An introduction to time series models

SLS Meeting
November 19th, 2018

Matt Nunes, University of Bath

Time series analysis

Time series analysis simply refers to the analysis of data collected / indexed over time. Such data is observed in a wide range of scientific areas of interest, e.g. industrial process monitoring, climate modelling, official statistics.

In particular, our aim is to build **realistic models** of such data which account for possible complex **temporal dependencies.**

Time series analysis

Time series analysis simply refers to the analysis of data collected / indexed over time. Such data is observed in a wide range of scientific areas of interest, e.g. industrial process monitoring, climate modelling, official statistics.

In particular, our aim is to build **realistic models** of such data which account for possible complex **temporal dependencies**.

Analysis tasks after modelling include

- forecasting (prediction)
- classification / distinguishing series
- detection of changes, identifying patterns or periodicities etc.

Time series analysis

Time series analysis simply refers to the analysis of data collected / indexed over time. Such data is observed in a wide range of scientific areas of interest, e.g. industrial process monitoring, climate modelling, official statistics.

In particular, our aim is to build **realistic models** of such data which account for possible complex **temporal dependencies**.

Analysis tasks after modelling include

- forecasting (prediction)
- classification / distinguishing series
- detection of changes, identifying patterns or periodicities etc.

Notation:

- A (real-valued, stationary) time series will be denoted by $\{X_t\}_{t\in\mathbb{Z}}$, with a corresponding realisation of X_t being x_t.

Stationarity

In order to do inference, it is often assumed some sort of invariance of time series, i.e. the statistical characteristics of the series do not change over time (**stationarity**).

Types of stationarity:

- **First order**: The mean of the time series is the same over time
- **Strict stationarity**: For any finite sequence of integers $t_1, \ldots, t_k$ and shift h, the distribution of $\{X_{t_1}, \ldots, X_{t_k}\}$ is the same as $\{X_{t_1+h}, \ldots, X_{t_k+h}\}$.
  (considered strong assumption, hard to check in practice).

Stationarity

In order to do inference, it is often assumed some sort of invariance of time series, i.e. the statistical characteristics of the series do not change over time (**stationarity**).

Types of stationarity:

- **First order**: The mean of the time series is the same over time
- **Strict stationarity**: For any finite sequence of integers $t_1, \ldots, t_k$ and shift h, the distribution of $\{X_{t_1}, \ldots, X_{t_k}\}$ is the same as $\{X_{t_1+h}, \ldots, X_{t_k+h}\}$.
  (considered strong assumption, hard to check in practice).
- **Second order / covariance / weak stationarity**: If the mean is constant for all t and if for any t and h, $\gamma_X(h) = \text{cov}(X_t, X_{t+h})$ only depends on the lag difference k.

(Note: Strict stationarity & $\mathbb{E}(|X_t|^2) < \infty$ implies second order stationarity).

Stationarity

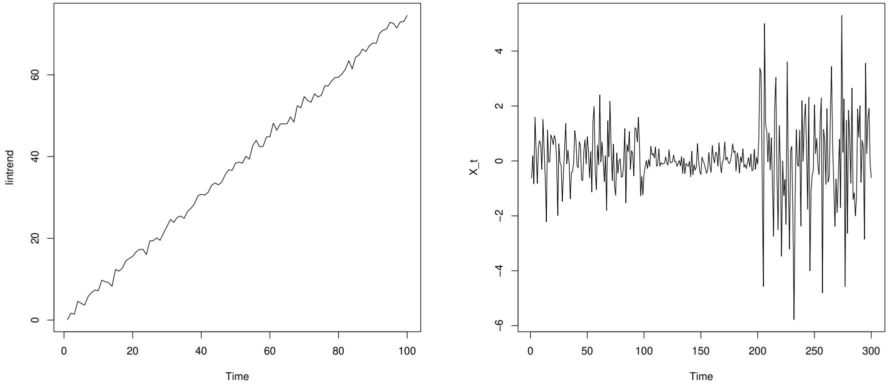


Figure: Types of (non)stationarity: linear trend (left); non-constant variance (right).

