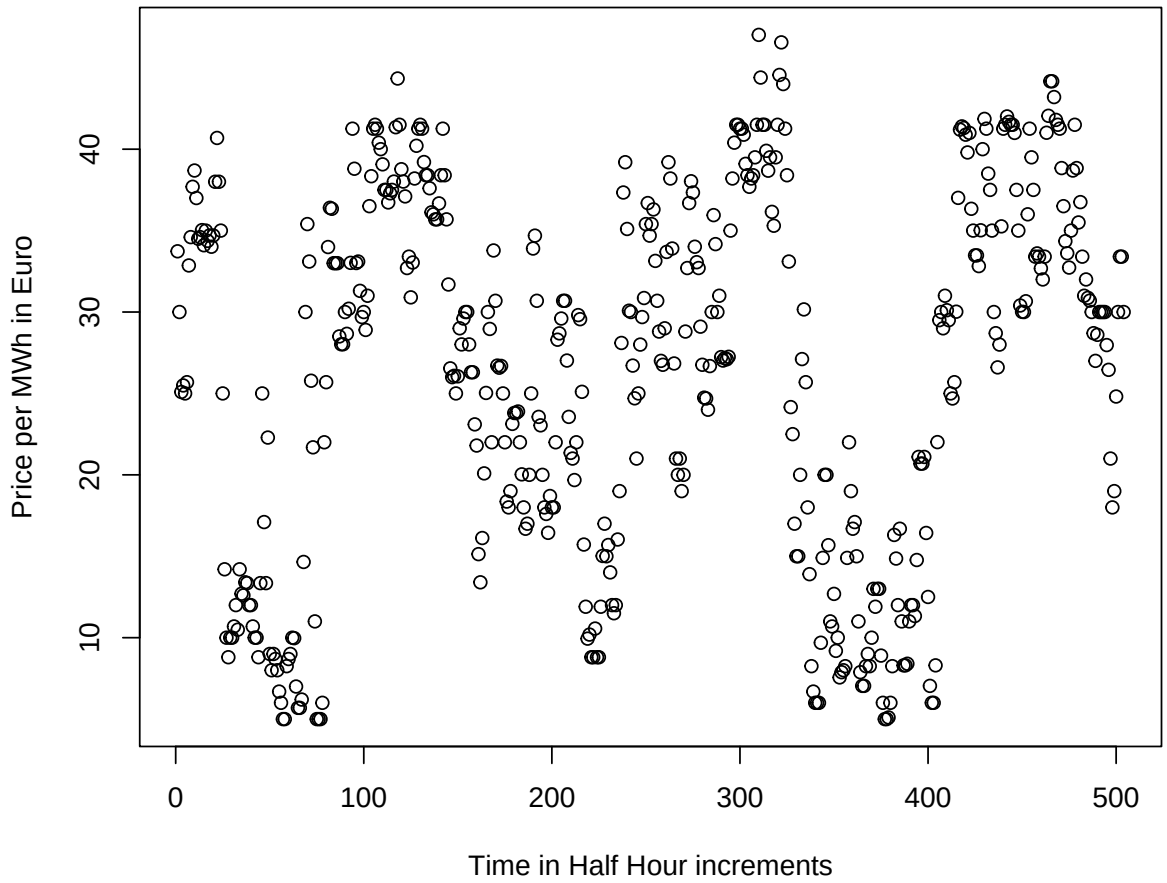
Functional Time Series Analysis in R

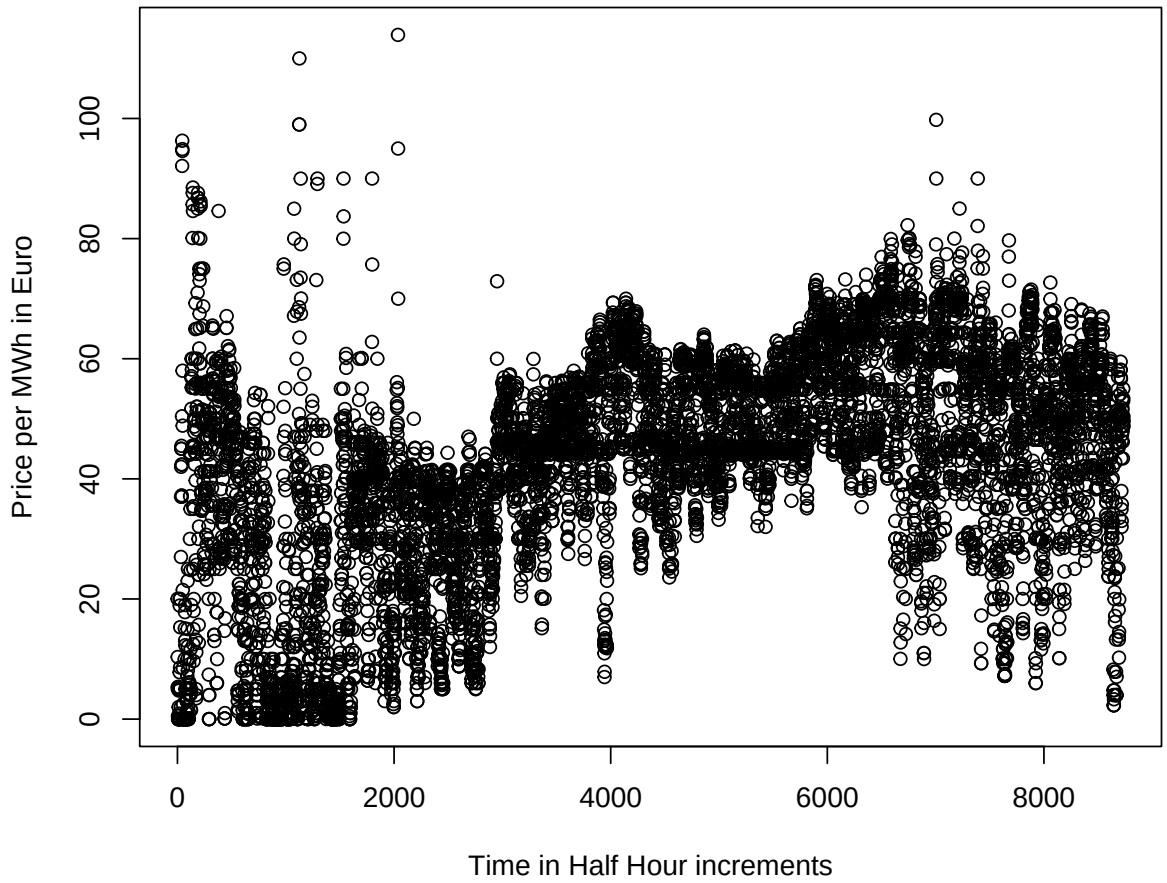
Greg Rice

Department of Statistics and Actuarial Science, University of Waterloo

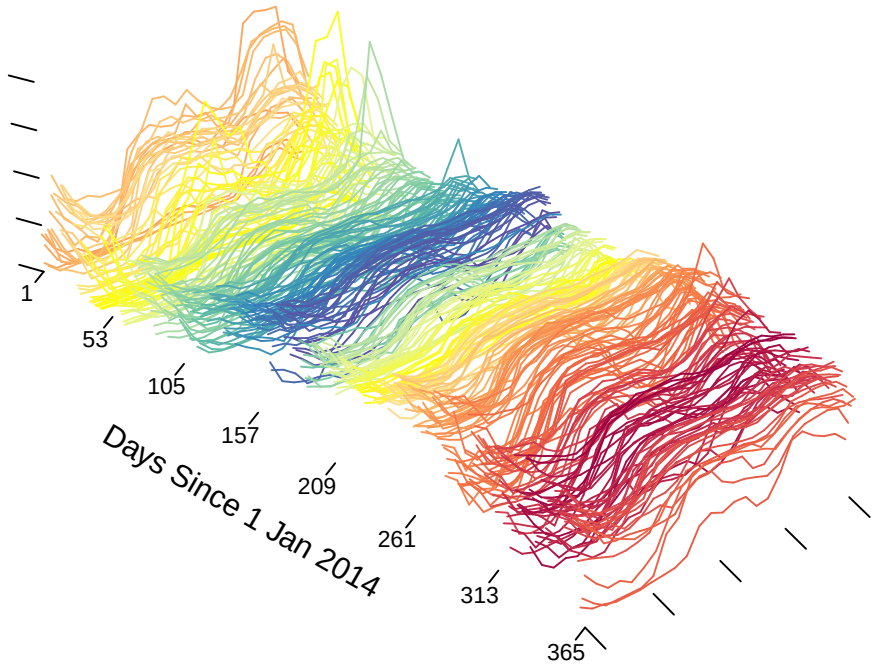
SSC Short Course, Sunday May 25th, 2025

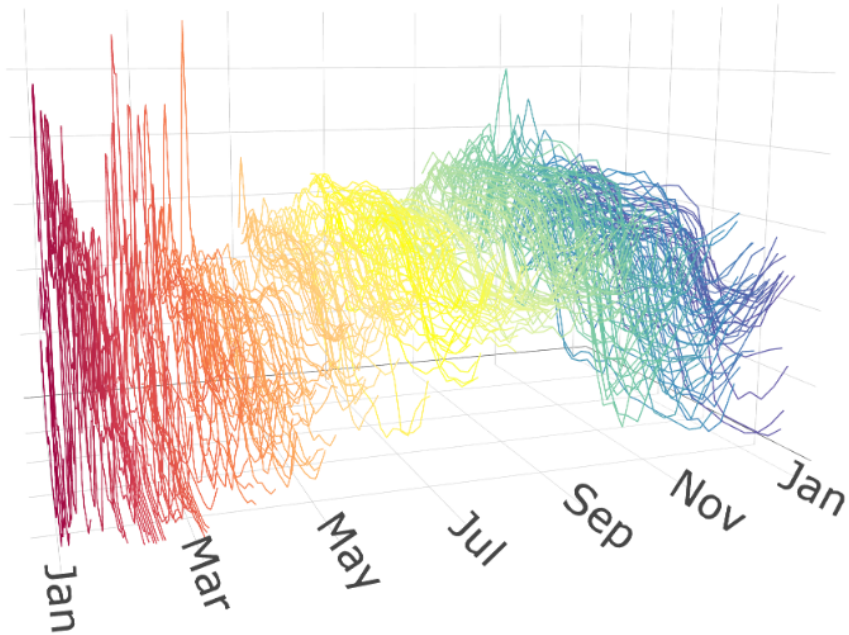


Hourly electricity spot price in Spain for three weeks in 2014.

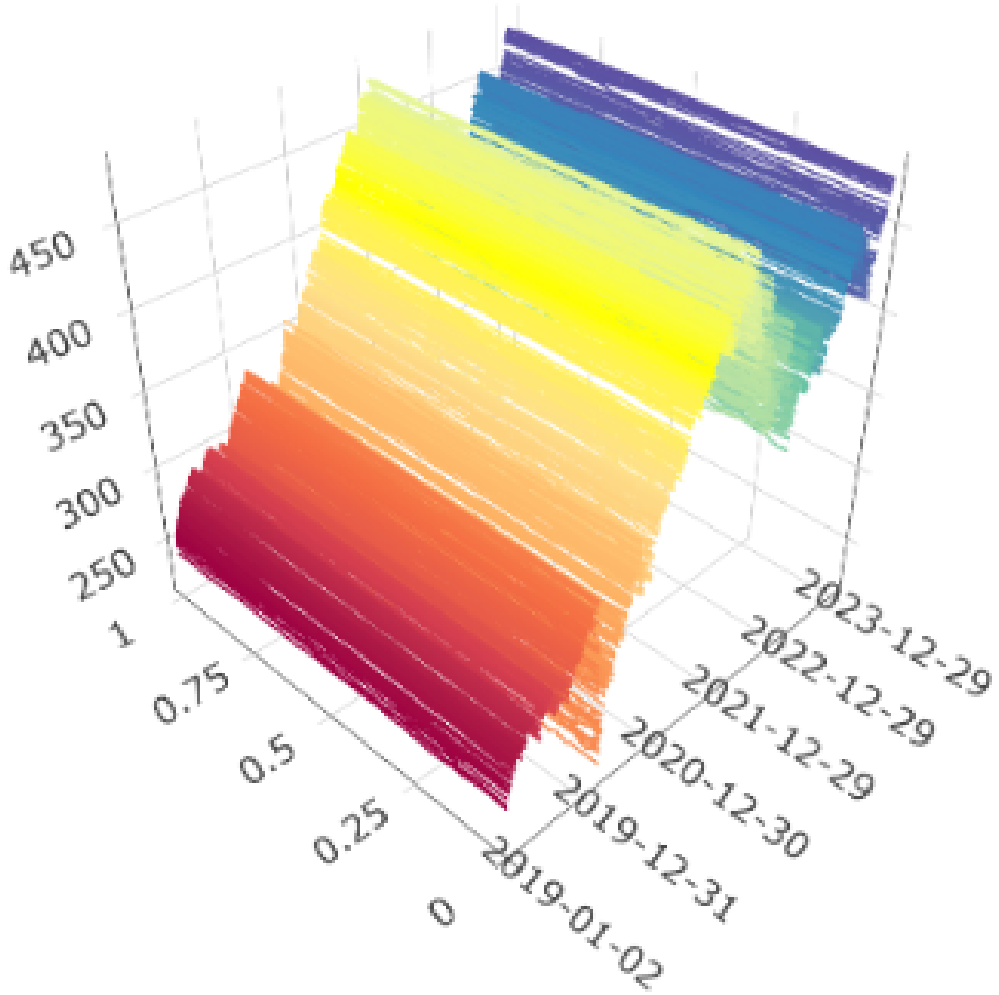


Hourly electricity spot price in Spain during the year 2014.

