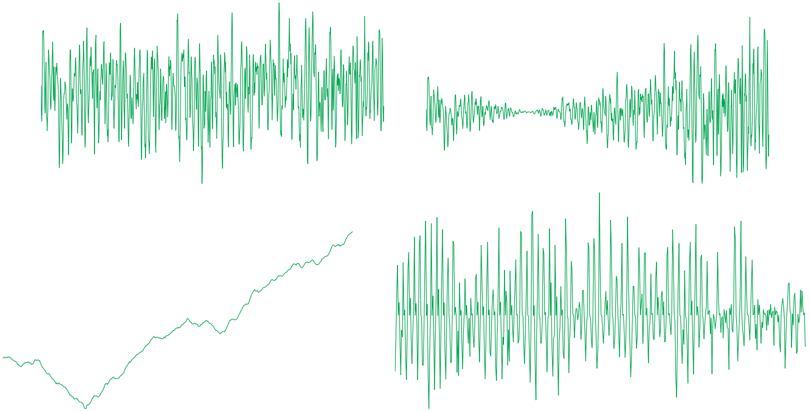
Estimation & Dependence in Space & Time
Lecture 1: Stationary Time Series

Sofia Olhede

EPFL

February 2, 2024

Different time series characteristics

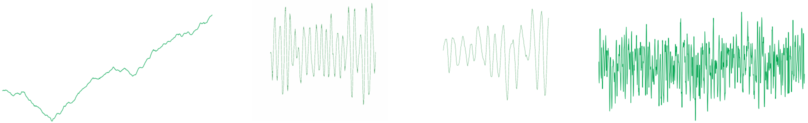


Mean reverting? Seasonal? Changing trend?

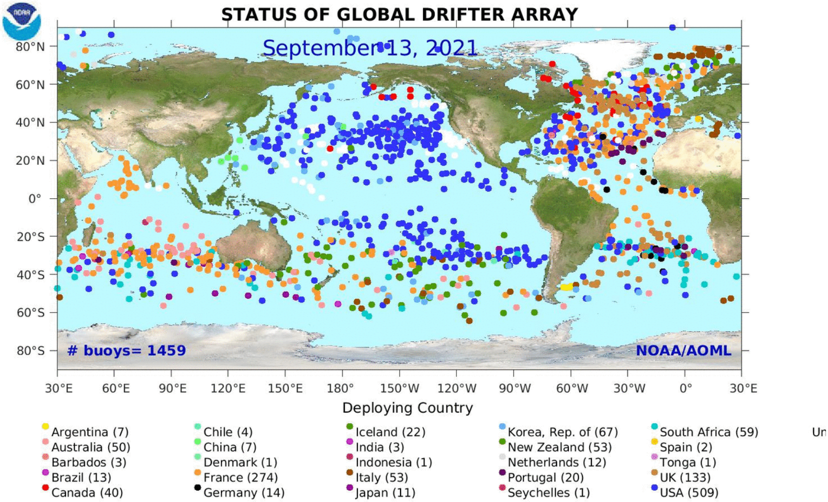
Unobserved Components Models

- In econometrics for example, the notion that a time series is an aggregation of different phenomological behaviours is common.
- The unobserved components model was championed by Harvey & Koopman as a basis for state-space modelling.
- Thus

$$\text{time series} = \text{Trend} + \text{Cyclical} + \text{Seasonal} + \text{Irregular}.$$



Examples


STATUS OF GLOBAL DRIFTER ARRAY
September 13, 2021
80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
80°S
buoys= 1459
NOAA/AOML
30°E
60°E
90°E
120°E
150°E
180°
150°W
120°W
90°W
60°W
30°W
0°
30°E
Deploying Country
Argentina (7)
Australia (50)
Barbados (3)
Brazil (13)
Canada (40)
Chile (4)
China (7)
Denmark (1)
France (274)
Germany (14)
Iceland (22)
India (3)
Indonesia (1)
Italy (53)
Japan (11)
Korea, Rep. of (67)
New Zealand (53)
Netherlands (12)
Portugal (20)
Seychelles (1)
South Africa (59)
Spain (2)
Tonga (1)
UK (133)
USA (509)

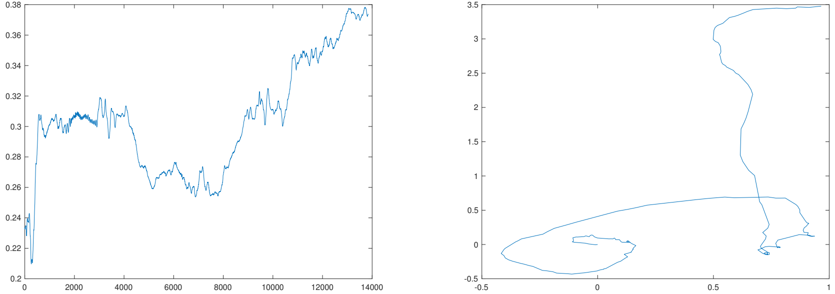
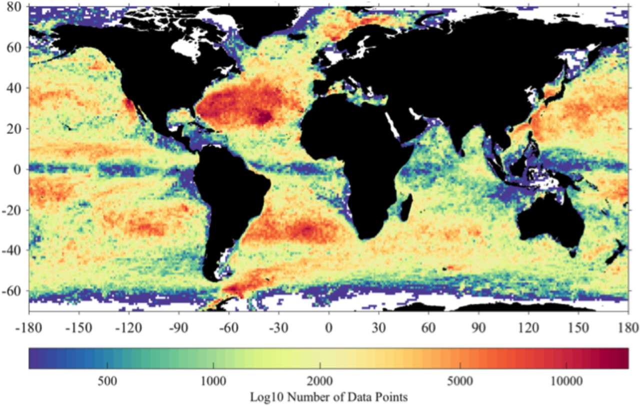


Figure: Measurements of velocity (cm/s) and position from a probe in the Global Drifter Programme.

Data density



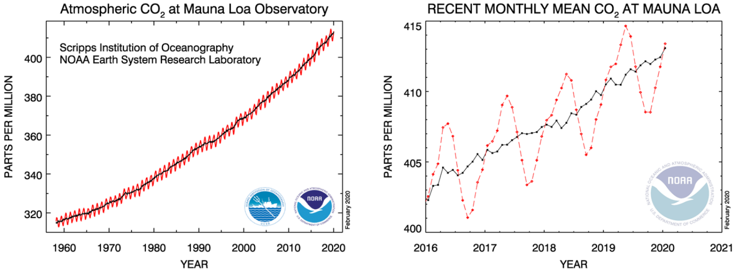
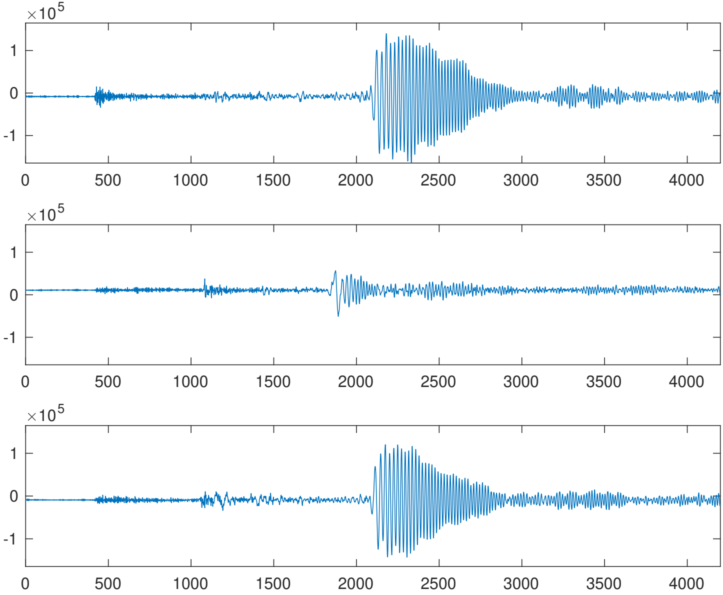


Figure: CO_2 measurements from Mauna Lao observatory.


×10^5
1
0
-1
0
500
1000
1500
2000
2500
3000
3500
4000
×10^5
1
0
-1
0
500
1000
1500
2000
2500
3000
3500
4000
×10^5
1
0
-1
0
500
1000
1500
2000
2500
3000
3500
4000

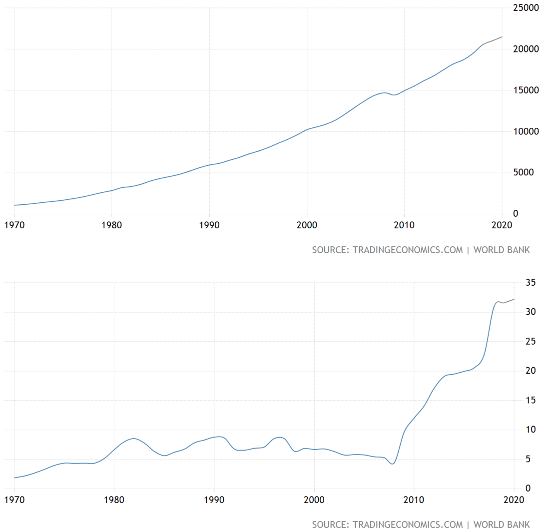
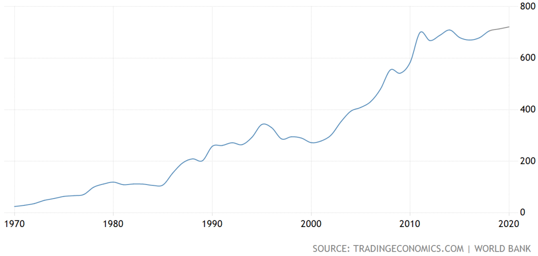


Figure: Gross domestic product (USA), (Zimbabwe).