Some popular time series models: AR(p)

Motivation: Recall from linear regression, we predict a response Y given some covariates X_j, so we model Y_i as

$$Y_i = \sum_{j=1}^{p} a_j X_{ij} + \varepsilon_i,$$

with $\mathbb{E}(\varepsilon_i | X_{ij}) = 0$ and typically ε_i and X_{ij} independent.

For time series, we can similarly predict a future observation from the current and past observations

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t.$$

This is the **autoregressive model (of order p)**.

Some popular time series models: MA(q)

Let $\{X_t\}$ be a time series. We say X_t has a **moving average of order q** (MA(q) for short) representation if

$$X_t = \sum_{j=0}^{q} \psi_j \varepsilon_{t-j},$$

where $\{\varepsilon_t\}$ are IID random variables with zero mean and finite variance (i.e. white noise).

In other words, the series is modelled as a linear combination of the previous noise.

Some popular time series models: MA(q)

Let $\{X_t\}$ be a time series. We say X_t has a **moving average of order q** (MA(q) for short) representation if

$$X_t = \sum_{j=0}^{q} \psi_j \varepsilon_{t-j},$$

where $\{\varepsilon_t\}$ are IID random variables with zero mean and finite variance (i.e. white noise).

In other words, the series is modelled as a linear combination of the previous noise.

We can combine autoregressive and moving average models to form ARMA models.

Some (partial) justification for ARMA processes

Suppose we have an AR(1) process. We have

$$\begin{aligned} x_t &= \phi x_{t-1} + \varepsilon_t \\ &= \phi(\phi x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \vdots \\ &= \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}. \end{aligned}$$

In other words, an AR(1) process can be expressed as a linear combination of elements from the noise process ε_t.

Some (partial) justification for ARMA processes

Theorem

Any zero-mean nondeterministic covariance-stationary process x_t can be decomposed as

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \nu_t,$$

where ε_t is a finite variance white noise process, $\sum_j \psi_j^2 < \infty$ and ε_t is independent of ν_t for all $t \in \mathbb{Z}$.

Some (partial) justification for ARMA processes

Theorem

Any zero-mean nondeterministic covariance-stationary process x_t can be decomposed as

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \nu_t,$$

where ε_t is a finite variance white noise process, $\sum_j \psi_j^2 < \infty$ and ε_t is independent of ν_t for all $t \in \mathbb{Z}$.

- This implies that the dynamic of any purely nondeterministic covariance-stationary process can be arbitrarily well approximated by an ARMA process.
- The decomposition of a series by the Wold representation may not be the best description of the process.

ARMA processes: model selection

- Looking at the autocorrelation function (ACF) and partial autocorrelation function can give an idea about how to choose model AR and MA orders (look for where the plots “cut off”)

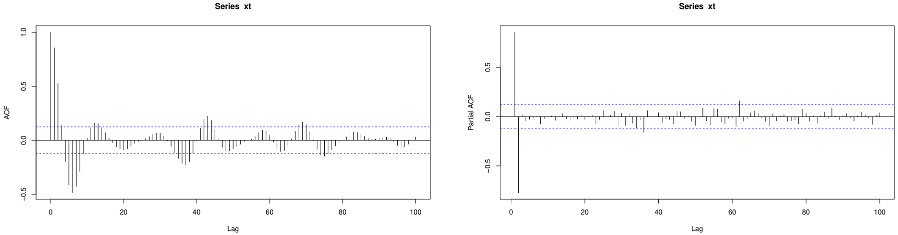


Figure: ACF and PACF of an AR(2) process; notice the characteristic “cut off” and damped exponential pattern of the plots.

- we can also use model selection procedures like the AIC.