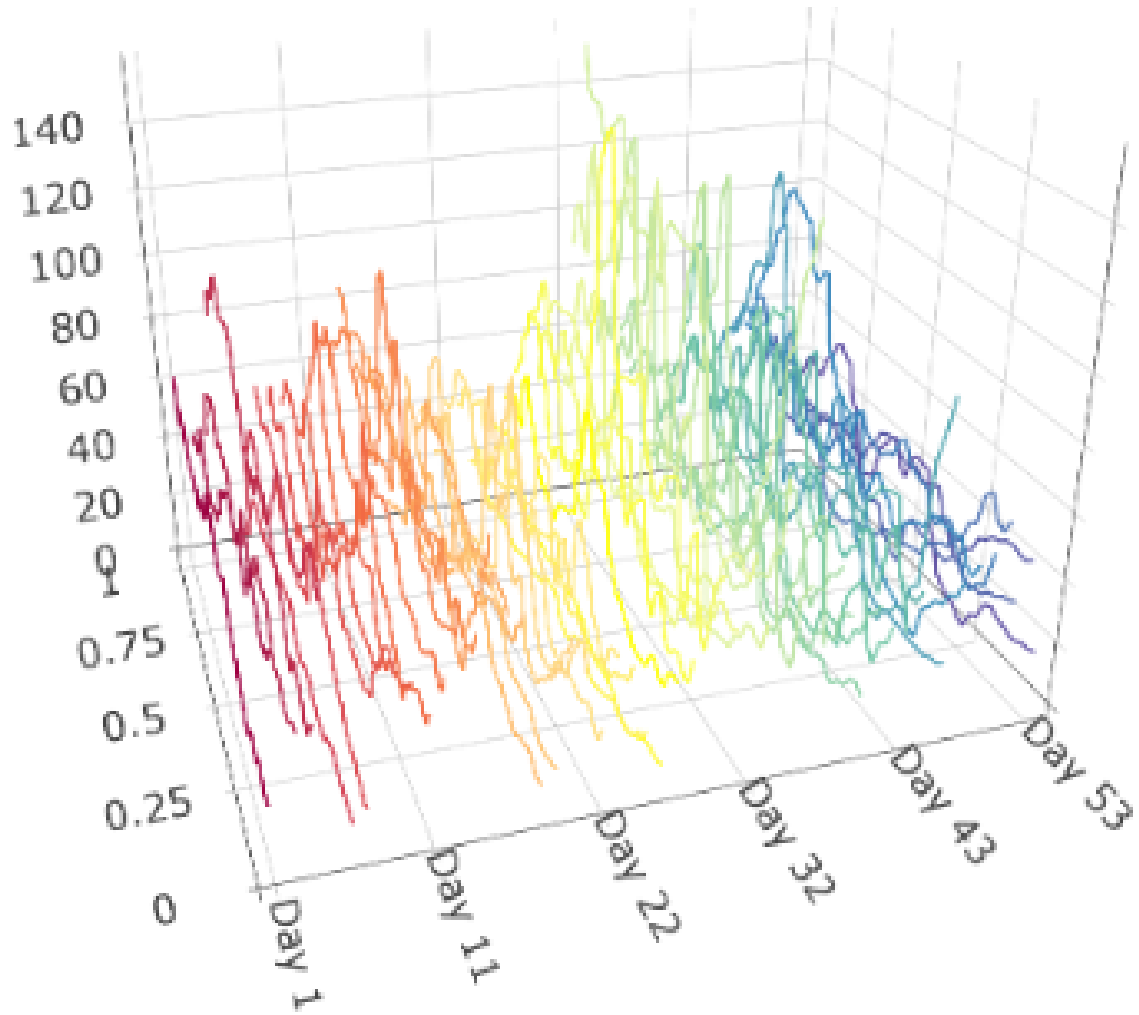

1
53
105
157
209
261
313
365
Days Since 1 Jan 2014



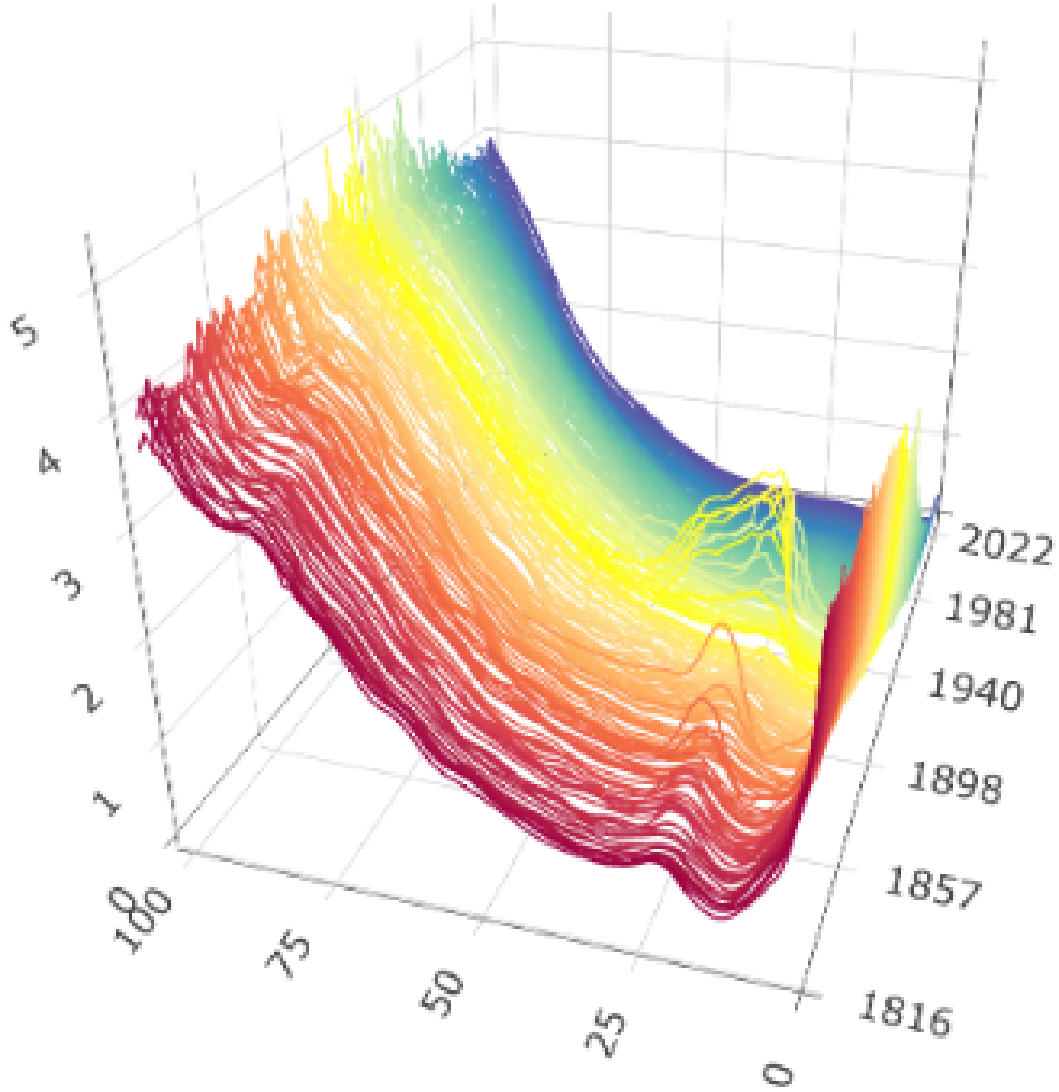
Another angle of hourly electricity spot price in Spain during the year 2014 visualized as a functional time series.



An FTS constructed from one minute resolution intraday prices from the S&P 500 .



Levels of particulate matter pollution in parts per million measured every 30 minutes in a Graz, Austria, visualized as a daily functional time series.



Log mortality rates for French males across ages 0 to 100 viewed as a functional time series over the years 1816 to 2022.

From a theoretical perspective we think of a time series as being a realization of length T from a discretely indexed *stochastic process*. I.e.

$$X_1, ..., X_T \sim \text{ Partial stretch observed from } \{X_i,\ i \in \mathbb{Z}\}$$

$$\mathbf{X}_1, ..., \mathbf{X}_T \in \mathbb{R}^d \sim \text{ Partial stretch observed from } \{\mathbf{X}_i \in \mathbb{R}^d,\ i \in \mathbb{Z}\}$$

When we think of data as being "functional" in nature, we think of them as taking their values in some space of functions rather than in $\mathbb{R}^d$. In the electricity spot price curve example

$$X_i : [0, 24] \mapsto \mathbb{R}.$$

Note: for functions defined over a compact interval we may as well assume the interval is $[0, 1]$.

Question: What are reasonable classes of functions $X_i : [0, 1] \mapsto \mathbb{R}$ might take its value in?

"Famous Function Spaces"

$$C[0,1] = \{f : [0,1] \mapsto \mathbb{R},\ f \text{ is continuous}\}.$$

$$L^2[0,1] = \left\{ f : [0,1] \mapsto \mathbb{R},\ \int_0^1 f^2(x)dx < \infty \right\}.$$

$$W^{1,2} = \left\{ f : [0,1] \mapsto \mathbb{R},\ \int_0^1 f^2(x)dx + \int_0^1 [f']^2(x)dx < \infty \right\}.$$

Definition

For a given function space $\mathcal{H}$, e.g. any of the above three examples, a *functional time series* (FTS) $X_1, ..., X_T$ is an observed stretch of data of length T of a function-valued stochastic process $\{X_i \in \mathcal{H},\ i \in \mathbb{Z}\}$.

The choice of space that we view the data as residing in is consequential in that it suggests how we might measure distance between functions.

$$f, g \in C[0,1], \quad \|f-g\|_\infty = \sup_{x \in [0,1]} |f(x) - g(x)|.$$

$$f, g \in L^2[0,1], \quad \|f-g\|_2 = \left(\int_0^1 [f(x) - g(x)]^2 dx \right)^{1/2}.$$

What ends up being a critical distinction between these spaces is that:

- $L^2[0,1]$ with distance measure $\|\cdot\|_2$ is a *separable Hilbert Space*.
- $C[0,1]$ with distance measure $\|\cdot\|_\infty$ (or even with distance $\|\cdot\|_2$) is NOT a separable Hilbert space.

What is so special about separable Hilbert space? Aside from the properties of linearity and completeness,

1. The distance measure is generated by an *inner product*:

$$\langle f, g \rangle = \int_0^1 f(t)g(t)dt, \qquad \|f\|_2^2 = \langle f, f \rangle.$$

   Introduces "geometry" as in $\mathbb{R}^d$, and the notion of *orthogonality*.

2. *Separability* in this setting is equivalent to the existence of a *complete orthonormal system* (CONS) of basis functions ϕ_i, $i \in \{1, 2, ...\}$ with 1) $\langle \phi_i, \phi_j \rangle = \mathbb{1}\{i = j\}$, i.e. the functions are orthogonal and have norm one, and 2) for any $f \in L^2[0, 1]$,

$$f(t) = \sum_{j=1}^{\infty} \langle f, \phi_j \rangle \phi_j(t).$$

These taken together mean that $L^2[0,1]$ is "close" in a sense to $\mathbb{R}^d$. If for instance we were ever able to make the simplification

$$f(t) = \sum_{j=1}^{\infty} \langle f, \phi_j \rangle \phi_j(t) \approx \sum_{j=1}^{d} \langle f, \phi_j \rangle \phi_j(t),$$

$(\langle f, \phi_1 \rangle, ..., \langle f, \phi_d \rangle)^\top \in \mathbb{R}^d$, and we would be in more familiar territory.