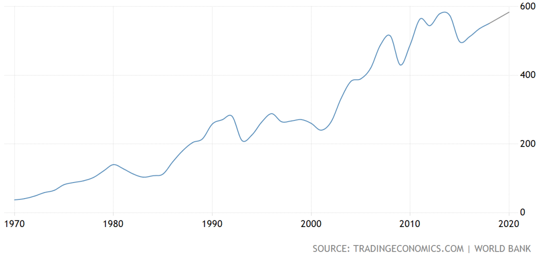
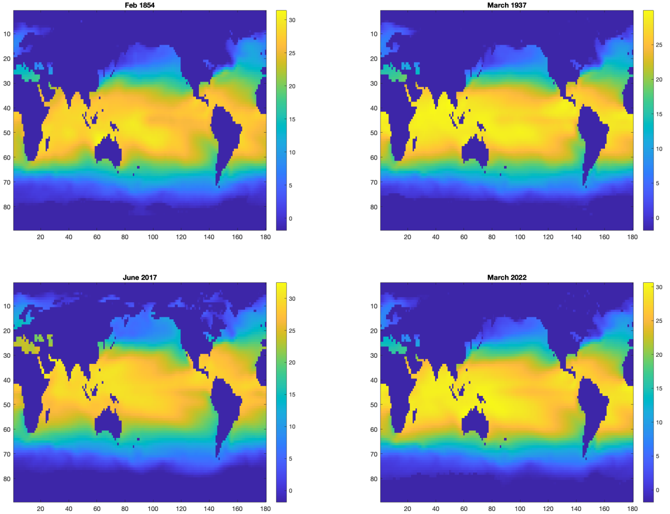


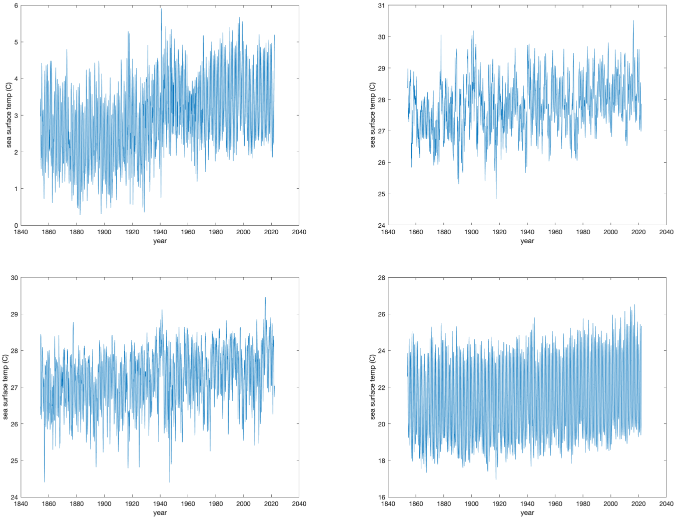
Figure: Gross domestic product (Switzerland) and (Sweden).

Spatio-temporal Data



How do we check if there have been processing errors?

Spatio-temporal Data II



Mean reverting? Seasonal? Changing trend?

Setting and Notation

What is a time series?

- Informally, a time series X_t is just data recorded over time.
- We shall use the word 'time series' to mean both **the data**, and **the process** from which the data is a realisation.
- More formally, we think of X_t as a stochastic process, namely as a family of random variables $\{X_t : \ t \in T\}$ defined on a probability space $(\Omega, \mathscr{F}, P)$.
- In time series analysis the index (or parameter) set T is a set of time points, very often $\mathbb{R}$ or $\Delta\mathbb{Z}$ (or a subset of them).

What is a time series practically?

- Whilst it is mathematically useful to think of processes with infinite index sets, in practice we can only make finitely many observations.
- Therefore, the set of observations X_t we actually record are in some set of time points $T' \subset T$.
- Normally T' is a discrete set (often with a regular sampling interval) $\{0, \Delta, \ldots, (N-1)\Delta\}$.
- The time series may also be recorded over an interval $T' = [0, T_0]$ (though it obviously cannot be stored digitally in this way directly).

Definition

Let $\mathscr{F}$ be the set of all vectors $\{\mathbf{t} = (t_1, \ldots, t_n)^T \in T^n : \ t_1 < t_2 < \cdots < t_n, \ n = 1, 2, \ldots\}$. Then the (finite-dimensional) distribution functions of $\{X_t \ t \in T\}$ are the functions $\{F_{\mathbf{t}}(\cdot), \ \mathbf{t} \in T\}$ defined for $\mathbf{t} = (t_1, \ldots, t_n)^T$ by

$$F_{\mathbf{t}}(\mathbf{x}) = \Pr(X_{t_1} \le x_1, \ldots, X_{t_n} \le x_n), \quad \mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n.$$

Theorem (Kolmogorov's theorem)

The probability distribution functions $\{F_{\mathbf{t}}(\cdot), \ \mathbf{t} \in \mathscr{F}\}$ are the distribution functions of a given stochastic process if and only if for any natural number n and $\mathbf{t} \in \mathscr{F}$ and $1 \le i \le n$ we have

$$\lim_{x_i \to \infty} F_{\mathbf{t}}(\mathbf{x}) = F_{\mathbf{t}(i)}(\mathbf{x}(i)),$$

where we have defined $\mathbf{t}(i)$ and $\mathbf{x}(i)$ as the $(n-1)$-component vectors obtained by deleting the ith component of $\mathbf{t}$ and $\mathbf{x}$ respectively.

Stationarity I

Definition ((Weak) Stationarity)

The time series $\{X_t\}$ is said to be second-order/weak or covariance stationary if for all $n \geq 1$ for any $t_1, \ldots, t_n \in T$ and for all τ such that $t_1 + \tau, \ldots, t_n + \tau \in T$ all the joint moments of order 1 and 2 of $X_{t_1}, \ldots, X_{t_n}$ exist, are all finite and equal to the corresponding joint moments of $X_{t_1+\tau}, \ldots, X_{t_n+\tau}$.

- In fact this corresponds to $\mathbb{E}\{X_t\} = \mu$, $\mathbb{V}\text{ar}\{X_t\} = \sigma^2$ and $\mathbb{E}\{X_{t_1} X_{t_2}\} = \mathbb{E}\{X_{t_1+\tau} X_{t_2+\tau}\}$. One may deduce from this that $\mathbb{E}\{X_{t_1} X_{t_2}\}$ is a function of $t_2 - t_1$ only.

Stationarity II

- We can go beyond the first two moments and define strong stationarity.

Definition ((Strong) Stationarity)

The time series $\{X_t\}$ is said to be completely/strong or strictly stationary if for all $n \geq 1$ for any $t_1, \ldots, t_n \in T$ and for all τ such that $t_1 + \tau, \ldots, t_n + \tau \in T$ the joint distribution of $X_{t_1}, \ldots, X_{t_n}$ is the same as $X_{t_1+\tau}, \ldots, X_{t_n+\tau}$.

- Note that second order stationary $\not\Rightarrow$ strictly stationary (in general). Strict stationarity $\not\Rightarrow$ 2nd order stationarity (in general). For example iid student t with non-finite variance.

Time domain analysis

Autocovariance

Normally we handle finite collections of random variables. To understand them better we often compute their covariance matrix. For a time series $\{X_t\}$ the extension of the covariance matrix corresponds to the autocovariance function. If the process is stationary and discrete time, this can be reduced to the autocovariance sequence.

Definition (The autocovariance sequence)

For a discrete time second-order stationary process $\{X_t\}$ we define the autocovariance sequence (ACVS) by

$$\gamma_\tau = \mathbb{C}\text{ov}\{X_t, X_{t+\tau}\} = \mathbb{C}\text{ov}\{X_0, X_\tau\}.$$

Properties of the autocovariance I

(i) Note that τ is the lag.

(ii) $\gamma_0 = \mathbb{V}\mathrm{ar}\{X_t\} = \sigma^2$ and $\gamma_{-\tau} = \gamma_\tau$.

(iii) The auto-correlation sequence (ACS) is defined as

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}.$$

(iv) As ρ_τ is a correlation it follows that $|\rho_\tau| \leq 1$. This implies that

$$\frac{|\gamma_\tau|}{\gamma_0} \leq 1 \Rightarrow |\gamma_\tau| \leq \gamma_0 = \mathbb{V}\mathrm{ar}\{X_t\}.$$

Properties of the autocovariance II

(v) The sequence $\{\gamma_\tau\}$ is positive semi-definite, that is for all $n \geq 1$ for any $t_1, \ldots, t_n \in T$ and for any real numbers $a_1, \ldots, a_n$ we have

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \gamma_{j-k} a_j a_k \geq 0.$$

(Follows easily by noting this equals the variance of the random variable $W = \sum_{j=1}^{n} a_j X_j$)

(vi) The covariance matrix of $\mathbf{X}$ is Toeplitz.