Integrated models for (trend) nonstationarity

Now suppose $x_t = \mu_t + y_t$, with y_t a stationary process. For example, suppose the mean is a random walk, i.e. $\mu_t = \mu_{t-1} + \nu_t$, with ν_t stationary.
Then the differenced the series

$$\nabla x_t = x_t - x_{t-1} = \nu_t + \nabla y_t$$

is made up of stationary components and thus is stationary.

Integrated models for (trend) nonstationarity

Now suppose $x_t = \mu_t + y_t$, with y_t a stationary process. For example, suppose the mean is a random walk, i.e. $\mu_t = \mu_{t-1} + \nu_t$, with ν_t stationary.

Then the differenced the series

$$\nabla x_t = x_t - x_{t-1} = \nu_t + \nabla y_t$$

is made up of stationary components and thus is stationary.

More generally, if μ_t is a kth order polynomial, $\mu_t = \sum_{j=0}^{k} \beta_j t^j$, then $\nabla^k y_t$ is stationary.

Integrated models for (trend) nonstationarity

Now suppose $x_t = \mu_t + y_t$, with y_t a stationary process. For example, suppose the mean is a random walk, i.e. $\mu_t = \mu_{t-1} + \nu_t$, with ν_t stationary.
Then the differenced the series

$$\nabla x_t = x_t - x_{t-1} = \nu_t + \nabla y_t$$

is made up of stationary components and thus is stationary.

More generally, if μ_t is a kth order polynomial, $\mu_t = \sum_{j=0}^{k} \beta_j t^j$, then $\nabla^k y_t$ is stationary.

This leads to the **integrated ARMA model**: a process x_t is said to be ARIMA(p,d,q) if $\nabla^d x_t$ is ARMA(p,q).

Integrated models for (trend) nonstationarity

Example: Stationary process with a linear trend:

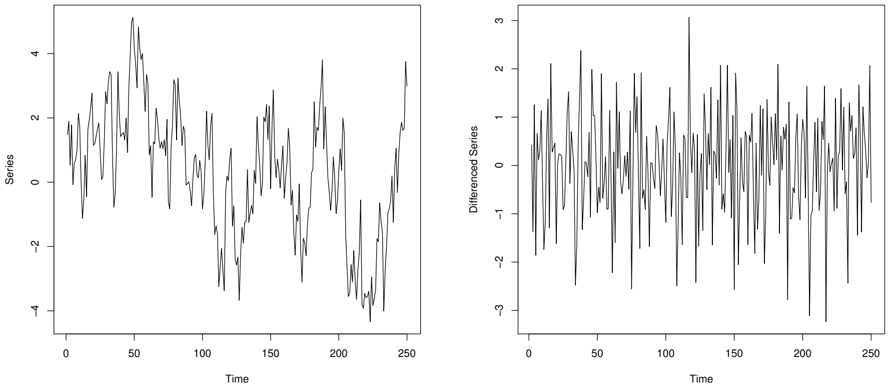


Figure: Effect of differencing: original series (left); differenced series (right).

Modelling seasonality

We can also extend the models we've seen to seasonal components, in a similar manner to integrated models.

Suppose a seasonal cycle lasts for s timepoints, i.e. the behaviour of the series is similar at a lag of s. Then if we difference the series **at lag s**,

$$y_t = \nabla_s x_t = x_t - x_{t-s},$$

this will remove the seasonality.

Modelling seasonality

We can also extend the models we've seen to seasonal components, in a similar manner to integrated models.

Suppose a seasonal cycle lasts for s timepoints, i.e. the behaviour of the series is similar at a lag of s. Then if we difference the series **at lag s**,

$$y_t = \nabla_s x_t = x_t - x_{t-s},$$

this will remove the seasonality.

More generally, we can have a ARIMA for the seasonal part of a time series $(P, D, Q)_s$, with the same interpretation of the orders.

Modelling seasonality

We can also extend the models we've seen to seasonal components, in a similar manner to integrated models.