Takeaway: Dimension reduction is critical!

Remark: All separable Hilbert spaces are "isometrically isomorphic". I.e. if $\mathcal{H}_A, \mathcal{H}_B$ are separable Hilbert spaces with norms $\|\cdot\|_A$, $\|\cdot\|_B$, there exists a bijection $z : \mathcal{H}_A \mapsto H_B$ so that for all $f, g \in \mathcal{H}_A$, $\|f - g\|_A = \|z(f) - z(g)\|_B$. As a result all separable Hilbert spaces are in a sense equivalent to $L^2[0,1]$. Other examples of separable Hilbert spaces:

1. Sobolev space $W^{1,2}$
2. ℓ^2
3. $L^2[0,1]^{\otimes p}$

Remark: Analyzing FTS when we think of the data as being elements of a more general Banach space, e.g. $C[0,1]$, poses some significant challenges from a theoretical perspective. Many problems remain open in this area.

In most cases in which FTS analysis is applied (but not all!), the observed data are discrete measurements of an underlying process. Such raw data can often be represented as

$$X_i(t_{i,j}), \quad i \in \{1, ..., T\}, \quad j \in \{1, ..., n_i\}, \quad 0 \leq t_{i,1} < \cdots < t_{i,n_i} \leq 1.$$

We often begin the analysis by completing the data to a full curve using smoothing and interpolation techniques.

Linear interpolation: Define a function $\hat{X}_i : [0,1] \mapsto \mathbb{R}$ so that for any t such that $t_{i,j} \leq t \leq t_{i,j+1}$,

$$\hat{X}_i(t) = \frac{X_i(t_{i,j+1}) - X_i(t_{i,j})}{t_{i,j+1} - t_{i,j}}[t - t_{i,j}] + X_i(t_{i,j}).$$

Basis representations/Smoothing:

$$\hat{X}_i(t) = \sum_{j=1}^{K} c_i \phi_j(t) = \mathbf{c}^\top \mathbf{\Phi}(t),$$

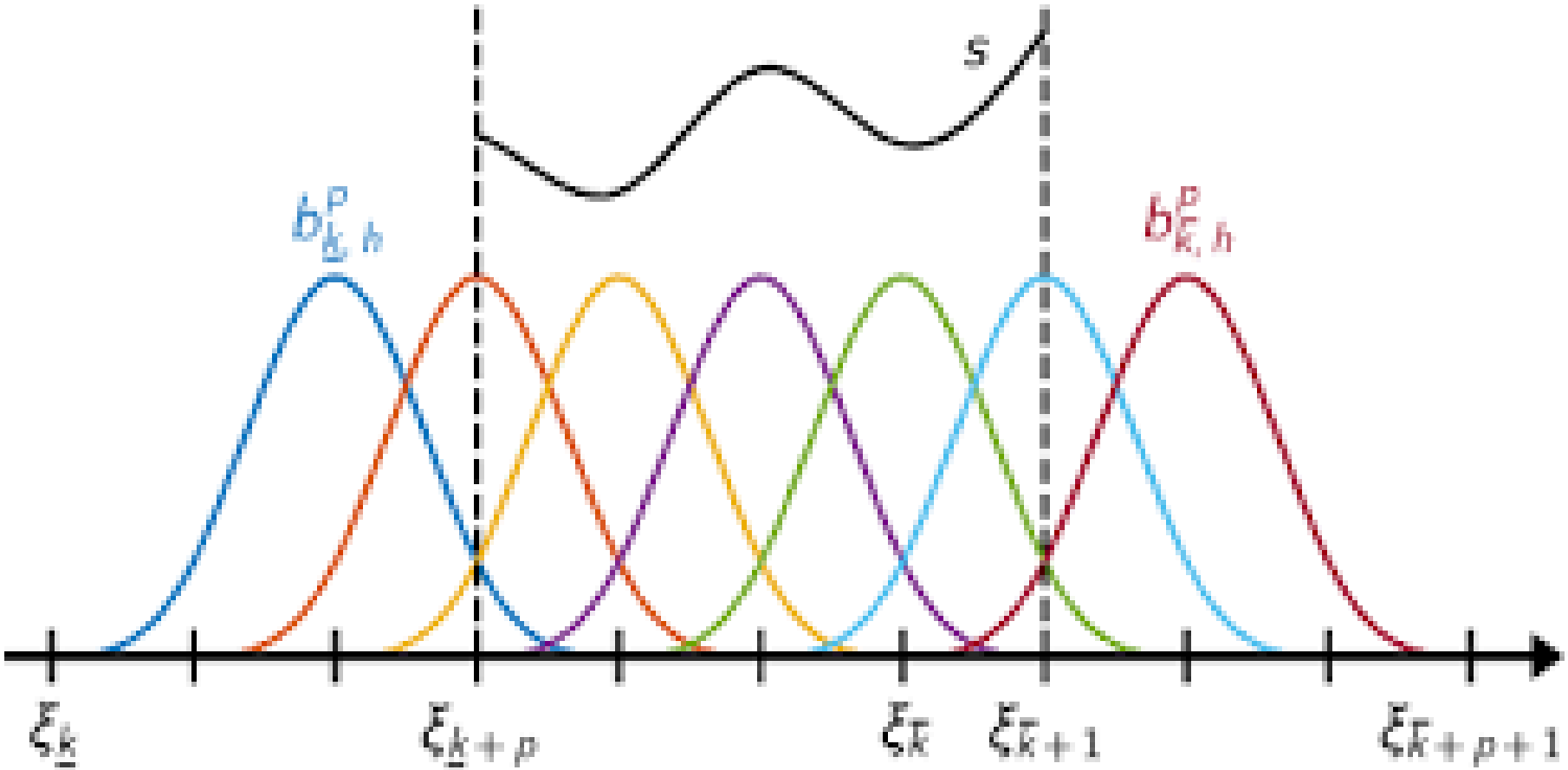
- ϕ_j is a CONS
- K is a tuning parameter
- $\mathbf{c} = (c_1, ..., c_K)^\top$
- $\mathbf{\Phi}(t) = (\phi_1(t), ..., \phi_K(t))^\top$.

Standard basis choices are:

- Fourier basis

$$\phi_j(t) = \sqrt{2}\cos(2\pi jt) + \sqrt{2}\sin(2\pi jt),$$

- Orthogonal b-spline basis



A plot of cubic B-spline polynomials based on equally spaced knots.

After choosing the basis, we often wish to estimate $\mathbf{c}$ in order to minimize

$$\text{SSE}_{i,K}(\mathbf{c}) = \sum_{j=1}^{n_i} [X_i(t_{i,j}) - \mathbf{c}^\top \mathbf{\Phi}(t_{i,j})]^2.$$

Generally if $K \geq n_i$ this can be minimized at $\text{SSE}_{i,K}(\mathbf{c})$ so the data are interpolated. This can might be OK, but can also lead to "overfitting".

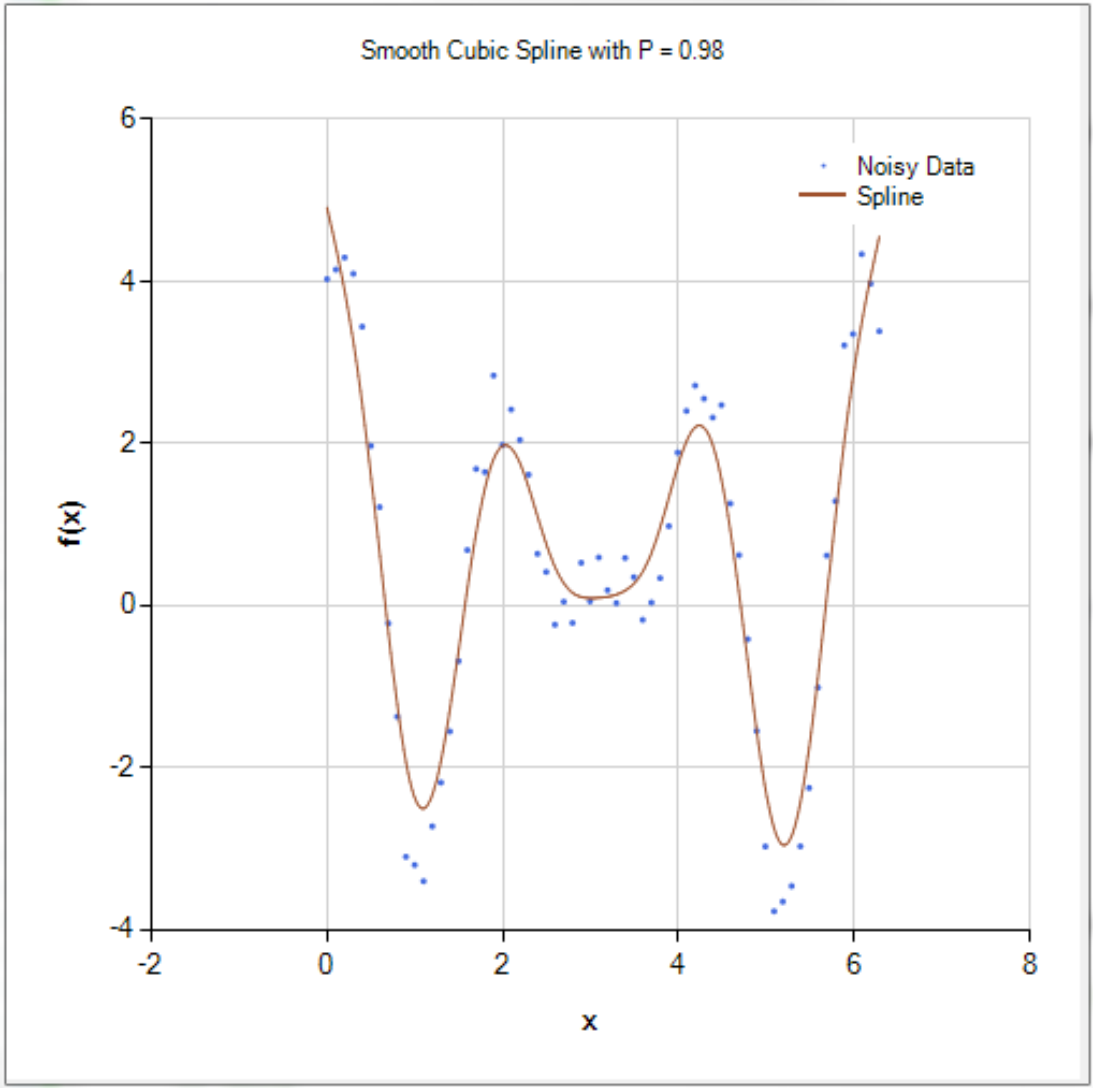
To prevent overfitting we sometimes penalize the "complexity" or "smoothness" of the curve using e.g.

$$\text{PEN}_{i,K}(\mathbf{c}) = \int_0^1 \left[\frac{d^2}{dt^2}\mathbf{c}^\top \mathbf{\Phi}(t)\right]^2 dt,$$