(a) Time Domain Models of First and Second Order Structure

- To model the distribution of $\mathbf{X}$ we hence posit forms for $\boldsymbol{\mu}_T$ and $\boldsymbol{\Sigma}_T$.
- **Second Order Stationarity**, corresponds to assuming:

$$\mu_t \quad = \quad \mu < \infty \quad \forall t, \quad [\boldsymbol{\Sigma}_T]_{ij} = \gamma_{t_i - t_j} < \infty. \qquad (1)$$

- **Strict Stationarity** corresponds to assuming the distribution of any finite sample, i.e. the distribution of $\mathbf{X}$, is equivalent to that of a time-shifted sample, for say shift $\tau \in \mathbb{N}$, $[X_{t_1+\tau}, \dots Z_{t_N+\tau}]$.
- For Gaussian Processes these two are equivalent.
- Directly specifies the model for the data.

Generating Mechanism

Definition (Gaussian process)

A process $\{X_t\}$ is a Gaussian if for every set of indices $T = \{t_1, \ldots, t_N\}$ the vector $\mathbf{X} = [X_{t_1}, \ldots X_{t_N}]$ is a vector-valued Gaussian random variable, i.e. for some mean $\boldsymbol{\mu}_T = [\mu_{t_1} \ldots \mu_{t_N}]$ and co-variance $\boldsymbol{\Sigma}_T$ the joint distribution of $\mathbf{Z}$ is given by:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_T)^T \boldsymbol{\Sigma}_T^{-1} (\mathbf{x} - \boldsymbol{\mu}_T)\right) \quad (2)$$

The joint distribution is then fully specified by noting $\boldsymbol{\mu}_T$ and $\boldsymbol{\Sigma}_T$.

Inferences:

- Inferences are sometimes made based on the time domain (log)-likelihood:

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\log(|\boldsymbol{\Sigma}_T(\boldsymbol{\theta})|) - \frac{1}{2}[\mathbf{X} - \mu\mathbf{1}]^T \boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta})[\mathbf{X} - \mu\mathbf{1}]. \tag{3}$$

- This is tedious to calculate as $\boldsymbol{\Sigma}_T$ is often very non-sparse.
- Other methods have been proposed to speed up calculations.

White noise

- An example of a stationary process is a white noise, also known as a purely random process. This corresponds to a sequence $\{X_t\}$ of uncorrelated RVs such that

$$\mathbb{E}(X_t) = \mu, \quad \mathbb{V}\text{ar}(X_t) = \sigma^2 \quad \forall t.$$

In this case

$$\gamma_\tau = \begin{cases} \sigma^2 & \text{if} \quad \tau = 0 \\ 0 & \text{otherwise} \end{cases},$$

or equivalently

$$\rho_\tau = \begin{cases} 1 & \text{if} \quad \tau = 0 \\ 0 & \text{otherwise} \end{cases}.$$

White noise is a building block for other time series models.

Moving average

- q-th order moving average (MA) process $\mathrm{MA}(q)$. This takes the form

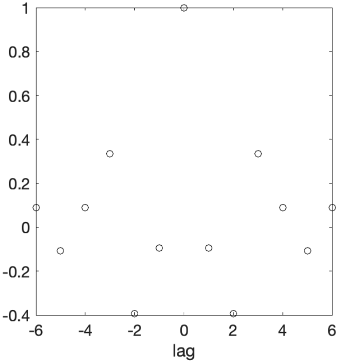
$$X_t = \mu - \theta_{0,q}\epsilon_t - \theta_{1,q}\epsilon_{t-1} - \cdots - \theta_{q,q}\epsilon_{t-q} = \mu - \sum_{j=0}^{q} \theta_{j,q}\epsilon_{t-j},$$

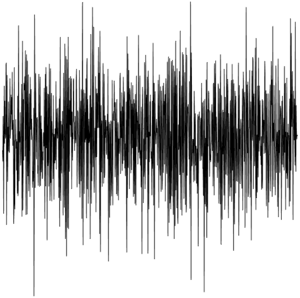
  and $\theta_{j,q}$ are constants ($\theta_{0,q} = -1$ and $\theta_{q,q} \neq 0$.)
- Also $\{\epsilon_t\}$ is normally assumed to be zero-mean white noise.
- We can calculate the moments of this process. We find that $\mathbb{E}\, X_t = \mu$.

$$\mathbb{C}\mathrm{ov}\{X_t, X_{t-\tau}\} = \mathbb{C}\mathrm{ov}\left\{\mu - \sum_{j=0}^{q} \theta_{j,q}\epsilon_{t-j}, \mu - \sum_{l=0}^{q} \theta_{l,q}\epsilon_{t-\tau-l}\right\}$$

$$= \mathbb{C}\mathrm{ov}\left\{\sum_{j=0}^{q} \theta_{j,q}\epsilon_{t-j}, \sum_{l=0}^{q} \theta_{l,q}\epsilon_{t-\tau-l}\right\} = \sum_{j=0}^{q}\sum_{l=0}^{q} \theta_{j,q}\theta_{l,q}\,\mathbb{C}\mathrm{ov}\{\epsilon_{t-j}, \epsilon_{t-\tau-l}\}$$

$$= \sigma_\epsilon^2 \sum_{j=0}^{q}\sum_{l=0}^{q} \theta_{j,q}\theta_{l,q}\delta_{t-j,t-\tau-l} = \begin{cases} \sigma_\epsilon^2 \sum_{j=0}^{q} \theta_{j,q}\theta_{j-\tau,q} & \text{if} \quad |\tau| \leq q \\ 0 & \text{otherwise} \end{cases}$$


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-6
-4
-2
0
2
4
6
lag



The Wold Decomposition

Theorem (The Wold Decomposition Theorem)

Assume that Z_t is a stationary zero-mean process. Then it admits representation

$$Z_t = \sum_{j=0}^{\infty} h_\tau \epsilon_{n-\tau} + \mu_t = U_n + \mu_n, \tag{4}$$

where $h_0 \equiv 1$, $\sum_j h_j^2 < \infty$, and ϵ_n is a zero-mean process satisfying:

1. $\mathbb{E}[\epsilon_n \epsilon_m] = \sigma_\epsilon^2 \delta_{n,m}$,
2. $\mathbb{E}[\epsilon_n \mu_m] = 0$, *for all n and m,*
3. μ_n *is purely deterministic.*

The Wold Decomposition II

- A function is purely deterministic if given its past it can be perfectly predicted.
- The Wold decomposition theorem historically important for the analysis of stationary processes,
- It can be used to justify the approximation of an arbitrary stationary time series via a truncated sum, or an MA model:

$$Z_t - \mu = \sum_{j=0}^{p} h_\tau \epsilon_{n-\tau}.$$

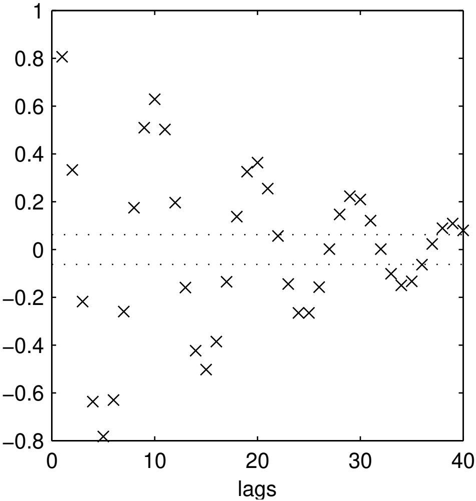
- If ϵ_n and ϵ_m are **independent** rather than **uncorrelated**, then Z_t is a linear process.

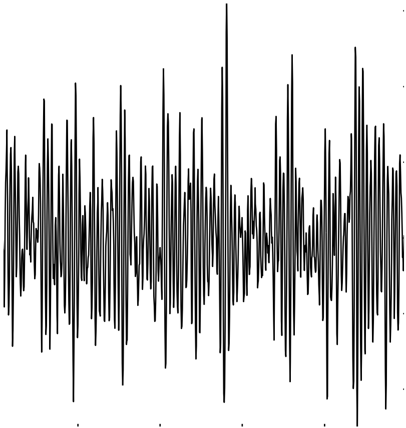
Autoregressive process

- pth order autoregressive process $\mathrm{AR}(p)$. This takes the form of

$$X_t = \phi_{1,p}X_{t-1} + \phi_{2,p}X_{t-2} + \cdots + \phi_{p,p}X_{t-p} + \epsilon_t.$$

  Here the set $\{\phi_{j,p}\}$ are constants, and ϵ_t is a white-noise process.
- In contrast with the moving average process, we have constraints on $\{\phi_{j,p}\}$ to obtain a stationary process.


1
0.8
0.6
0.4
0.2
0
–0.2
–0.4
–0.6
–0.8
0
10
20
30
40
lags



ARMA

Next process combined two mechanisms

- Auto-regressive Moving Average Process ARMA(p, q). This is specified by

$$X_t = \phi_{1,p} X_{t-1} + \cdots + \phi_{p,p} X_{t-p} + \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \cdots - \theta_{q,q}\epsilon_{t-q},$$

  for $t = 0, \pm 1, \pm 2, \pm 3, \ldots$.