Suppose a seasonal cycle lasts for s timepoints, i.e. the behaviour of the series is similar at a lag of s. Then if we difference the series **at lag s**,

$$y_t = \nabla_s x_t = x_t - x_{t-s},$$

this will remove the seasonality.

More generally, we can have a ARIMA for the seasonal part of a time series $(P, D, Q)_s$, with the same interpretation of the orders.

Putting this together a flexible model is the **SARIMA model**: $ARIMA(p, d, q) \times ARIMA(P, D, Q)_s$.

This allows for nonseasonal and seasonal components.

Modelling seasonality

Example: `co2` time series representing monthly CO_2.

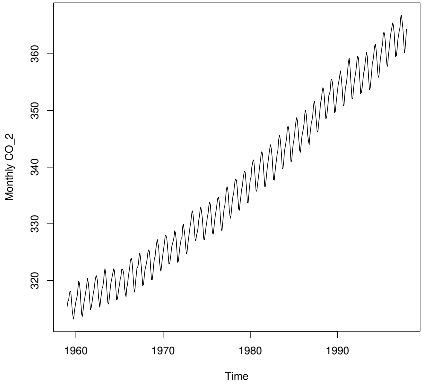


Figure: Original series, featuring trend and yearly seasonality.

Modelling seasonality

Example: Time series representing monthly CO_2: first difference (∇x_t), and seasonal difference $\nabla_{12}\nabla x_t$.

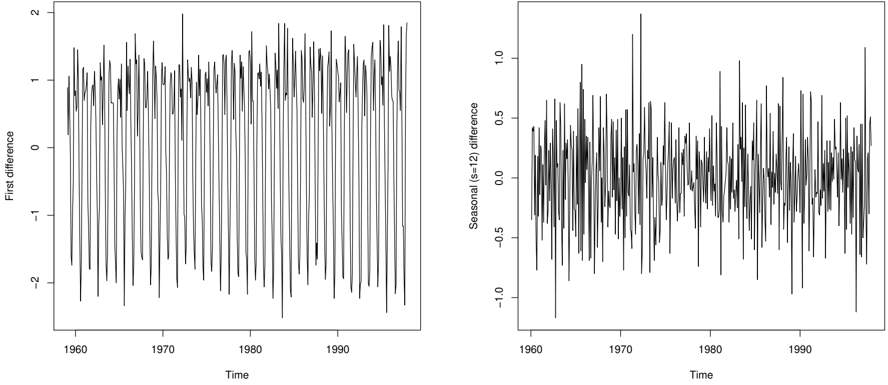


Figure: Effect of differencing: First difference (left); further (s=12) differenced series (right).

Model fitting (estimating coefficients)

- Model fitting in ARMA models is generally done using maximum likelihood estimation, subject to the constraints that the coefficients satisfy stationarity conditions.
- Usually, (due to the autocorrelation etc) we have to resort to numerical methods to maximise the likelihood or use a state space approach (Kalman filter).
- As usual, one can incorporate prior belief on the structure of the model and use a Bayesian formulation.

Spectral analysis: frequency domain representations

In many applications, time series will exhibit periodicities or oscillations, which may occur at differing rates.

These periodicities may be difficult to discern in the time domain.

Spectral / frequency domain analysis aims to capture these features, and provide extra insight and properties of data.

Spectral analysis: frequency domain representations

In many applications, time series will exhibit periodicities or oscillations, which may occur at differing rates.

These periodicities may be difficult to discern in the time domain.

Spectral / frequency domain analysis aims to capture these features, and provide extra insight and properties of data.

Main idea:

decompose a (stationary) series in terms of sinusoids at different frequencies ω_j with random, uncorrelated amplitudes.

Spectral analysis: frequency domain representations

Suppose $X_t = \sum_{j=1}^{k} A_j \sin(2\pi\omega_j t) + B_j \cos(2\pi\omega_j t)$, with A, B uncorrelated, mean zero, with variance σ_j^2 (mixture of sinusoids at different frequencies and amplitudes).