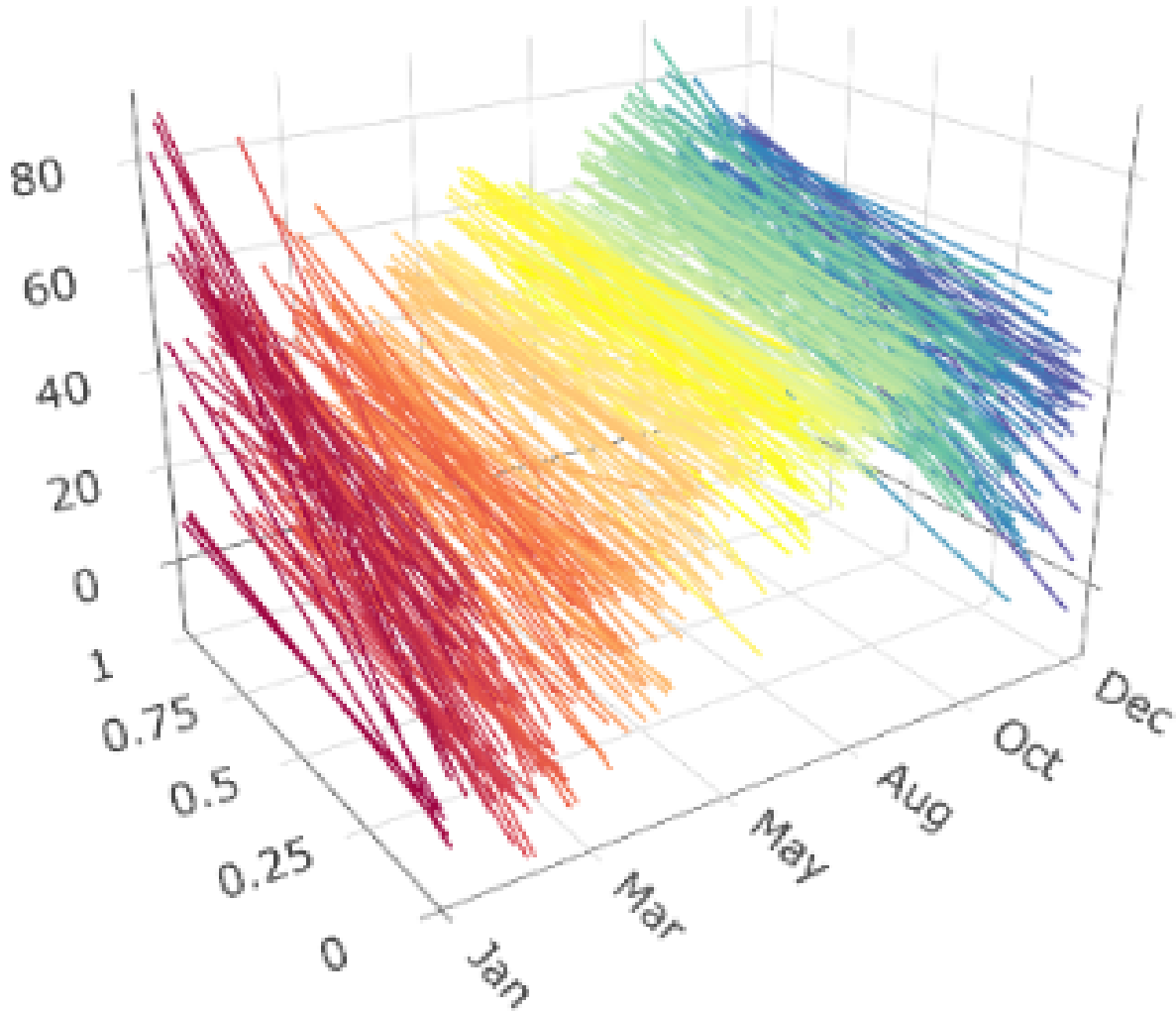
and then choose $\mathbf{c}$ to minimize

$$\text{SSE}_{i,K}(\mathbf{c}) + \lambda \text{PEN}_{i,K}(\mathbf{c}),$$

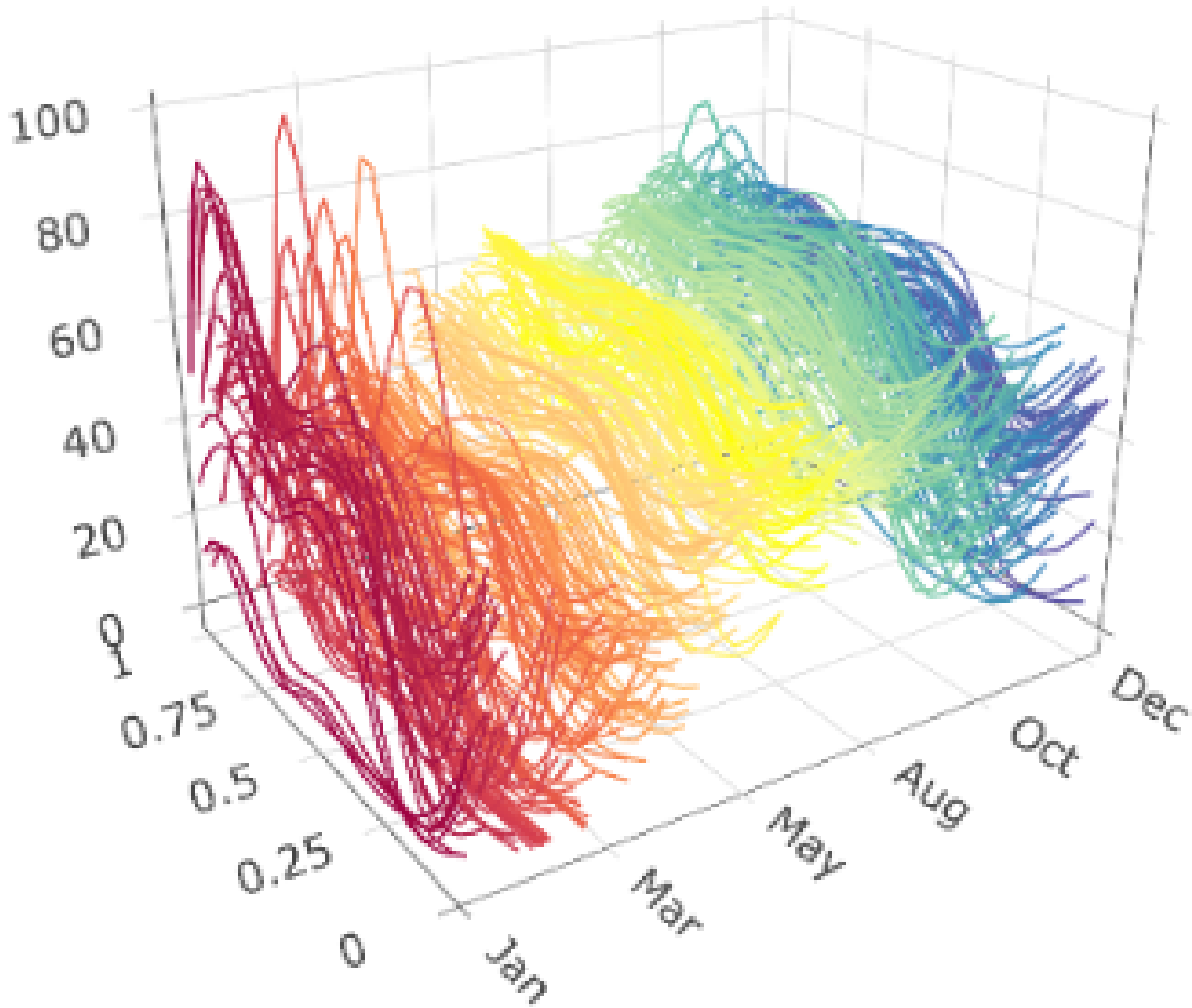
Larger λ leads to more smooth curves, whereas smaller λ leads to a curve that more closely interpolates the data.



Example of a functional data object created using penalized cubic splines.



Electricity data smoothed with b-splines and $\lambda = 1$.



Electricity data smoothed with b-splines and $\lambda = 0.0001$.

STOP FOR EXAMPLES

When beginning to analyze an FTS, we often consider computing some summary values and performing dimension reduction. Many of these tasks are carried out by thinking of the FTS as being *stationary.*

Definition

An FTS $\{X_i \in \mathcal{H},\ i \in \mathbb{Z}\}$ is (strictly) stationary if for each $p \in \mathbb{N}$, and $i_1, ..., i_p, h \in \mathbb{Z}$, and all (Borel) subsets $B_1, ..., B_p \subset \mathcal{H}$,

$$P(X_{i_1} \in B_1, ..., X_{i_p} \in B_p) = P(X_{i_1+h} \in B_1, ..., X_{i_p+h} \in B_p).$$

"The distribution of the data is the same no matter where we look"

For a stationary FTS X_i the *mean function* of the series is

$$E[X_i(t)] = \mu(t).$$

We typically estimate the mean function using the sample mean function

$$\bar{X}_T(t) = \frac{1}{T}\sum_{i=1}^{T} X_i(t).$$

The *covariance operator* of a stationary FTS is defined by

$$C(x)(t) := E[\langle X_i - \mu, x\rangle (X_i - \mu)(t)], \quad x \in L^2[0,1], \tag{1}$$

and is well defined so long as $E\|X_i\|_2^2 < \infty$. The covariance operator is also characterized by the *covariance kernel*

$$c(t,s) = E[(X_i(t) - \mu(t))(X_i(s) - \mu_i(s))],$$

via the relation

$$C(x)(t) = \int_0^1 c(t,s)x(s)ds.$$

These quantities are most commonly estimated using the empirical covariance kernel

$$\widehat{c}(t,s) = \frac{1}{T}\sum_{i=1}^{T}[X_i(t) - \bar{X}_T(t)][X_i(s) - \bar{X}_T(s)],$$

and the empirical covariance operator

$$\widehat{C}(x)(t) = \int_0^1 \hat{c}(t,s)x(s)ds.$$

The eigenfunctions of C, denoted by $v_j, j \geq 1$, are often called the *functional principal components* (FPCs), i.e.

$$C(v_j) = \lambda_j v_j.$$

These functions are orthogonal and define a CONS of $L^2[0,1]$. The *Karhunen–Loéve (KL) expansion*. It takes the form

$$X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_{ij} v_j(t), \quad \xi_{ij} = \langle X_i - \mu, v_j \rangle. \tag{2}$$

The *scores* ξ_{ij} satisfy $E\xi_{ij} = 0$, $E\xi_{ij}^2 = \lambda_j$, $E[\xi_{ij}\xi_{ij'}] = 0$ for $j \neq j'$.

This KL expansion is "optimal" in the following sense: if we consider the truncated expansion

$$X_i^{(d)}(t) = \mu(t) + \sum_{j=1}^{d} \xi_{ij} v_j(t)$$

this "finite dimensional" representation of X_i is optimal in that it minimizes

$$E\|X_i - X_i^{(d)}\|_2^2$$

among all possible choices of the functions $v_1, ..., v_d$.

The FPCs v_j and the eigenvalues λ_j are estimated by $\hat{v}_j$ and $\hat{\lambda}_j$ defined as the solutions to the equations

$$\widehat{C}(\hat{v}_j)(t) = \hat{\lambda}_j \hat{v}_j(t), \quad j \in \{1, ..., T\}.$$

Each curve X_i can then be approximated by a linear combination of a finite set of the estimated FPCs $\hat{v}_j$, i.e.

$$X_i(t) \approx \bar{X}_T + \sum_{j=1}^{d} \hat{\xi}_{ij} \hat{v}_j(t),$$

The vector of the sample scores,

$$[\hat{\xi}_{i1}, \ldots, \hat{\xi}_{ip}]^\top,$$

encodes the shape of X_i to a good approximation.