- Harmonic (Sinusoidal) process

$$X_t = r^{-1} A(\omega)\cos(\nu t + \Theta(\omega)),$$

  for $t = 0, \pm 1, \pm 2, \pm 3, \ldots$.

How do we estimate time series models?

- In practice, we will only have access to (part of) a single realization.
- We will use a time-average to give time replication.
- Assume that the autocovariance satisfies $\sum_\tau |\gamma_\tau| < \infty$. and define

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i.$$

- What are the properties of this estimator?

$$\mathbb{E}\{\bar{X}\} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\{X_i\} = \mu.$$

- So $\bar{X}$ is unbiased as an estimator of μ.

- What about the variance? We say that $\bar{X}$ will converge to μ in mean square if

$$\lim_{N\to\infty} \mathbb{V}\text{ar}\{\bar{X}\} = 0.$$

- How do we figure this out?
- We calculate

$$\begin{aligned}\mathbb{V}\text{ar}\{\bar{X}\} &= \mathbb{E}\{(\bar{X}-\mu)^2\} \\ &= \mathbb{E}\{(\frac{1}{N}\sum_{i=1}^{N} X_i - \mu)^2\} \\ &= \frac{1}{N^2}\sum_i \sum_j \mathbb{E}(X_i-\mu)(X_j-\mu).\end{aligned}$$

How can we simplify this?

- We need to acknowledge the correlation in the process. If the covariance was σ^2 everywhere then we could not have mean square convergence.
- We find that

$$
\begin{aligned}
\mathbb{V}\text{ar}\{\bar{X}\} &= \frac{1}{N^2}\sum_i\sum_j \mathbb{E}(X_i-\mu)(X_j-\mu) \\
&= \frac{1}{N^2}\sum_i\sum_j \gamma_{j-i} \\
&= \frac{1}{N^2}\sum_{\tau=-(N-1)}^{N-1}(N-|\tau|)\gamma_\tau \\
&= \frac{1}{N}\sum_{\tau=-(N-1)}^{N-1}\frac{(N-|\tau|)}{N}\gamma_\tau
\end{aligned}
$$

- We now need the Césaro summability theorem which says that if $\sum_{\tau=-(N-1)}^{N-1} \gamma_\tau$ converges to a limit then $\sum_{\tau=-(N-1)}^{N-1} \frac{(N-|\tau|)}{N}\gamma_\tau$ converges to the same limit.
- Thus

$$\begin{aligned}
\lim_{N\to\infty} N \cdot \mathbb{V}\mathrm{ar}\{\bar{X}\} &= \lim_{N\to\infty} \sum_{\tau=-(N-1)}^{N-1} \frac{(N-|\tau|)}{N}\gamma_\tau \\
&= \lim_{N\to\infty} \sum_{\tau=-(N-1)}^{N-1} \gamma_\tau \\
&= \sum_{\tau=-\infty}^{\infty} \gamma_\tau = C^\gamma,
\end{aligned}$$

say.

- The assumption of absolute summability of $\{\gamma_\tau\}$ implies that $\{X_t\}$ has a purely continuous spectrum with sdf:

$$S(f) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau e^{2i\pi f \tau}.$$

- Thus it follows that

$$S(0) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau .$$

- Thus

$$\lim_{N\to\infty} N\,\mathbb{V}\mathrm{ar}\{\bar{X}\} = S(0),$$

- and so

$$\mathbb{V}\mathrm{ar}\{\bar{X}\} \asymp \frac{S(0)}{N}.$$

- We just showed that the sample mean was consistent, e.g. $\overline{X} \xrightarrow{P} \mu$, if the spectrum was a purely continuous spectrum.
- This is something we would expect for an iid sample, but as we saw due to the decay of the covariance sequence the pure continuity of the spectrum followed.
- Seems like a general idea: "when can we replace a sample average by a population average"? But what about the correlation? Does it matter? Does it change things?
- For example, consider the AR(1) process:

$$X_t = \phi X_{t-1} + \varepsilon_t,$$

for $X_0 \sim N(0, \frac{1}{1-\phi^2})$. Can we always average this? Do other statistical operations? What happens as ϕ changes?

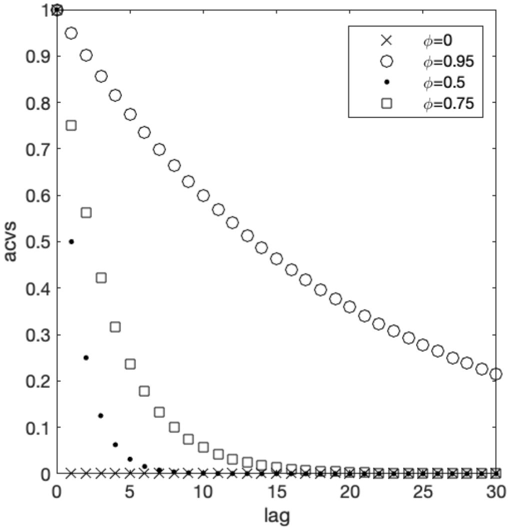


Figure: ACVS of the AR(1) with different values of ϕ.

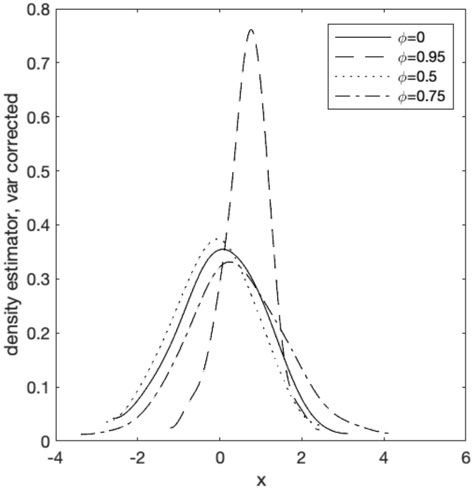


Figure: Density estmates from the AR(1) with different values of ϕ.

- We can now formalize this idea:

Definition ((Mean) Ergodic)

The time series $\{X_t\}$ is said to be mean ergodic if its first and second moments are finite and

$$\lim_{N\to\infty} \bar{X} \xrightarrow{P} \mathbb{E}\, X_t.$$

- The funny squiggle $\xrightarrow{P}$ means "converges in probability" and informally it implies that the mean stabilizes as the expectation tends to a constant and the variance goes to zero.
- The concept can be generalized to the dth moment for $d \geq 2$, not just for the mean.
- The informal understanding of this is "Sample means converge to population means," or "temporal averages are equivalent to population averages".
- Ergodicity and stationarity are not equivalent. The former concept is popular in econometrics.

Other Moments

- We also estimate other moments by method of moments, e.g. replacing the population moments with sample moments. Thus we take

$$\widehat{\gamma}_{\tau}^{(p)} = \frac{1}{N} \sum_{t=0}^{N-|\tau|-1} \{X_t - \hat{\mu}\}\{X_{t+\tau} - \hat{\mu}\}.$$

- This is a shrinkage estimator. Why?

The frequency domain

- For discrete–time stationary stochastic processes assume that X_t is a real-valued discrete time stationary process with zero mean.
- There exists an orthogonal increment process $\{Z(f)\}$ on $[-\frac{1}{2}, \frac{1}{2}]$ such that

$$X_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi ft} dZ(f). \tag{5}$$

This equality holds in the mean–square sense. The process $\{Z(f)\}$ has properties

1. $\mathbb{E}\{dZ(f)\} = 0$ for $|f| \leq 1/2$.
2. $\mathbb{E}\{|dZ(f)|^2\} = dS^{(I)}(f)$ for $S^{(I)}(f)$ the integrated spectrum.
3. For any two disjoint frequencies $f_1 \neq f_2$ $\mathbb{C}\text{ov}\{dZ(f_1), dZ(f_2)\} = 0$. This orthogonality of the increment process is a very useful result.

- Note that the covariance of two complex random variables Z_1 and Z_2 is defined as

$$\mathbb{C}\text{ov}\{Z_1, Z_2\} = \mathbb{E}\{(Z_1 - EZ_1)(Z_2 - EZ_2)^*\}.$$

- The complementary covariance or relation is defined as

$$\mathbb{R}\text{el}\{Z_1, Z_2\} = \mathbb{E}\{(Z_1 - EZ_1)(Z_2 - EZ_2)\}.$$

Definition

A Complex–Proper Process A Complex–Proper Process Z_t satisfies that its relation-sequence $r(\tau)$ is

$$r(\tau) = \mathbb{R}\text{el}\{Z_t, Z_{t-\tau}\} = 0.$$

- Note that

$$\begin{aligned} X_t &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi ft} dZ(f) \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi ft + i\arg\{dZ(f)\}} |dZ(f)|. \end{aligned} \tag{6}$$

- We note that if X_t is real–valued with zero–mean then

$$\begin{aligned}
\gamma_\tau &= \mathbb{C}\text{ov}\{X_t, X_{t+\tau}\} \\
&= \mathbb{E}\{X_t X_{t+\tau}\} \\
&= \mathbb{E}\{X_t^* X_{t+\tau}\} \\
&= \mathbb{E}\left\{\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-2i\pi ft} dZ(f) \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi f'(t+\tau)} dZ(f')\right\} \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-2i\pi ft} e^{2i\pi f'(t+\tau)} dS^{(I)}(f)\delta(f-f') \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi f\tau} dS^{(I)}(f).
\end{aligned}$$

Thus the integrated spectrum determines the autocovariance for a stationary process.