Then,

$$\gamma(h) = \sum_{j=1}^{k} \sigma_j^2 \cos(2\pi\omega_j h).$$

(This follows from the uncorrelatedness of A_j and B_j).

Spectral analysis: frequency domain representations

Suppose $X_t = \sum_{j=1}^{k} A_j \sin(2\pi\omega_j t) + B_j \cos(2\pi\omega_j t)$, with A, B uncorrelated, mean zero, with variance σ_j^2 (mixture of sinusoids at different frequencies and amplitudes).
Then,

$$\gamma(h) = \sum_{j=1}^{k} \sigma_j^2 \cos(2\pi\omega_j h).$$

(This follows from the uncorrelatedness of A_j and B_j). In particular, setting $h = 0$, we have

$$\text{var}(X_t) = \gamma(0) = \sum_{j=1}^{k} \sigma_j^2.$$

In other words, we can decompose the autocovariance / variance of the process via the sinusoidal components of the series X_t.

The spectral density

Definition

Let X_t be a stationary process. Then if the autocovariance is absolutely summable (i.e. $\sum_{h=-\infty}^{\infty} \gamma(h) < \infty$), then it has the representation[a]

$$\gamma(h) = 2\pi \int_0^{1/2} \cos(2\pi\omega h) f(\omega) d\,\omega \quad h \in \mathbb{N},$$

as the inverse transform of the **spectral density**

$$f(\omega) = 2\gamma(0) + 4 \sum_{h=1}^{\infty} \gamma(h) \cos(2\pi\omega h) \quad \text{for} \quad \omega \in (0, 1/2).$$

[a]Other definitions exist which differ by the range of the sum / integral and scaling factor 2π.

The spectral density: some comments

Some comments of the spectral density:

- Interpretation: A stationary time series can be (approximately) expressed as a random linear combination of sines and cosines at different frequencies).
- The spectral density is positive
- The spectral density contains **the same information** as the autocovariance, just expressed differently (cf. Parseval's theorem).
- The spectral density is even and periodic (hence we can restrict our attention to e.g. $\omega \in (0, 1/2)$).

The spectral density: examples

1. Since white noise is an uncorrelated process, then $\gamma(0) = \sigma^2$ and is zero for $h \neq 0$. Hence

$$f_{WN}(\omega) = 2\gamma(0) + 4\sum_{h=1}^{\infty} \gamma(h)\cos(2\pi\omega h) = 2\gamma(0) = 2\sigma^2,$$

i.e. the spectral density of white noise is constant for all frequencies.

The spectral density: examples

1. Since white noise is an uncorrelated process, then $\gamma(0) = \sigma^2$ and is zero for $h \neq 0$. Hence

$$f_{WN}(\omega) = 2\gamma(0) + 4\sum_{h=1}^{\infty} \gamma(h)\cos(2\pi\omega h) = 2\gamma(0) = 2\sigma^2,$$

   i.e. the spectral density of white noise is constant for all frequencies.

2. Let X_t be an AR(1) process with parameter ϕ. Then it can be shown that

$$f(\omega) = \frac{\sigma^2}{1 - 2\phi\cos(2\pi\omega) + \phi^2}.$$

   ($\phi > 0 \leftrightarrow$ low frequencies, $\phi < 0 \leftrightarrow$ high frequencies).

Example

Example: AR(1) with $\phi = 0.9$.

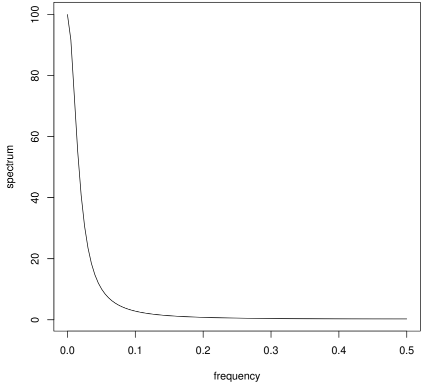


Figure: Theoretical spectrum of AR(1) process with $\phi = 0.9$.