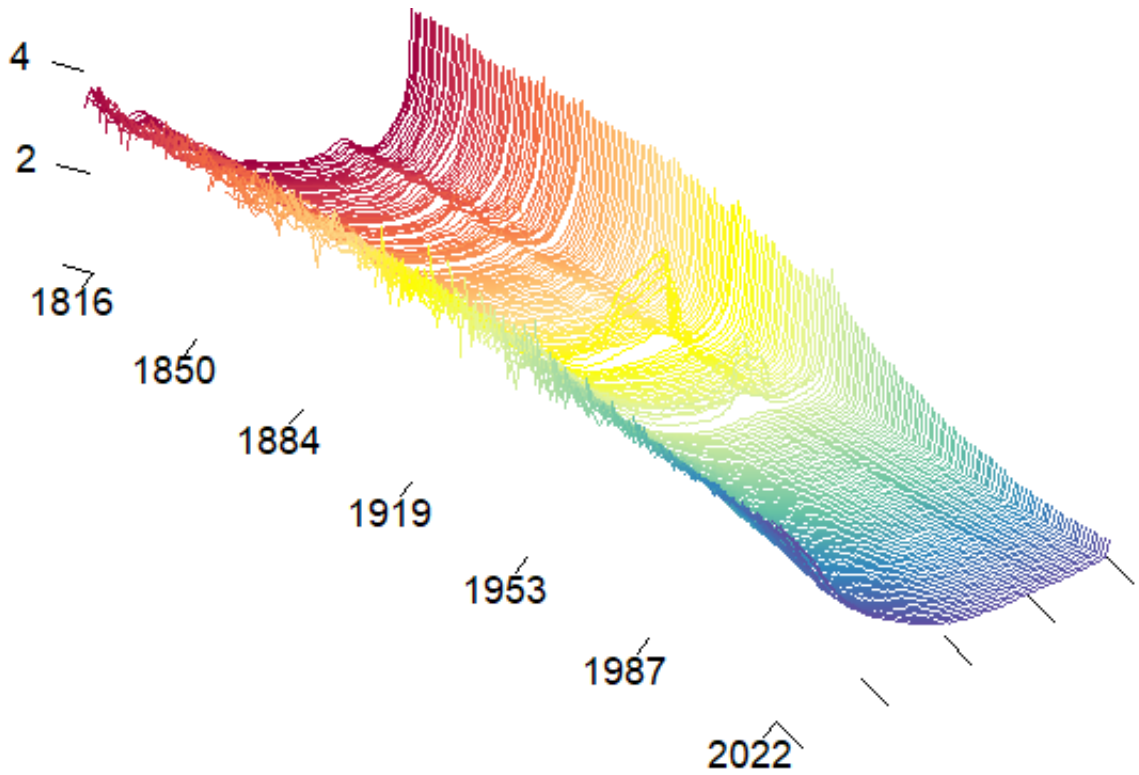
When using this dimension reduction technique is to select d. This is most commonly done using the *Total Variation Explained (TVE)* approach: we choose d so that the percentage

$$\text{TVE}_d = \frac{\hat{\lambda}_1 + \cdots + \hat{\lambda}_d}{\int_0^1 \widehat{c}(t,t)dt} = \frac{\hat{\lambda}_1 + \cdots + \hat{\lambda}_d}{\sum_{i=1}^{\infty} \hat{\lambda}_i}$$

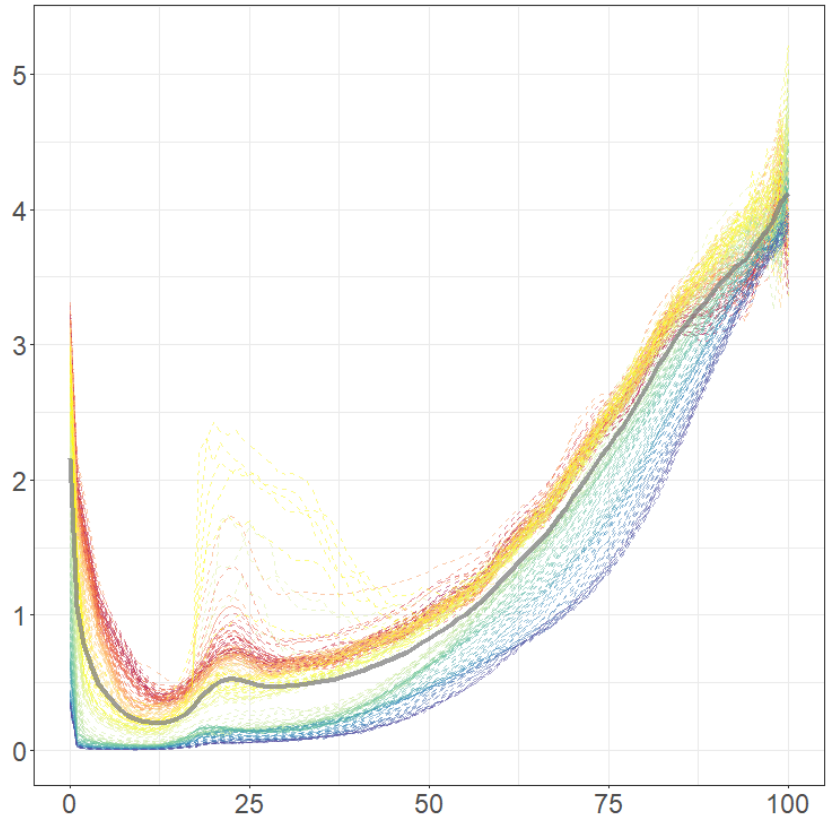
is sufficiently large. For example a common criterion is to choose

$$d = \inf\left\{p \ : \text{TVE}_p \geq 0.95\right\}.$$

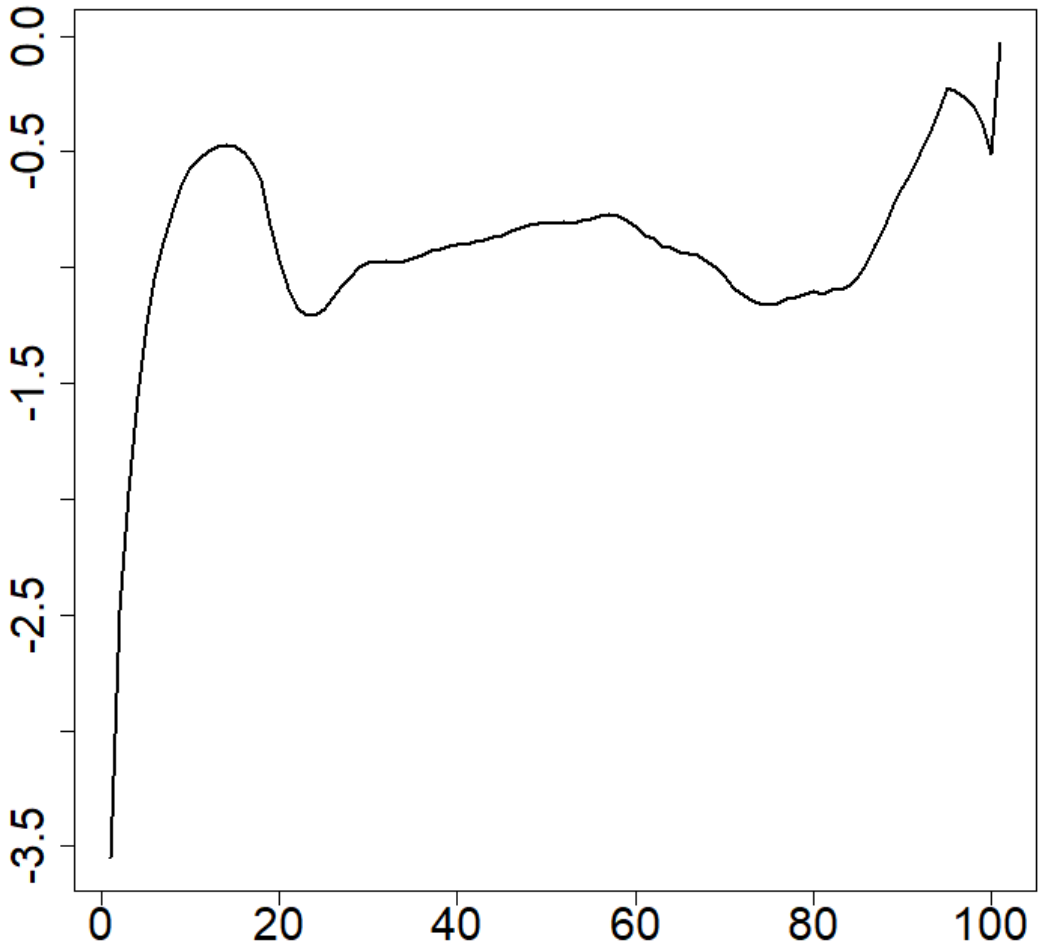
In many examples this leads to a reasonably small value of d. Tuning the threshold for selecting d and analyzing the effects this has on any conclusion drawn from the data is an important consideration.



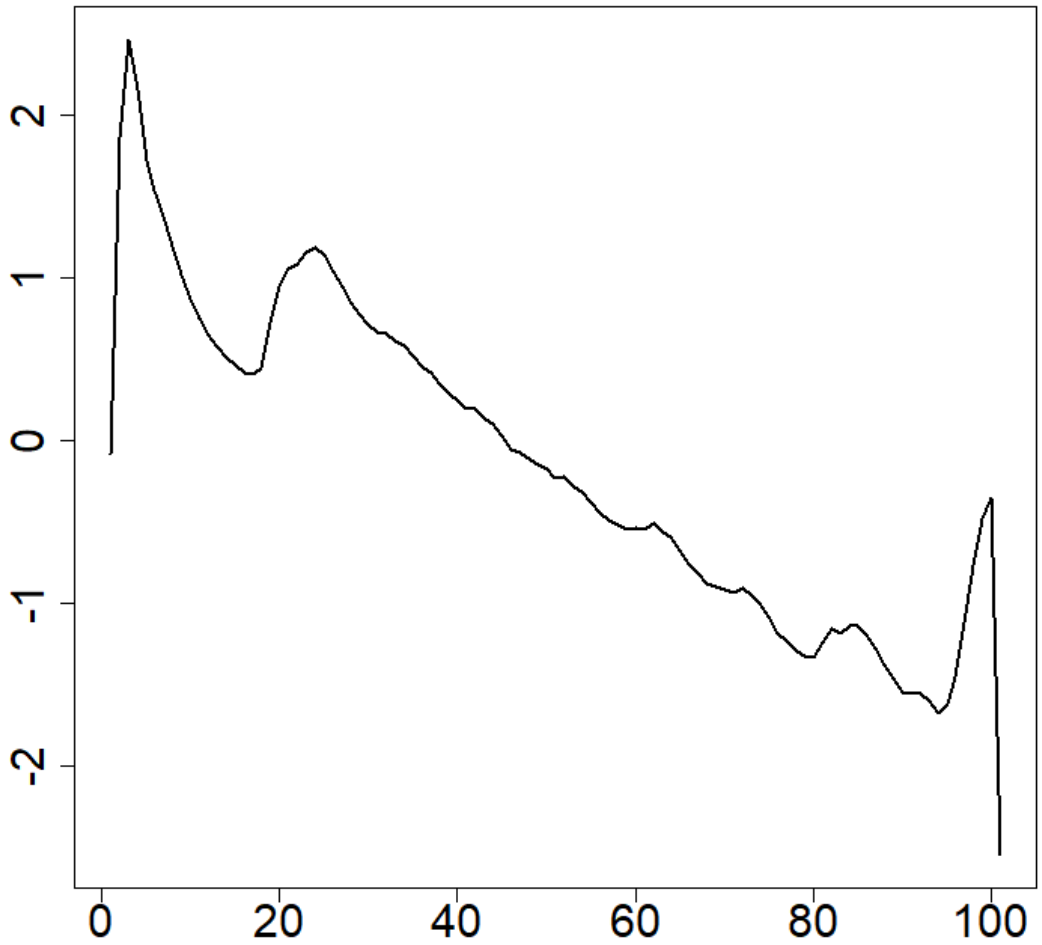
Plot of FTS of French Log mortality-rate data.



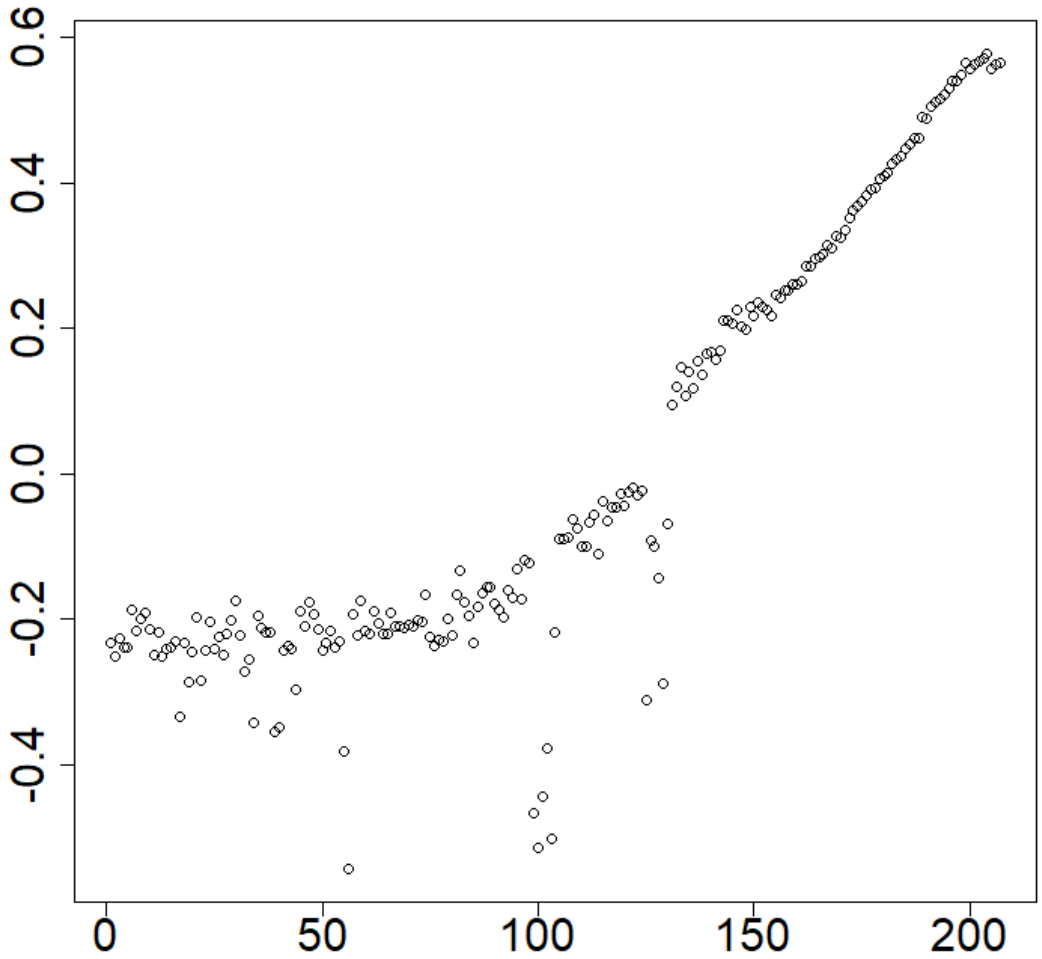
Plot of FTS French mortality data with the sample mean function of the series $\bar{X}_T(t)$.



