S+c} C_{\boldsymbol{X}_T}(\boldsymbol{x}),$$

where Φ^i_j is an T-vector of zeros with Φ^i in jth position.

$$\hat{X}_{T+h} = P_G X_{T+h} = \Phi^{a+1}_{T-S+c} \boldsymbol{X}_T = \Phi^{a+1} X_{T-S+c}$$

Based on KOE, 1-step ahead predictor of X_{T+1} is:

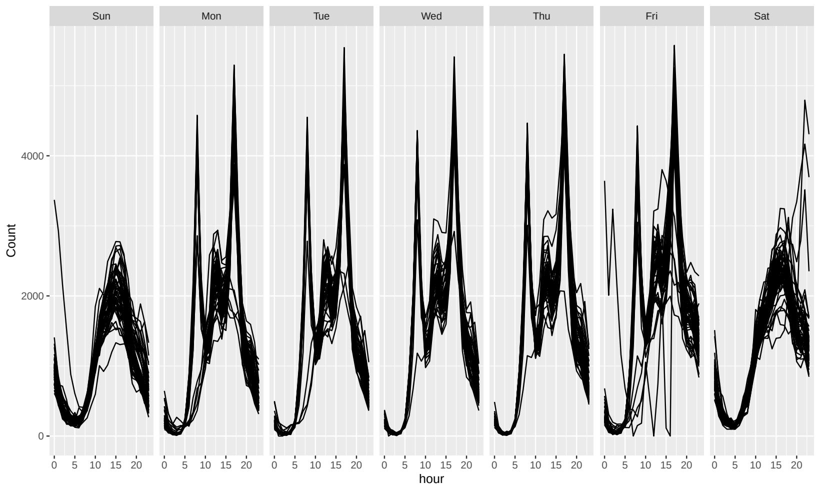
$$\hat{X}_{T+1} = \sum_{i=1}^{k_T} < X_{T-S+1}, \hat{z}_{\alpha,i} > \hat{C}_S(\hat{z}_{\alpha,i}),$$

where $$\hat{z}_{\alpha,i} = \sum_{j=1}^{q} \hat{u}_j^{-1/2} \langle v_{\alpha,i}, \hat{\nu}_j \rangle \hat{\nu}_j + \alpha v_{\alpha,i}.$$

Select q by cumulative variance method and set $k_T = q$.

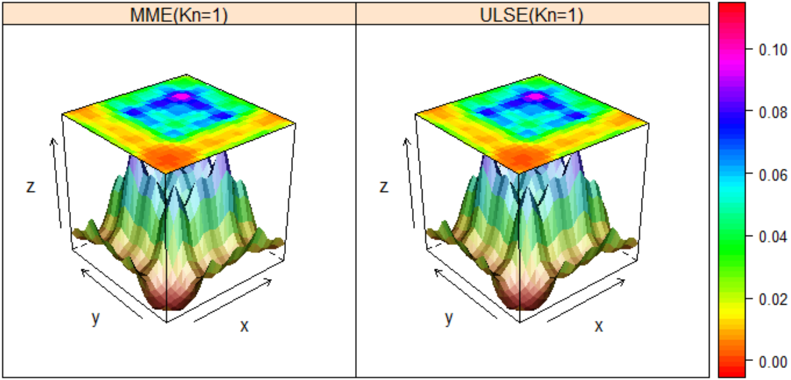
Outline

Application: pedestrian counts



Application: pedestrian counts

The estimated kernel of the autocorrelation operator using MME and ULSE methods.



Application: pedestrian counts

1-step predictors for the last 7 days of the dataset

	MAE		RMSE	
k_T	MME	ULSE	MME	ULSE
1	198.7	197.9	201.7	201.0
2	202.8	99.1	205.6	99.7
3	315.3	199.8	319.7	207.0
4	418.4	155.8	423.4	157.5
5	508.0	267.5	515.0	301.6
6	645.6	168.0	655.2	169.5

Outline

SFAR$(P)_S$ processes

Definition

A sequence $\{X_t; t \in \mathbb{Z}\}$ of functional random variables is said to be a seasonal functional autoregressive process of order P with seasonality S if

$$X_t - \mu = \Phi_1 (X_{t-S} - \mu) + \cdots + \Phi_P (X_{t-PS} - \mu) + \varepsilon_t,$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is H-white noise, $\mu \in H$, and $\Phi_1, \ldots, \Phi_P \in \mathcal{L}(H)$, with $\Phi_P \neq 0$.

$\text{SFAR}(P)_S$ processes

Let $Y_t = (X_t, X_{t-S}, \ldots, X_{t-PS+S})'$, $\varepsilon_t' = (\varepsilon_t, 0, \ldots, 0)'$, and

$$\mathbf{\Phi} = \begin{bmatrix} \Phi_1 & \Phi_2 & \ldots & \Phi_P \\ I & 0 & \ldots & 0 \\ 0 & I & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix},$$

where I and 0 denote identity and zero operator on H.

Lemma

If X is a $\text{SFAR}(P)_S$ associated with associated with $(\varepsilon, \phi_1, \ldots, \phi_P)$, then Y is a $\text{SFAR}(1)_S$ with values in the product Hilbert space H^P associated with $\left(\varepsilon', \boldsymbol{\phi}\right)$.

SFAR$(P)_S$ processes

Theorem

Let X_n be a SFAR$(P)_S$ zero-mean process associated with $(\varepsilon, \phi_1, \phi_2, \ldots, \phi_P)$. Suppose that there exist $\nu \in H$ and $\alpha_1, \ldots, \alpha_P \in \mathbb{R}$, $\alpha_P \neq 0$, such that $\phi_j(\nu) = \alpha_j \nu_j$, $j = 1, \ldots, P$ and $E\langle \varepsilon_0, \nu \rangle^2 > 0$. Then, $\left(\langle X_t, \nu \rangle,\ t \in \mathbb{Z}\right)$ is a SAR(P) process, i.e.,

$$\langle X_t, \nu \rangle = \sum_{j=1}^{P} \alpha_j \langle X_{t-jS}, \nu \rangle + \langle \varepsilon_t, \nu \rangle, \qquad t \in \mathbb{Z}.$$

SFAR$(P)_S$ processes

Theorem

If X is a standard SFAR$(P)_S$ process, then

$$C_h = \sum_{j=1}^{P} \phi_j C_{h-jS}, \qquad h = 1, 2, \dots,$$

$$C_0 = \sum_{j=1}^{P} \phi_j C_{jS} + C_\varepsilon,$$

where C_ε is the covariance operator of the innovation process ε.