Spectral estimation: the periodogram

Recall that the discrete Fourier transform is defined as

$$d(\omega_j) = T^{-1/2} \sum_{t=1}^{T} x_t e^{2\pi i \omega_j t},$$

for equally spaced *Fourier frequencies* ω_j.

Spectral estimation: the periodogram

Recall that the discrete Fourier transform is defined as

$$d(\omega_j) = T^{-1/2} \sum_{t=1}^{T} x_t \mathrm{e}^{2\pi \mathrm{i} \omega_j t},$$

for equally spaced *Fourier frequencies* ω_j. The **periodogram** is then defined as

$$I_T(\omega_j) = |d(\omega_j)|^2 = T^{-1} \left| \sum_{t=1}^{T} x_t \mathrm{e}^{2\pi \mathrm{i} \omega_j t} \right|^2,$$

where $\omega_j = \frac{j}{2n_\omega},\ j = 0, \ldots, n_\omega = \lceil \frac{T+1}{2} \rceil$.

Spectral estimation: the periodogram

Recall that the discrete Fourier transform is defined as

$$d(\omega_j) = T^{-1/2} \sum_{t=1}^{T} x_t e^{2\pi i \omega_j t},$$

for equally spaced *Fourier frequencies* ω_j. The **periodogram** is then defined as

$$I_T(\omega_j) = |d(\omega_j)|^2 = T^{-1} \left| \sum_{t=1}^{T} x_t e^{2\pi i \omega_j t} \right|^2,$$

where $\omega_j = \frac{j}{2n_\omega}$, $j = 0, \ldots, n_\omega = \lceil \frac{T+1}{2} \rceil$.

- However, the periodogram is an *inconsistent* estimator of the spectrum (i.e. $\text{var}(I_T(\omega_j)) \nrightarrow 0$ as $T \to \infty$), and so the periodogram is usually smoothed to remedy this.

Periodogram examples

Let $X_t = 2\cos(2\pi 6t/100) + 4\cos(2\pi 10t/100) + 6\cos(2\pi 40t/100)$.

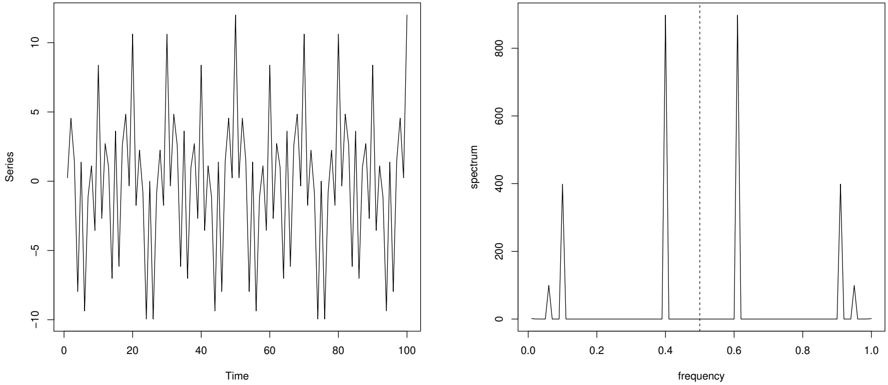


Figure: Periodogram of X_t, featuring three periodicities at distinct frequencies ("full" frequency range).

Periodogram examples

Let $X_t = 2\cos(2\pi 6t/100) + 4\cos(2\pi 10t/100) + 6\cos(2\pi 40t/100)$.