- If in fact $S^{(I)}(f)$ is differentiable so that $\frac{dS^{(I)}(f)}{df} = S(f)$ i.e. $dS^{(I)}(f) = S(f)df$ then it follows

$$\gamma_\tau = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi f\tau} S(f)df.$$

- $S(f)$ is called the spectral density of X_t.
- In fact it transpires that $\mathbb{E}\{|dZ(f)|^2\} = S(f)df$.
- To understand this fully, we need to study Fourier theory.
- As a reminder we note that a square summable deterministic sequence $\{g_t\}$ has as a Fourier representation

$$g_\tau = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi f\tau} G(f)df.$$

This is the Inverse Fourier Transform.

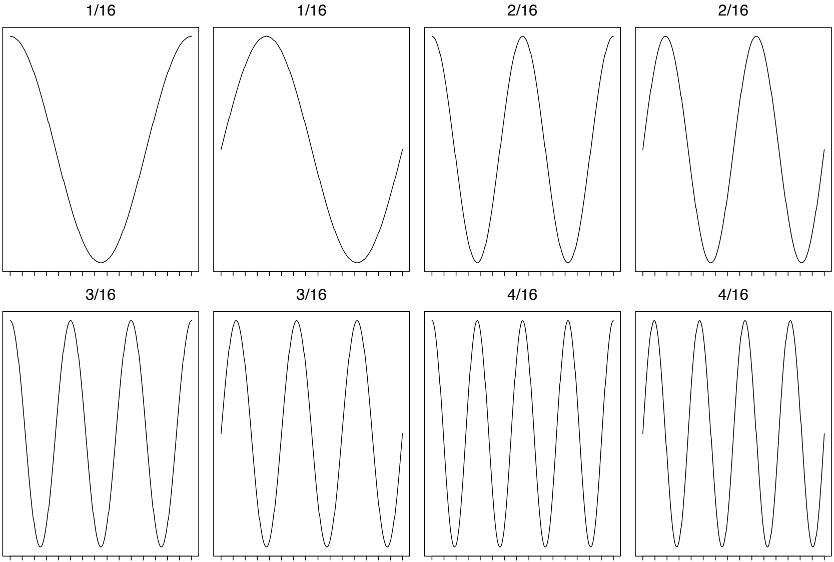
- The forward Fourier Transform is given by

$$G(f) = \sum_{\tau} g_{\tau} e^{-2i\pi f \tau}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2}.$$

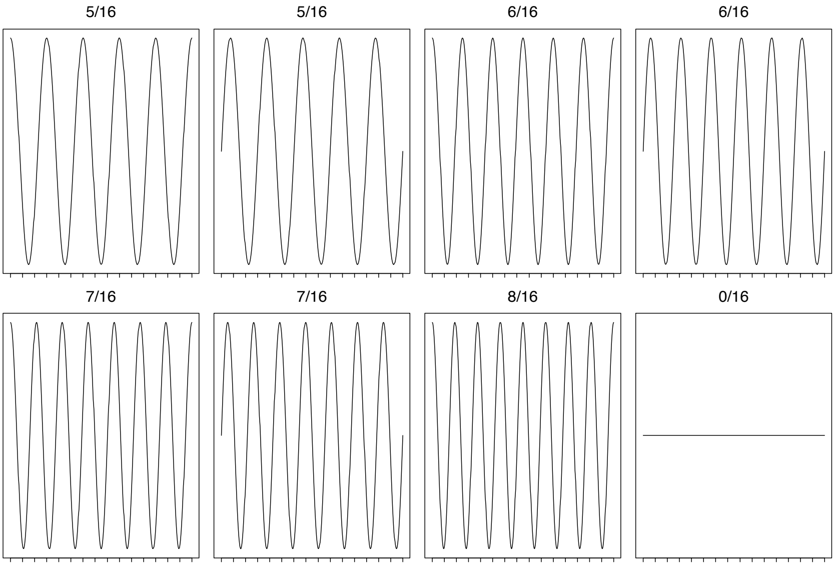
- Note that in this definition we are using discrete time and continuous frequency. This will have consequences in our analysis as we shall see later.
- $\{g_\tau\}$ and $G(f)$ form a Fourier transform pair. This is a bijection; namely $\{g_\tau\}$ is given by $G(f)$, and vice versa. We write $g_\tau \leftrightarrow G(f)$.
- In this manner, the autocovariance sequence $\{\gamma_\tau\}$ forms a Fourier transform pair with the spectral density function $S(f)$, and so

$$\gamma_\tau = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi f \tau} S(f) df \tag{7}$$

$$S(f) = \sum_{\tau} e^{-2i\pi f \tau} \gamma_\tau. \tag{8}$$



Slides from Emma McCoy.



Slides from Emma McCoy.

- In various books people use $\omega = 2\pi f$ instead of f. This is called angular frequency rather than frequency. I don't like it, because it causes factors of 2π to fly around all over the place, that are easily missed/forgotten, but it is equivalent.
- Furthermore $S(f)df$ is the average contribution over all realizations of the process to the power from components with frequencies in a small interval around f. The variance or "power" of X_t is

$$\gamma_0 = \mathbb{V}\text{ar}\{X_t\} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi f 0} S(f) df. \tag{9}$$

- We note the following properties of the spectral density function (sdf, $S(f)$) and the integrated spectrum ($S^{(I)}(f)$):
  - (i) The integrated spectrum, is the integrated spectrum

$$S^{(I)}(f) = \int_{-\frac{1}{2}}^{f} S(f')df'.$$

  - (ii) $0 \leq S^{(I)}(f) \leq \sigma^2 = \mathbb{V}\text{ar}\{X_t\}$.
  - (iii) $S^{(I)}(-\frac{1}{2}) = 0$ and $S^{(I)}(\frac{1}{2}) = \mathbb{V}\text{ar}\{X_t\}$.
  - (iv) $f < f' \Rightarrow S^{(I)}(f) \leq S^{(I)}(f')$. Also $S(-f) = S(f)$.
- Since $S^{(I)}(f)$ is quite similar to a probability distribution function we have the following theorem.

- We can characterise a stochastic process by its integrated spectrum.

Theorem (Lebesgue decomposition theorem)

Any integrated spectrum $S^{(I)}(f)$ can be written as

$$S^{(I)}(f) = S_1^{(I)}(f) + S_2^{(I)}(f) + S_3^{(I)}(f),$$

where the three contributions are all non-negative, non-decreasing functions with $S_1^{(I)}(-\frac{1}{2}) = 0$ for $j = 1, 2, 3$ and

(1) *$S_1^{(I)}(f)$ is an absolutely continuous function, i.e. its derivates exist for almost all f and is equal almost everywhere to its spectral density function (sdf) $S(f)$ such that*

$$S_1^{(I)}(f) = \int_{-\frac{1}{2}}^{f} S(f')df'.$$

(2) *The function $S_2^{(I)}(f)$ is a step function with jumps of size $\{p_l\}$ at the points $\{f_l\}_l$ where f_l are frequencies of pure sinusoids.*

(3) $S_3^{(I)}(f)$ is a continuous singular function (pathological and generally of no practical use).

- We can then characterise some common scenarios in terms of this decomposition:

(a) This case corresponds to $S_1^{(I)}(f) \geq 0$ and $S_2^{(I)}(f) \equiv 0$. In this case we say that $\{X_t\}$ has a purely continuous spectrum. Note that as $S_1^{(I)}(f)$ is absolutely continuous and non-decreasing (often increasing). Hence its derivative $S(f)$ is absolutely integrable (see for example Titchmarsh, The Theory of Functions). But note that if $S(f)$ is absolutely integrable $\int_{-\frac{1}{2}}^{\frac{1}{2}} \cos(2\pi f'\tau)S(f')df' \rightarrow 0$ $\int_{-\frac{1}{2}}^{\frac{1}{2}} \sin(2\pi f'\tau)S(f')df' \rightarrow 0$ as $\tau \rightarrow \infty$.

- But then (de Moivre) as $\tau \rightarrow \infty$:

$$\gamma_\tau = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi f\tau} S(f)df = \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos(2\pi f'\tau)S(f)df + i\int_{-\frac{1}{2}}^{\frac{1}{2}} \sin(2\pi f'\tau)S(f)df \rightarrow 0.$$

- We have discussed estimating properties of $\{X_t\}$ where this is a discretely indexed stochastic process, e.g. $t \in \mathbb{N}$.
- But actually, in real-life problems, it is often convenient to imagine this is a sample from a continuous time process $\{X(t)\}$, $t \geq 0$.
- Thus a discrete process $\{X_t\}$ is usually obtained by sampling a continuous time-process at equal time intervals $\Delta t > 0$.
- For a given sampling interval $\Delta t > 0$ and an arbitrary starting point t_0 we can define a discrete time process as

$$X_t = X(t_0 + t\Delta t), \quad t = 0, \pm 1, \pm 2, \ldots .$$

- If $\{X(t)\}$ is a stationary process with say sdf $S_{X(t)}(f)$ and autocovariance function

$$\gamma(\tau) = \mathbb{C}\text{ov}\{X(t), X(t+\tau)\},$$

then $\{X_t\}$ is also a stationary process with sdf $S_{X_t}(f)$ and autocovariance sequence $\{\gamma_\tau\}$.