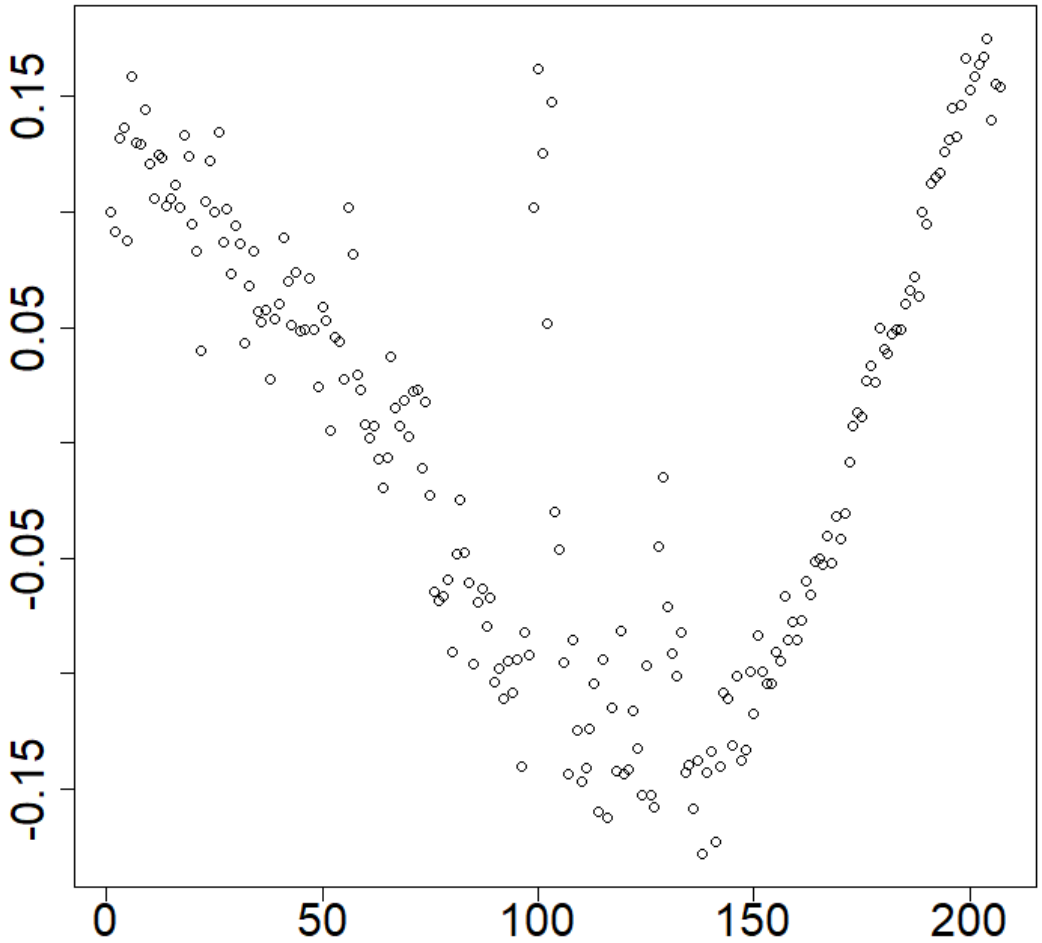
Plot of first FPC $\hat{v}_1(t)$.



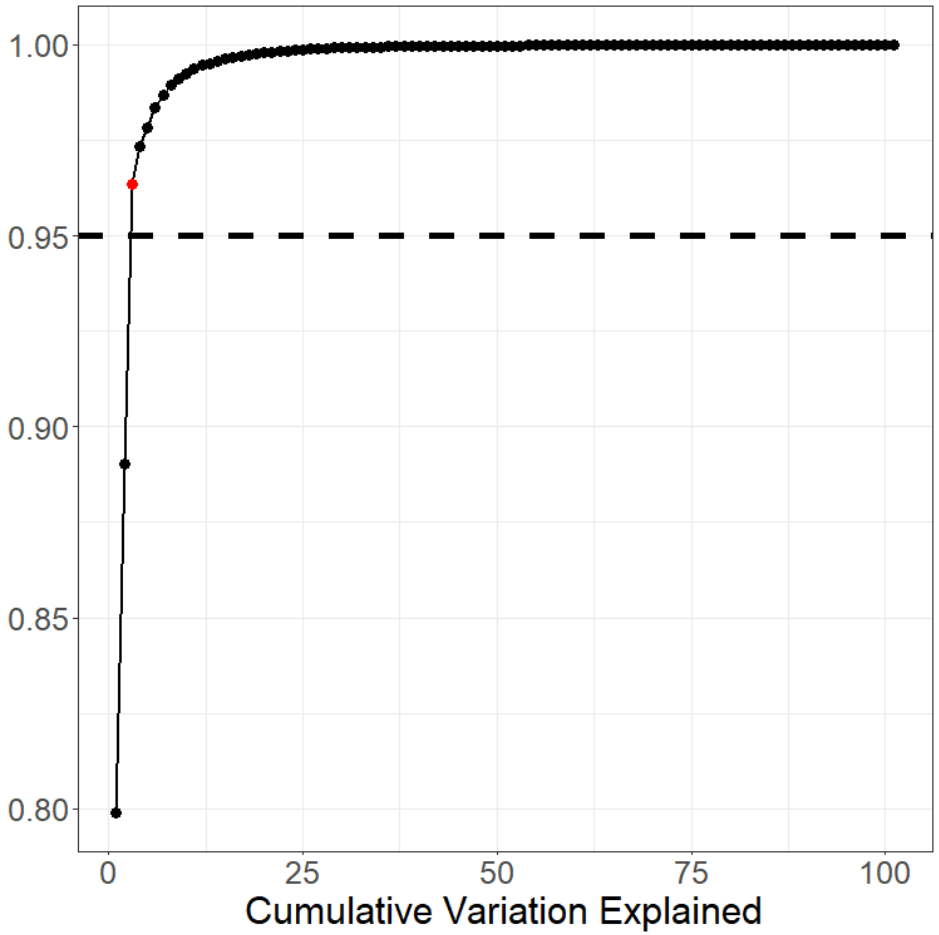
Plot of second FPC $\hat{v}_2(t)$.



Plot of the scores from PC 1.



Plot of the scores from PC 2.



Scree plot of TVE_d versus d for PCA with the FTS French mortality data.

STOP FOR EXERCISES

Forecasting FTS:

Suppose that we are given a functional time series (FTS)

$$\{X_i(t),\ i \in \{1, ..., T\},\ t \in [0, 1]\}$$

Our goal is to predict the curve X_{T+h}, as well as quantify the uncertainty in such a forecast.

Although there are many methods to forecast FTS, we will focus on what is sometimes called the "Hyndman-Ullah" method. We start with the PCA approximation

$$X_i(t) \approx X_i^{(J)}(t) = \hat{\mu}(t) + \sum_{j=1}^{J} \xi_{i,j} \hat{v}_j(t).$$

The time dynamics of the series are characterized by the J scalar time series $\xi_{i,1}, \ldots, \xi_{i,J}$.

Procedure:

1. Forecast the scalar time series $\xi_{T+h,1}, ..., \xi_{T+h,J}$, which we denote $\hat{\xi}_{T+h,1}, ..., \hat{\xi}_{T+h,J}$.
2. Forecast of X_{T+h}

$$\hat{X}_{T+h}(t) = \hat{\mu}(t) + \sum_{j=1}^{J} \hat{\xi}_{T+h,j} \hat{v}_j(t).$$

3. In order to quantify the uncertainty in the forecasts we simulate potential future values $\xi^{(b)}_{T+h,1}, ..., \xi^{(b)}_{T+h,J}$, for $b \in \{1, ..., B\}$. These are used to simulate the FTS as

$$\hat{X}^{(b)}_{T+h}(t) = \hat{\mu}(t) + \sum_{j=1}^{J} \xi^{(b)}_{T+h,j} \hat{v}_j(t), \quad b \in \{1, ..., B\}.$$

A pointwise in t 95% prediction interval for $X_{T+h}(t)$ is constructed as

$$X_{T+h}(t) \in (\hat{q}(0.025), \hat{q}(0.975)),$$

where $\hat{q}(\alpha)$ is the α sample quantile of $\hat{X}^{(b)}_{T+h}(t)$, $b \in \{1, ..., B\}$.

Standard methods to forecast the component series are:

1. SARIMA models: `autoarima`
2. Exponential smoothing models: `ets`

Two important examples are the "strong white noise model",

$$x_t = \mu + w_t \quad \text{strong white noise model}$$

$$x_t = \mu_0 + t\mu_1 + \sum_{j=1}^{t} w_j \quad \text{Random walk with drift model}$$

In the case of the first model the best predictor of x_n given past values of the series is the mean value μ, which can be estimated from $x_1, \ldots, x_{n-1}$ by the sample mean $\bar{x}$.

Let the *backshift operator* be defined by

$$B^j x_t = x_{t-j}, \quad j \geq 0.$$

The random walk model can be rewritten as

$$x_t = \mu_1 + x_{t-1} + w_t \iff \nabla x_t = x_t - x_{t-1} = (1 - B)x_t = \mu_1 + w_t$$

which makes clear that the best forecast for x_n based on $x_{n-1}, ..., x_1$ is $x_{n-1} + \mu_1$. This model is called an "integrated model".

ARIMA models in essence choose between these two models while also allowing for additional serial correlation in the series.

Given a white noise sequence $\{w_t\ ,\ t \in \mathbb{Z}\}$, a moving average process of order q (MA(q)) is

$$x_t = w_t + \theta_1 B w_t + \cdots + \theta_q B^q w_t = \theta(B) w_t,$$

$$\theta(x) = 1 + \theta_1 x + \cdots + \theta_q x^q.$$

An *AutoRegressive process of order p*, denoted $AR(p)$, is defined by

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t \iff \phi(B) x_t = w_t$$

$$\phi(x) = 1 - \phi_1 x - \cdots - \phi_p x^p,$$

An ARMA (autoregressive moving average) model is of the form

$$\phi(B)x_t = \theta(B)w_t.$$

Definition

x_t is said to follow an SARIMA (Seasonal Autoregressive Integrated Moving Average) model of orders p, d, q, P,D,Q and seasonal period s if

$$\Phi_P(B^s)\phi(B)(1-B^s)^D(1-B)^d x_t = \Theta(B^s)\theta(B)w_t.$$

This is abbreviated $x_t \sim SARIMA(p,d,q) \times (P,D,Q)_s$.

Notes on SARIMA models and `auto.arima`:

1. The seasonal period s is typically supplied by the practitioner. This is typically achieved in `R` by specifying the `frequency` parameter of at time series object.
2. `auto.arima` first selects the differencing degrees D and d via applying a KPSS stationarity tests to the time series x_t and ∇x_t. The ARMA orders p, q, P, and Q are selected using the AIC. Model is fit using MLE.
3. By simulating the error process w_t and iterating the SARIMA recursion we can simulate approximate future values of the series $x_{T+h}^{(b)}$.