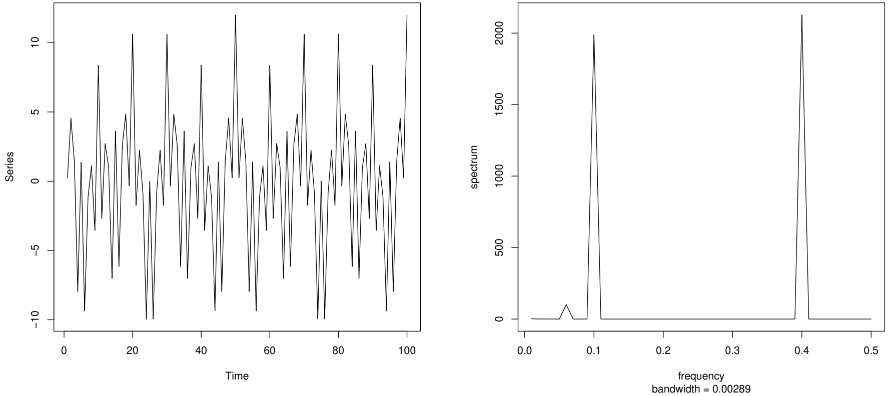
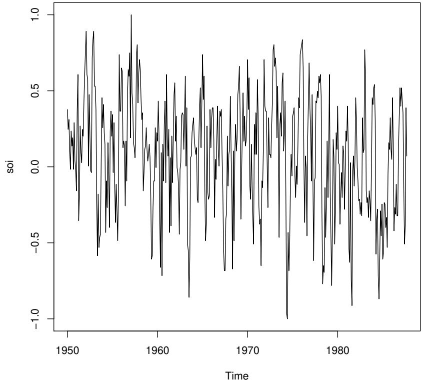


Figure: Periodogram of X_t, featuring three periodicities at distinct frequencies.

Periodogram examples

Let X_t be the `soi` (Southern Oscillation Index) series (below).



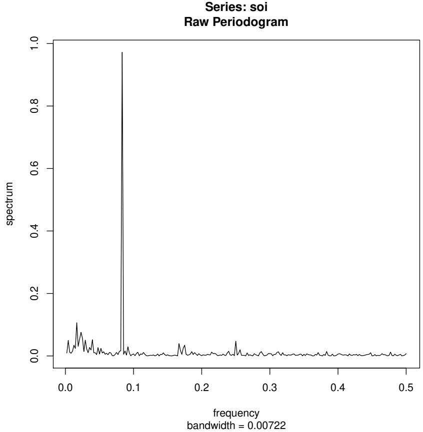


Figure: Periodogram of the Southern Oscillation Index.

Spectral estimation

There are many ways to perform spectral estimation. Nonparametric spectral estimation methods are derived from the spectrum definition, e.g.

- Average (frame) periodogram/ Welch's Method

- Blackman-Tukey estimator (using a weighted average of periodogram values)

Parametric methods assume some sort of model / form for the spectrum,e.g. AR Spectral approximation.

Forecasting

There are many ways to forecast a time series, depending on your intuition and the model. For example, one could use

- a naive estimator: $\hat{y}_{t+1} = y_t$
- a moving average: $\hat{y}_{t+1} = \frac{1}{K}\sum_{k=1}^{K} y_{t+1-k}$
- exponential moving average: $\alpha y_t + (1-\alpha)\hat{y}_t$
- if there are trend and seasonal components, these can also be taken into account by using similar procedures, or using the model form

Other Remarks and considerations

There are many other issues in time series which are relevant. Here are some

- non-Gaussian errors: either transform (e.g. via log) or use a count process model
- addition of covariates (exogenous variables) straightforward
- Vector time series models: multivariate extensions (VARIMA) which include dependence between series
- second order nonstationarity: ARCH models, time-varying coefficients, locally stationary models
- In `R`, see the `base`, `forecast`, `VTS` packages.

References

Chatfield, C. (2004) *The Analysis of Time Series: An Introduction. Chapman & Hall/CRC Pr I Llc.*

Hyndman, R. J. and Athanasopoulos, G. (2018) Forecasting: Principles and Practice. OTexts.

Koopmans, L. H. (1995) The Spectral Analysis of Time Series. Academic Press.

Shumway, R. H. and Stoffer, D. S. (2017) Time Series Analysis and Its Applications. With R Examples. Springer.