- We start with the spectral representation for $\{X(t)\}$, which is

$$X(t) = \int_{-\infty}^{\infty} e^{2i\pi ft} dZ_{X(t)}(f).$$

- Then we have that

$$\begin{aligned}
X_t &= X(t_0 + t\Delta t) \\
&= \int_{-\infty}^{\infty} e^{2i\pi f(t_0 + t\Delta t)} dZ_{X(t)}(f) \\
&= \sum_{k=-\infty}^{\infty} \int_{\frac{2k-1}{2\Delta t}}^{\frac{2k+1}{2\Delta t}} e^{2i\pi f(t_0 + t\Delta t)} dZ_{X(t)}(f) \\
&= \sum_{k=-\infty}^{\infty} \int_{\frac{2k-1}{2\Delta t}}^{\frac{2k+1}{2\Delta t}} e^{2i\pi f t_0} e^{2i\pi f t\Delta t)} dZ_{X(t)}(f).
\end{aligned}$$

- We now do a change of variable $f \mapsto f' + \frac{k}{\Delta t}$.

Then we have

$$
\begin{aligned}
X_t &= \sum_{k=-\infty}^{\infty} \int_{\frac{-1}{2\Delta t}}^{\frac{1}{2\Delta t}} e^{2i\pi\left(f'+\frac{k}{\Delta t}\right)t_0} e^{2i\pi\left(f'+\frac{k}{\Delta t}\right)t\Delta t} dZ_{X(t)}\left(f'+\frac{k}{\Delta t}\right) \\
&= \int_{-\frac{1}{2\Delta t}}^{\frac{1}{2\Delta t}} e^{2i\pi f' t\Delta t} \sum_{k=-\infty}^{\infty} e^{2i\pi\left(f'+\frac{k}{\Delta t}\right)t_0} dZ_{X(t)}\left(f'+\frac{k}{\Delta t}\right) \\
&= \int_{-\frac{1}{2\Delta t}}^{\frac{1}{2\Delta t}} e^{2i\pi f t\Delta t} dZ(f).
\end{aligned}
$$

Thus we end up with the regular spectral representation. $\{dZ(f)\}$ is the orthogonal increment for the discrete time process. We find

$$
\mathbb{E} dZ(f) = \mathbb{E} \sum_{k=-\infty}^{\infty} e^{2i\pi\left(f+\frac{k}{\Delta t}\right)t_0} dZ_{X(t)}\left(f+\frac{k}{\Delta t}\right) = 0
$$

- By the orthogonality of $dZ_{X(t)}(f)$ and $dZ_{X(t)}(f')$ it follows that $\mathrm{Cov}\{dZ(f), dZ(f')\} = 0$. (If $f \neq f'$).
- We can thus take $dZ(f)$ to be the increments of the orthogonal process $Z(f)$ in the spectral representation for $X(t)$.
- The integrated spectrum for $\{X_t\}$ is

$$\begin{aligned} dS^{(I)}_{X_t}(f) &= \mathbb{E}\{|dZ(f)|^2\} \\ &= \sum_{k=-\infty}^{\infty} \mathbb{E}\left\{\left|dZ_{X(t)}\left(f + \frac{k}{\Delta t}\right)\right|^2\right\} \\ &= \sum_{k=-\infty}^{\infty} dS^{(I)}_{X(t)}\left(f + \frac{k}{\Delta t}\right) \quad |f| \leq \frac{1}{2\Delta t}. \end{aligned}$$

- Thus when $S^{(I)}_{X(t)}(f)$ is differentiable, so is also $S^{(I)}_{X_t}(f)$.
- And when that is the case we have aliasing or "fahltung".

$$S_{X_t}(f) = \sum_{k=-\infty}^{\infty} S_{X(t)}\left(f + \frac{k}{\Delta t}\right), \quad |f| \leq \frac{1}{2\Delta t}.$$

- The frequency $\frac{1}{2\Delta t}$ is called the Nyquist frequency.
- As we previously took $\Delta t = 1$ this yields $-\frac{1}{2} \leq f \leq \frac{1}{2}$.
- If $S_{X(t)}(f)$ is essentially zero for $|f| \geq \frac{1}{2\Delta t}$ we can expect a good correspondence between $S_{X_t}(f)$ and $S_{X(t)}(f)$ for $|f| \leq \frac{1}{2\Delta t}$.
- This basically is equivalent to assuming $S_{X(t)}\left(f + \frac{k}{\Delta t}\right) \approx 0$ for $k = 1, 2, \ldots$.
- If $S_{X(t)}(f)$ is large for $|f| \geq \frac{1}{2\Delta t}$ then the correspondence is poor so $S_{X_t}(f)$ tells us nothing about $S_{X(t)}(f)$.

- We recall the relationship

$$S(f) = \Delta t \sum_{\tau=-\infty}^{\infty} \gamma_\tau e^{-2i\pi f \tau \Delta t}$$

- We can therefore produce a spectral estimator from $\widehat{\gamma}_\tau^{(p)}$ by appending the sequence with zeros.

$$\begin{aligned}
\widehat{S}^{(p)}(f) &= \Delta t \sum_{\tau=-(N-1)}^{N-1} \widehat{\gamma}_\tau^{(p)} e^{-2i\pi f \tau \Delta t} \\
&= \frac{\Delta t}{N} \sum_{\tau=-(N-1)}^{N-1} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|} e^{-2i\pi f \tau \Delta t} \\
&= \left| \frac{\sqrt{\Delta t}}{\sqrt{N}} \sum_{j=1}^{N} X_j e^{-2i\pi f j \Delta t} \right|^2 .
\end{aligned}$$

- Note that $\widehat{S}^{(p)}(f)$ is defined over $\left[-\frac{1}{2}, \frac{1}{2}\right]$ and $\widehat{S}^{(p)}(f) \leftrightarrow \widehat{\gamma}_\tau^{(p)}$. This mirrors $S^{(p)}(f) \leftrightarrow \gamma_\tau^{(p)}$.
- Further we have

$$\widehat{\gamma}_\tau^{(p)} = \int_{-\frac{1}{2\Delta t}}^{\frac{1}{2\Delta t}} \widehat{S}^{(p)}(f) e^{2i\pi f \Delta t \tau} df$$

- Ideally as an estimator we would have
  (a) $\mathbb{E}\widehat{S}^{(p)}(f) = S(f)$,
  (b) Var $\widehat{S}^{(p)}(f) \to 0$ as $N \to \infty$ (consistency)
  (c) $\text{Cov}\left\{\widehat{S}^{(p)}(f), \widehat{S}^{(p)}(f')\right\} = 0$ for $f \neq f'$

- However instead we find that
  (a) $\mathbb{E}\widehat{S}^{(p)}(f) = S(f)$, is approximately valid
  (b) is false
  (c) holds approximately if f and f' have a particular form, the so-called Fourier frequencies.
- Furthermore, considering the expected value we arrive at

$$\mathbb{E}\widehat{S}^{(p)}(f) = \Delta t \sum_{\tau=-(N-1)}^{N-1} \mathbb{E}\widehat{\gamma}_{\tau}^{(p)} e^{-2i\pi f \tau \Delta t}$$
$$= \Delta t \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) \gamma_{\tau} e^{-2i\pi f \tau \Delta t}$$

- Thus if we know $\{\gamma_\tau\}$ then we can work out $\mathbb{E}\widehat{S}^{(p)}(f)$.

- We obtain more insight if we consider

$$\widehat{S}^{(p)}(f) = \left| \frac{\sqrt{\Delta t}}{\sqrt{N}} \sum_{j=1}^{N} X_j e^{-2i\pi f j \Delta t} \right|^2$$

- We then define

$$J(f) = \frac{\sqrt{\Delta t}}{\sqrt{N}} \sum_{j=1}^{N} X_j e^{-2i\pi f j \Delta t}, \quad \widehat{S}^{(p)}(f) = |J(f)|^2.$$

- In fact, from one of the problem sheets

$$\mathbb{E}\widehat{S}^{(p)}(f) = \int_{-\frac{1}{2\Delta t}}^{\frac{1}{2\Delta t}} S(f')\mathcal{F}_N(f - f')df'$$

with $\mathcal{F}_N(f) = \frac{\sin^2(N\Delta t \pi f)}{\sin^2(\Delta t \pi f)}$. This is the Fejer kernel.

- Set $\Delta = 1$. Thus $\mathbb{E}\left\{\widehat{S}^{(p)}(f)\right\}$ is the convolution of $S(f)$ and the imaging kernel $\mathcal{F}_N(f)$. The properties of $\mathcal{F}_N(f)$ are
  - (a) For all integers $N \geq 1, \mathcal{F}_N(f) \to N$ as $f \to 0$.
  - (b) For $N \geq 1 f \in \left[-\frac{1}{2}, \frac{1}{2}\right], f \neq 0 \mathcal{F}_N(f) < \mathcal{F}_N(0)$.
  - (c) For $N \geq 1 f \in \left[-\frac{1}{2}, \frac{1}{2}\right], f \neq 0 \mathcal{F}_N(f) \to 0$ as $N \to \infty$ or $f \to \infty$.
  - (d) For any integer k such that $f_k = k/N$ (Fourier frequencies) $\mathcal{F}_N(f_k) = 0$.
  - (e) Normalization

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{F}_N(f) df = 1$$

- From (a), (c) and (e) it follows that as $N \to \infty \mathcal{F}_N(f)$ acts like a delta function.