Exponential smoothing was introduced in the late 1950's. The basic principle behind it is that for a time series $x_1, ..., x_n$, two extreme forecasts are:

$$\hat{x}_{n+1} = x_n \quad \longrightarrow \quad \text{random walk prediction}$$

$$\hat{x}_{n+1} = \bar{x} = \sum_{j=1}^{n} \frac{1}{n} x_j \quad \longrightarrow \quad \text{iid noise prediction.}$$

We notice that both of these forecasts are weighted linear combinations of the past values of the series $x_1, ..., x_n$. For general time series the optimal prediction might fall between these extremes.

Simple Exponential Smoothing: We forecast x_n with

$$\hat{x}_{n+1} = \alpha x_n + \alpha(1-\alpha)x_{n-1} + \alpha(1-\alpha)^2 x_{n-2} + \cdots$$

$0 \leq \alpha \leq 1$. This prediction may be restated as:

$$\begin{aligned} \text{Forecast Equation} \quad & \hat{x}_{n+1} = \ell_n \\ \text{Smoothing Equation} \quad & \ell_n = \alpha x_n + (1-\alpha)\ell_{n-1} = \ell_n(\alpha, \ell_0) \\ \text{Initial Condition} \quad & \ell_0 \end{aligned}$$

Here α and ℓ_0 are the scalar parameters defining this prediction, and can be estimated via least squares:

$$(\hat{\alpha}, \hat{\ell}_0) = \arg \min_{0 \leq a \leq 1, l_0} \sum_{i=2}^{n} (x_i - \ell_i(a, l_0))^2.$$

Linear Trend Exponential Smoothing: In order to make a forecast m steps ahead, we extrapolate the trend linearly as follows:

$$\begin{aligned}
\text{Forecast equation} \quad & \hat{x}_{n+m} = \ell_n + m b_n \\
\text{Level equation} \quad & \ell_n = \alpha x_n + (1-\alpha)(\ell_{n-1} + b_{n-1}) \\
\text{Trend/Slope equation} \quad & b_n = \beta(\ell_n - \ell_{n-1}) + (1-\beta) b_{n-1},
\end{aligned}$$

Scalar Parameters: $\alpha, \beta, \ell_0, b_0$ can once again be estimated using least squares.

Trend+Seasonal Exponential Smoothing (Holt-Winters ES): If a time series exhibits seasonality at period p, then we incorporate it into the forecast as follows. Letting $k = \lfloor (m-1)/p \rfloor$,

$$\begin{aligned}
\text{Forecast equation} \quad & \hat{x}_{n+m} = \ell_n + m b_n + s_{n+m-p(k+1)} \\
\text{Level equation} \quad & \ell_n = \alpha(x_n - s_{n-p}) + (1-\alpha)(\ell_{n-1} + b_{n-1}) \\
\text{Trend equation} \quad & b_n = \beta(\ell_n - \ell_{n-1}) + (1-\beta) b_{n-1} \\
\text{Seasonal equation} \quad & s_n = \gamma(x_n - \ell_{n-1} - b_{n-1}) + (1-\gamma) s_{n-p},
\end{aligned}$$

Scalar Parameters: $\alpha, \beta, \gamma, \ell_0, b_0, s_0, ..., s_{-p+1}$ can be estimated using least squares.

We note that each of these models may be rewritten in what is called "state-space" or "innovations" form. For example, the simple exponential smoothing model may be rewritten as:

$$\begin{aligned} \text{Observation equation} \quad & x_t = \ell_{t-1} + \varepsilon_t \\ \text{State equation} \quad & \ell_t = \ell_{t-1} + \alpha\varepsilon_t, \end{aligned}$$

where ε_t is an innovation sequence representing the residuals $x_t - \ell_{t-1}$. By assuming for instance that these residuals are Gaussian, one can once again easily simulate future values of the series $x_{T+h}^{(b)}$ as well as conduct model selection using AIC.

Notes on the `ets()` function in `R`

1. The `ets()` function from the `fpp2` package in `R` fits such exponential smoothing models using least squares and conducting model selection between standard, linear trend, and seasonal/Holt-Winters models using AIC.
2. We note this function also chooses among "multiplicative seasonality" versions of the same models. A multiplicative Holt-Winters model for a non-negative time series can be more appropriate if the seasonal/periodic fluctuations of the series increase (or decrease) as a function of the level of the series.

Show Examples

Time Series Cross-Validation:

A useful model selection tool is *time series cross-validation*. Suppose we wish to evaluate the quality of model g.

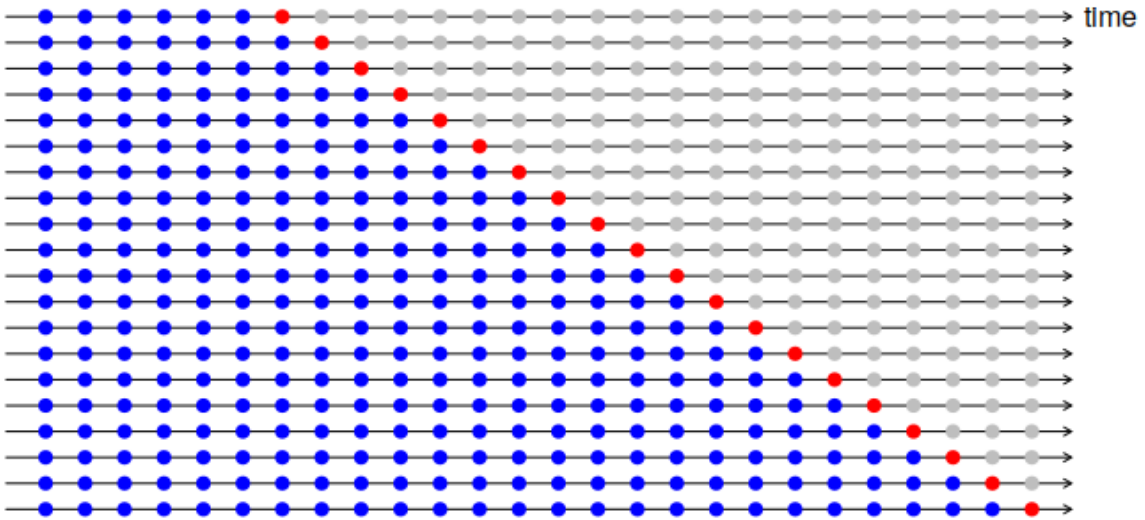
1. Select training, validation, and testing ranges, $1 \leq tr \leq v \leq test \leq T$, e.g. $(70\%, 15\%, 15\%)$. In forecasting we often forego testing.
2. For $j \in \{tr, ..., v\}$, forecast $\hat{X}_{j+1}$ based on the curves $X_1, ..., X_j$ using model g, or if we are interested in horizon h forecasting forecast $\hat{X}_{j+1}, ..., \hat{X}_{j+h}$. Calculate losses

$$L_j(g) = \int_0^1 [X_j(t) - \hat{X}_j(t)]^2 dt \text{ or } L_j(g) = \sum_{\ell=1}^{h} \int_0^1 [X_{j+\ell}(t) - \hat{X}_{j+\ell}(t)]^2 dt.$$

3. A CV loss score for the model g can be taken as

$$CV(g) = \frac{1}{(v - tr)} \sum_{j=tr+1}^{v} L_j(g).$$

Small values of $CV(g)$ suggest better performance.



'This type of time series cross-validation is often called "expanding window" cross-validation, since we forecast X_j at each stage using all of the previous data $X_1, ..., X_j$.

Show Examples

STOP FOR EXERCISES

One standard way to investigate the serial dependence of a time series is in terms of serial correlation. The *autocovariance operator at lag h*, $h \geq 0$, is defined by

$$\Gamma_h(x)(t) := \int \gamma_h(t,s)x(s)ds, \quad x \in L^2[0,1], \tag{3}$$

where $\gamma_h(t,s)$ is the *autocovariance kernel* defined by

$$\gamma_h(t,s) := E[(X_i(t) - \mu(t))(X_{i+h}(s) - \mu(s))], \quad t,s \in [0,1]. \tag{4}$$

Note: at lag zero, $\Gamma_0 = C$, where C is the covariance operator.

Given functional observations, $X_1, \ldots, X_T$, γ_h can be estimated using its sample counterpart

$$\hat{\gamma}_{T,h}(t,s) := \frac{1}{T}\sum_{i=1}^{T-h}(X_i(t) - \bar{X}_T(t))(X_{i+h}(s) - \bar{X}_T(s)), \quad 0 \leq h < T. \quad (5)$$

A simple graphical summary of the serial dependence in the series can be obtained by plotting the functional autocorrelation function (fACF)

$$\hat{\rho}_h = \frac{\|\hat{\gamma}_{T,h}\|_2}{\int \hat{\gamma}_{T,0}(t,t)dt}$$

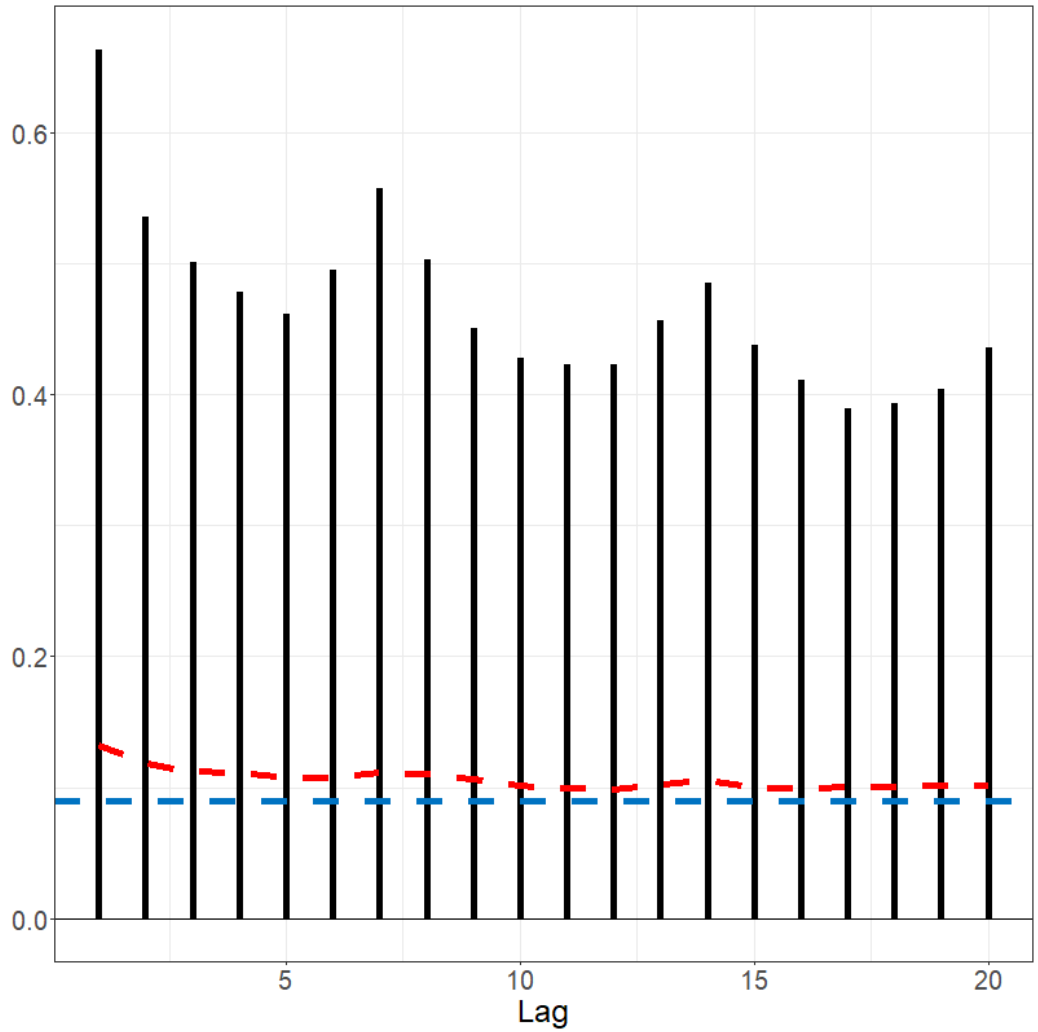
An upper prediction bound for this quantity assuming the data follow a white noise process is

$$\frac{\sqrt{Q_{IID,0.95}}}{\int \hat{\gamma}_{N,0}(t,t)dt} \quad \text{or} \quad \frac{\sqrt{Q_{h,0.95}}}{\int \hat{\gamma}_{N,0}(t,t)dt}.$$