- Since $S(f)$ is normally assumed continuous we can informally argue

$$\lim_{N\to\infty} \mathbb{E}\left\{\widehat{S}^{(p)}(f)\right\} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{F}_N(f-f')S(f')df'$$
$$\approx \int_{-\frac{1}{2}}^{\frac{1}{2}} \delta(f-f')S(f')df' = S(f)$$

- The performance of the periodogram depends on the time series being analysed.
- Examples. For white noise $S(f) = \sigma_\epsilon^2$ for $|f| \leq 1/2$.

$$\mathbb{E}\left\{\widehat{S}^{(p)}(f)\right\} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{F}_N(f-f')S(f')df'$$
$$= \sigma_\epsilon^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{F}_N(f-f')df'$$
$$= \sigma_\epsilon^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{F}_N(f')df' = \sigma_\epsilon^2$$

- Thus we may note that the periodogram is unbiased for all values of N.
- Let us study an AR(2) process. Define

$$10 \log_{10} \left\{ \frac{\max_f S(f)}{\min_f S(f)} \right\},$$

as the dynamic range of the spectrum.

- How can we reduce the bias of the spectral estimator.
- One method corresponds to tapering. The idea is to change $\mathcal{F}_N(f)$ to something that decays faster.
- We will form the product

$$\{h_t X_t, t = 1, \ldots, N\}.$$

- We refer to $\{h_t\}$ a data taper. We normally assume h_t is real-valued.

- We define

$$J(f)=\sum_{t=1}^{N}h_tX_te^{-2i\pi ft}.$$

- Starting from first principles we get

$$\begin{aligned}J(f)&=\sum_{t=1}^{N}h_t\int_{-\frac{1}{2}}^{\frac{1}{2}}dZ(f')e^{2i\pi f' t}e^{-2i\pi ft}\\&=\int_{-\frac{1}{2}}^{\frac{1}{2}}H(f-f')dZ(f'),\quad H(f)=\sum_{t=1}^{N}h_te^{-2i\pi ft}\end{aligned}$$

- Finally we define

$$\widehat{S}^{(d)}(f)=|J(f)|^2=\left|\sum_{t=1}^{N}h_tX_te^{-2i\pi ft}\right|^2$$

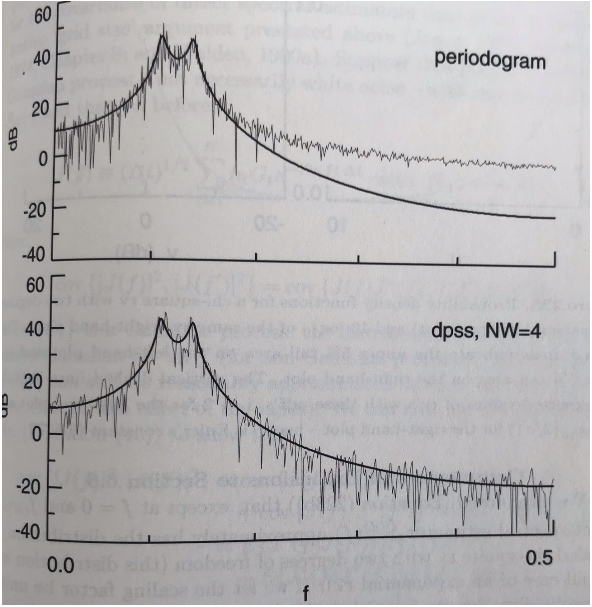


Image from Percival and Walden (1993).

- We can recover the periodogram by taking $h_t = \frac{1}{\sqrt{N}}$. In that case $\widehat{S}^{(d)}(f) = \widehat{S}^{(p)}(f)$.
- So

$$\begin{aligned}
\mathbb{E}\left\{\widehat{S}^{(d)}(f)\right\} &= \mathbb{E}\{J(f)J^*(f)\} \\
&= \mathbb{E}\left\{\int_{-\frac{1}{2}}^{\frac{1}{2}} H(f-f')dZ(f') \int_{-\frac{1}{2}}^{\frac{1}{2}} H^*(f-f'')dZ^*(f'')\right\} \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f-f')H^*(f-f'')\mathbb{E}\{dZ(f')dZ^*(f'')\} \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{H}(f-f')S(f')df'
\end{aligned}$$

- We take $\mathcal{H}(f) = |H(f)|^2$. Also chose to take $\sum_{t=1}^{N} h_t^2 = 1$.

- Assume you have a set of tapers $\{h_{t,k}\}$ for $t = 1, \ldots, N$ and $k = 0, \ldots, K-1$.
- We define a spectral estimate for each value of k. We therefore have

$$\widehat{S}_k^{(d)}(f) = \left| \sum_{t=1}^{N} h_{t,k} X_t e^{-2i\pi f t} \right|^2.$$

- Orthogonality implies that we assume

$$\sum_{t=1}^{N} h_{t,k} h_{t,j} = \delta_{j,k}.$$

- The simplest multitaper estimate is

$$\widehat{S}^{(mt)}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \widehat{S}_k^{(d)}(f).$$

- We can determine its expectation:

$$\mathbb{E}\left\{\widehat{S}^{(mt)}(f)\right\} = \frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\left\{\widehat{S}_k^{(d)}(f)\right\}$$
$$= \frac{1}{K}\sum_{k=0}^{K-1}\int_{-\frac{1}{2}}^{\frac{1}{2}}\mathcal{H}_k(f-f')S(f')df'$$
$$= \int_{-\frac{1}{2}}^{\frac{1}{2}}\overline{\mathcal{H}}(f-f')S(f')df'$$

- We call $\overline{\mathcal{H}}(f) = \frac{1}{K}\sum_{k=0}^{K-1}\mathcal{H}_k(f)$ the average kernel.
- We can note that

$$\mathrm{Var}\left\{\widehat{S}^{(mt)}(f)\right\} \approx \frac{S^2(f)}{K}$$

Frequency domain likelihoods

- Easy to state permitted auto-covariance functions using Bochner's theorem. A form for $\gamma_X(\tau)$ if

$$S(f;\boldsymbol{\theta}) = \sum_{\tau=-\infty}^{\infty} \gamma_X(\tau) e^{-2i\pi f \tau}. \quad (10)$$

- Note that

$$\gamma_X(0) \equiv \sigma^2 = \int_{-1/2}^{1/2} S(f)\, df. \quad (11)$$

- The population mean μ is estimated by $\hat{\mu}$ and removed from the data.
- Remaining structure found in $\gamma_X(\tau)$ which also needs to be estimated.

- We cannot determine $dZ(f)$.
- Instead the Discrete Fourier Transform (DFT) of sample $\mathbf{X}$ is calculated:

$$J(f) = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} [X_t - \widehat{\mu}] \exp(-2\pi i t f).$$

- From Brillinger we can note that with $f_k = \frac{k}{N}$ with appropriate cumulant conditions:

$$J(f_k) \overset{d}{=} N^C(0, S(f_k))$$

  where $N^C(\cdot,\cdot)$ denotes the complex proper distribution (see earlier).
- $\{J(f_k)\}_{k=1}^{\lfloor N/2 \rfloor - 1}$ are asymptotically uncorrelated (there is some funkiness with frequency grid), not quite the same as jointly normal.

- Define $\widehat{S}^{(p)}(f)(f) = |J(f)|^2$ then the discrete Whittle's likelihood can be written as:

$$\ell(\boldsymbol{\theta}) = -\frac{1}{N}\sum_{j=0}^{\lfloor N/2 \rfloor}\left\{\frac{\widehat{S}^{(p)}(f_j)}{S(f_j;\boldsymbol{\theta})} + \log(S(f;\boldsymbol{\theta}))\right\}$$

- Then maximum likelihood corresponds to

$$\widehat{\boldsymbol{\theta}}_{wl} = \arg_{\boldsymbol{\theta}} \max \ell(\boldsymbol{\theta}) = \arg_{\boldsymbol{\theta}} \min(-\ell(\boldsymbol{\theta})).$$

- Whittle likelihood can be implemented over a subset of all frequencies (Robinson), or with tapering (Thomson).

- When so we expect the Whittle likelihood to work well?
- It is an example of a composite likelihood (obtained by multiplying together likelihood for components (that are not necessarily independent), see B. Lindsay (1988)).
- What composite likelihoods could be formed? In time we might chop up region, or select a subset of spatial observations, partially conditioning on a nearby region (Vecchia (1988), has been added to by Katzfuss and Guinness (2021)).
- We could instead form a composite likelihood in frequency.
- For finite samples the Fourier coefficients are correlated, and their variance is not the spectrum.
- It is however very fast to compute.

Whittle's Likelihood II

- This also is an approximation to the time-domain likelihood (see review paper by Dzhaparidze [17] and).
- Relies on eigenvectors of circulant matrices etc.
- Is usually discretized and based on the Discrete Fourier Transform.
- A series of papers by Whittle [47, 48], establish the properties of this approximation, corresponding to the continuous Whittle likelihood. For certain families $\int \log\{S(f;\boldsymbol{\theta})\}\, df = 0$, and this motivates the simplified form without the log term.

Discussion

- This lecture introduced the notion of a probabilistic invariance; that of shift invariance, or temporal stationarity.
- We explored a number of consequence of this, a) temporally independent moments, b) the spectral representation and c) the Wold decomposition.
- Frequency domain estimation was explored.

References I

- L. Brieman, J. Friedman, R. Olshen and C. Stone,
  Classification and Regression Trees,
  Wadsworth, CA, 1984.
- D. Brillinger,
  Time Series: Data Analysis and Theory,
  SIAM, 2001.
- M. B. Priestley,
  Non-Linear and Non-Stationary Time Series,
  Academic Press, London, 1991.

References II

S. Adak,
Time-Dependent Spectral Analysis of Nonstationary Time Series,
J. Am. Stat. Assoc., vol. 93, pp. 1488–1501.

M. G. Amin,
Time and Lag Window Selection in Wigner-Ville Distribution,
in *Proc. IEEE Int. Conf. on Acoust. Speech and Signal Proc*, pp. 1529–1532, 1987.

R. G. Baraniuk and D. L. Jones,
Shear Madness – New Orthonormal Bases and Frames using Chirp Functions,
IEEE Transactions on Signal Processing, vol. 41, pp. 3543–3549, 1993.

M. Casdagli,
Chaos and Deterministic versus Stochastic Non-Linear Modelling,
J. Roy. Stat. Soc. B, vol. 54, pp. 303–328, 1992.

References III

G. Y. H. Chi,
Multiplicity and Representation Theory of Generalized Random Processes,
J. of Multivariate Analysis, vol. 1, pp. 412–432, 1971.

R. Coifman and M. Wickerhauser,
Entropy based Algorithms for Best Basis Selection,
IEEE Transactions on Information Theory, vol. 32, pp. 712–718, 1992.

J. Cooley, P. Lewis, and P. Welch,
The finite Fourier transform,
IEEE Trans. Audio Electroacoustics, vol. 17, pp. 77–85, 1969.

H. Cramér,
A contribution to the theory of stochastic process,
In *Proc. Second Berkeley Symp. Math. Statist. and Probability*, pp. 228–352, University of California, Berkeley, CA.

References IV

H. Cramér,
Structural and Statistical Problems for a class of Stochastic Processes
S. S. Wilks memorial lecture, Princeton University Press, Princeton, N. J.

G. R. Dargahi-Noubary, P. J. Laycock and T. Subba Rao,
Non-linear stochastic models for seismic events with applications in event identification,
Geophys. J. Roy. Astron. Soc., vol. 55, pp. 655–668, 1978.

R. Dahlhaus,
Fitting Time Series Models to Nonstationary Processes,
The Annals of Statistics, vol. 25, pp. 1–37, 1997.

R. Dahlhaus,
A likelihood approximation for locally stationary processes,
The Annals of Statistics, vol. 28, pp. 1762–1794, 2000.

References V

I. Daubechies,
Orthonormal Bases of Compactly Supported Wavelets,
Communications on Pure and Applied Mathematics, vol. 41, pp. 909–996, 1988.

K. O. Dzhaparidze and A.M. Yaglom,
Spectrum parameter estimation in time series analysis
in Developments in Statistics, volume 4, pages 1–96, Elsevier, 1983.

D. L. Donoho, S. Mallat, R. von Sachs and Y. Samuelides,
Locally Stationary Covariance and Signal Estimation with Macrotiles,
IEEE Trans. Signal Proc., vol. 51, pp. 614–627, 2003.

Y. Grenier,
Time-Dependent ARMA modeling of Nonstationary Signals,
IEEE Trans. Acoust., Speech and Signal Proc., vol. 31, pp. 899–911, 1983.

References VI

J. K. Hammond,
On the response of single and multi degree-of-freedom systems to non-stationary random excitations,
J. Sound & Vibr., vol. 7, pp. 393–416, 1968.

J. K. Hammond,
Evolutionary Spectra in Random Vibrations,
J. Roy. Stat. Soc. B, vol. 35, pp. 167–188, 1973.

T. Hida,
Canonical Representations of Gaussian Processes and their applications,
Mem. Coll. Sci. Kyoto Univ. Sec. 4, vol. 32, pp. 109–155, 1960.

Y. Hosoya and M. Taniguchi,
A central limit theorem for stationary processes and the parameter estimation of linear processes.
The Annals of Statistics, 10, 132–153, 1982.

References VII

J. Kampé de Feriet and F. N. Frenkiel,
Correlation and spectra of nonstationary random functions,
Math. Comp., vol. 10, pp. 1–21, 1962.

K. Karhunen,
Über lineare Methoden in der Wahrscheinlichkeitsrechnung,
Ann. Acad. Sci. Fenn Ser A, I. Math., vol. 37, pp. 3–79, 1947.

A. S. Kayhan, A. El-Jaroudi, L. F. Chaparro,
Evolutionary Periodogram for Nonstationary Signals,
IEEE Transactions on Signal Processing, vol. 42, pp. 1527–1536 1994.

A. S. Kayhan, A. El-Jaroudi, L. F. Chaparro,
Data Adaptive Evolutionary Spectral Estimation,
IEEE Transactions on Signal Processing, vol. 43, pp. 204–213, 1995.

T.-H. Li and H.-S. Oh,
Wavelet spectrum and its characterization property for random processes,
IEEE Transactions on Information Theory, vol. 48, pp. 2922–2937, 2002.

References VIII

- M. Loève,
  Fonctions aléatoires du secondre ordre,
  In *Processes Stochastiques et Movement Browien*, by P. Lévy, pp. 228–352, Gauthier-Villars, Paris.
- R. M. Loynes,
  On the Concept of the Spectrum for Non-Stationary Processes,
  J. Roy. Stat. Soc. B, vol. 30, pp. 1–30, 1968.
- S. Mallat, G. Papanicolau and Z. Zhang,
  Adaptive Covariance estimation for locally stationary processes,
  Annals of Statistics, vol. 26, pp. 1–47, 1998.
- W. Martin and P. Flandrin,
  Wigner-Ville Spectral-Ananlysis of Non-Stationary Processes,
  IEEE Trans. Acoust. Speech Signal Proc., vol. 33, pp. 1461–1470, 1985.

References IX

K. S. Miller,
Complex Gaussian Processes,
SIAM Review, vol. 11, pp. 544–567, 1969.

G. P. Nason, R. von Sachs and G. Kroisandt,
Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum,
J. Roy. Stat. Soc. B, vol. 62, pp. 271–292, 2000.

G. P. Nason and R. von Sachs,
Wavelets in Time-Series Analysis,
Philosophical Transactions: Mathematical, Physical and Engineering Sciences, vol. 357, pp. 2511-2526, 1999.

S. C. Olhede and P. J. Wolfe.
Network histograms and universality of blockmodel approximation.
Proceedings of the National Academy of Sciences 111.41 (2014): 14722-14727.

References X

H. C. Ombao, J. A. Raz, R. von Sachs and B. A. Malow,
Automatic Statistical Analysis of Bivariate Nonstationary Time Series,
J. Am. Stat. Assoc., vol. 96, pp. 543–560, 2001.

E. Parzen,
Spectral Analysis of asymptotically stationary time series,
Bull. Internat. Statist. Inst., vol. 39, pp. 87–103, 1962.

A. Pintore and C. C. Holmes,
Non-stationary covariance functions via spatially adaptive spectra,
Imperial College London Statistics Section Technical Report, 2004.

M. B. Priestley.
Evolutionary Spectra and Non-Stationary Processes,
Journal of the Royal Statistical Society, B, vol. 27, pp. 204–237, 1965.

Y. A. Rozanov,
Spectral Analysis of Abstract functions,
Theory Probab. Appl., vol. 4, pp. 271–287, 1959.

References XI

P. D. Sampson and P. Guttorp,
Nonparametric Estimation of Nonstationary Spatial Covariance Structure,
J. Am. Stat. Assoc., vol. 87, pp. 108–119, 1992.

Arthur Schuster,
On the Periodicities of Sunspots,
Philosophical Transactions of the Royal Society of London. Series A, Vol. 206, pp. 69-100, 1906.

D. Tjøstheim,
Spectral Generating Operators for Non-Stationary Operators,
Advances in Applied Probability, vol. 8, pp. 831–846, 1976.

M. K. Tsatsanis and G. B Giannakis,
Time-varying system identification and model validation using wavelets,
IEEE Transactions on Signal Processing, vol. 41, pp. 3512–3523, 1993.

References XII

M. K. Tsatsanis and G. B Giannakis,
Subspace methods for blind estimation of time-varying FIR channels,
IEEE Transactions on Signal Processing, vol. 45, pp. 3084–3093, 1997.

P. Whittle,
Estimation and information in stationary time series,
Arkiv för Matematik, vol. 2, pp. 423–434, 1952.

P. Whittle,
Gaussian Estimation in Stationary Time Series,
Bulletin of the International Statistical Institute, vol. 39, pp. 105–129, 1962.

Y. Zheng, Z. Lin and D. B. H. Tay,
Time-varying parametric system multiresolution identification by wavelets,
Int. J. of Syst. Sci., vol. 32, pp. 775–793, 2001.