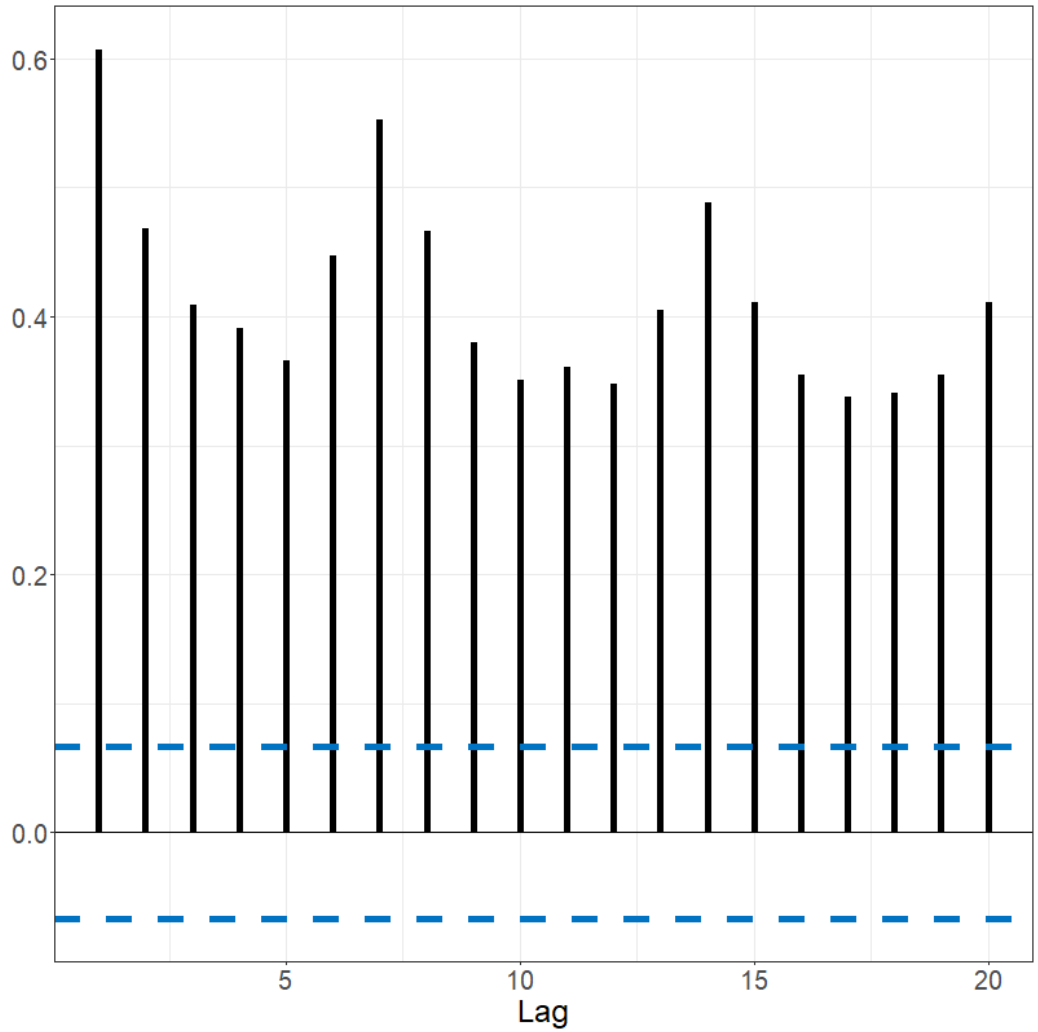
A robust graphical summary of the serial dependence in an FTS is the *functional spherical autocorrelation function* (fSACF)

$$\tilde{\rho}_h = \frac{1}{T}\sum_{i=1}^{T-h}\left\langle \frac{X_i - \tilde{\mu}}{\|X_i - \tilde{\mu}\|}, \frac{X_{i+h} - \tilde{\mu}}{\|X_{i+h} - \tilde{\mu}\|} \right\rangle, \qquad 0 \leq h < T,$$

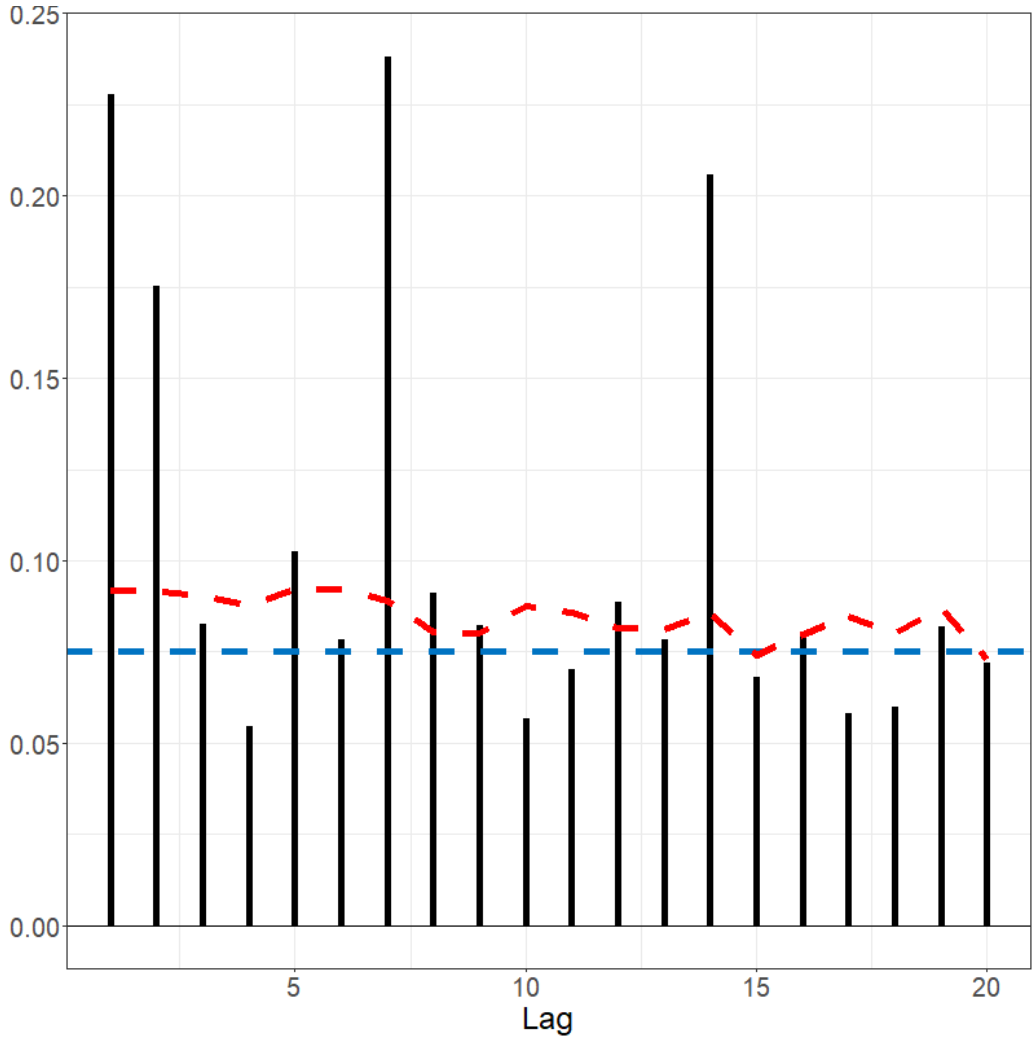
where $\tilde{\mu}$ is the estimated spatial median of the series.



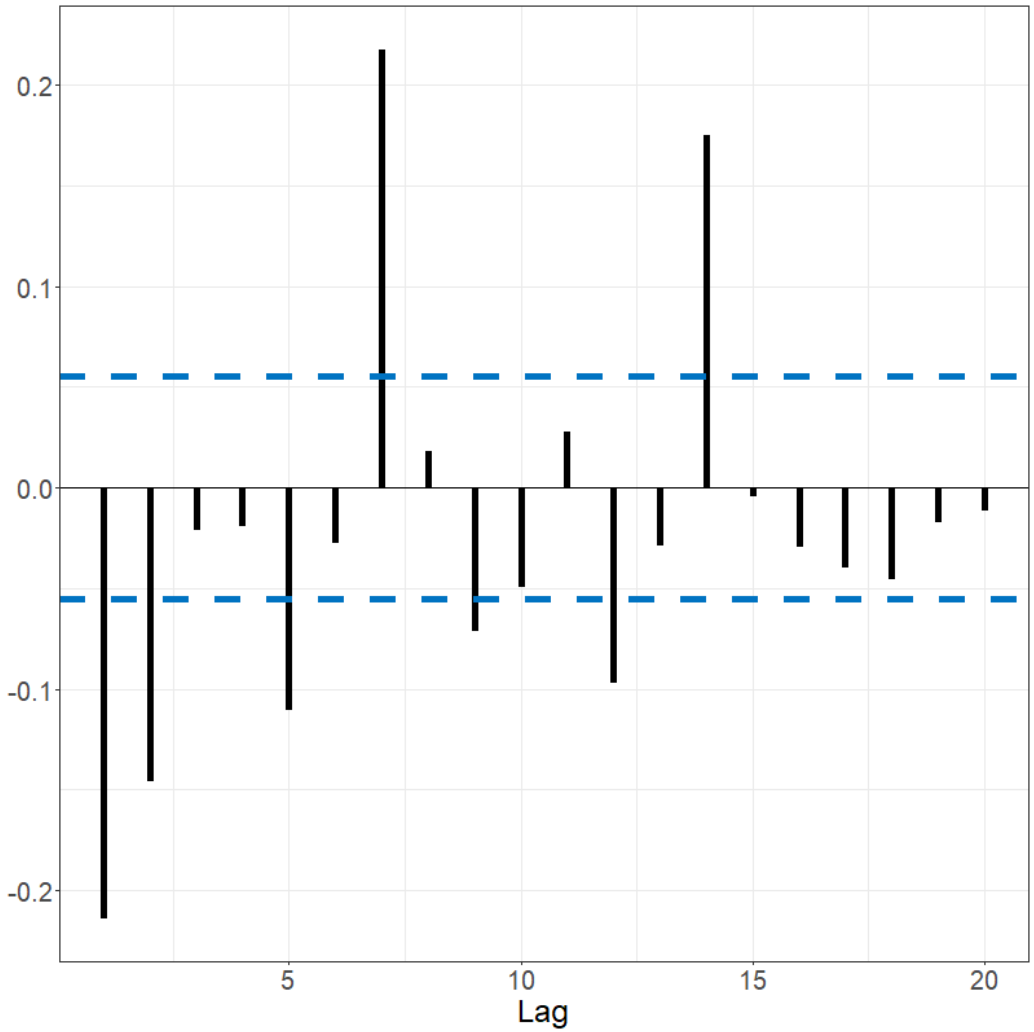
Autocorrelation in the electricity data as described by the fACF.



Autocorrelation in the electricity data as described by the fSACF.



Autocorrelation of the first differenced electricity data $X_i(t) - X_{i-1}(t)$ as described by the fACF.



Autocorrelation in the first differenced electricity data $X_i(t) - X_{i-1}(t)$ electricity data as described by the fSACF.

White noise tests:

$$X_i \text{ a white noise} \implies \hat{\rho}_h \approx 0 \text{ for } \textbf{all } h$$

.

$$\mathrm{KRS}_{T,H} = T \sum_{h=1}^{H} \|\hat{\gamma}_{T,h}\|^2,$$

A higher value of $\mathrm{KRS}_{T,H}$ suggests a potential departure of the observed series from a white noise process. $p-values$ are calculated assuming under the null that the series is a white noise.

Show Examples

Many procedures to analyze FTS assume at some level that the data are stationary, it is useful to have methods to evaluate the plausibility of that assumption.

$$S_T(x,t) = \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor Tx \rfloor} X_i(t)$$

CUSUM process:

$$Z_T(x,t) = S_T(x,t) - xS_T(1,t)\,.$$

When the data are non-stationary, the CUSUM process tends to fluctuate more with respect to the partial sample parameter x than when the data are stationary.

$$I_T = \int_0^1 \int_0^1 Z_T^2(x,t)dt\,dx\,,$$

and

$$M_T = \sup_{x\in[0,1]} \int_0^1 Z_T^2(x,t)dt.$$

$p-values$ of stationarity tests and change point detectors are based on these statistics and constructed under general stationarity conditions on the series.

Show Examples

STOP FOR EXERCISES