01/16

Time Series Analysis

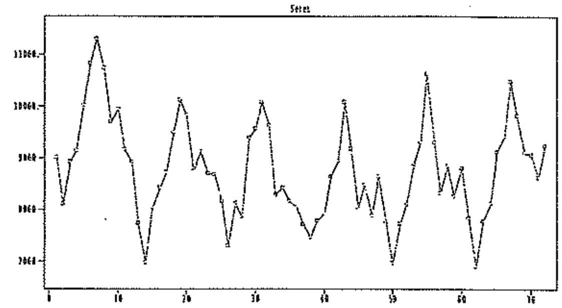
1. Introduction

Examine time series plots for:
- trend over time;
- seasonal/cyclical/periodic components;
- changing variability over time;
- "carry-over" effects from past;
- other systematic features.

1.1 Examples.

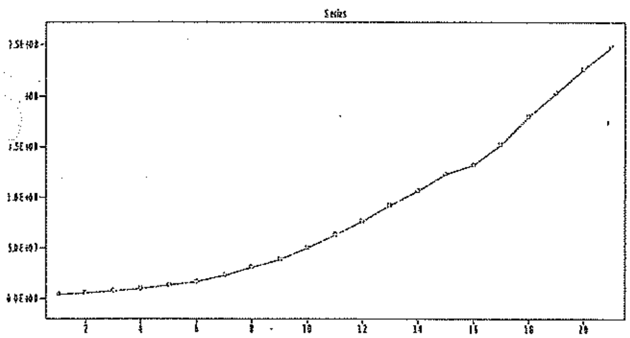
(Refer to B&D 2016 and the accompanying ITSM software, as well as R package **itsmr**.)

Accidental Deaths in U.S.A.
("deaths" in itsmr, "USAaccDeaths" in R base)
- Monthly totals: January 1973 to December 1978.
- Strong seasonal pattern: high in July, low in Feb.



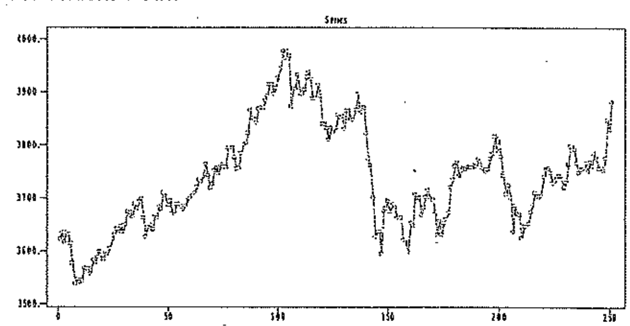
- strong seasonality of period 12 (monthly)

U.S.A. Population (uspop in R base)
- Ten-year intervals: 1790 to 1970. "uspop" in R.
- Exponential trend.
- Little or no random variability.



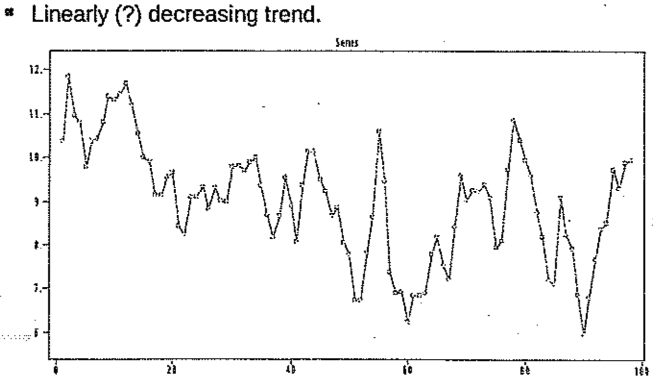
- strong exponentially growing trend

Dow-Jones Index
(DJAO2.TSM in B&D, similar to "dowj" in itsmr)
- Closing prices on 251 consecutive trading days, ending 08/26/94.
- A Random Walk?



- erratic (random walk?)

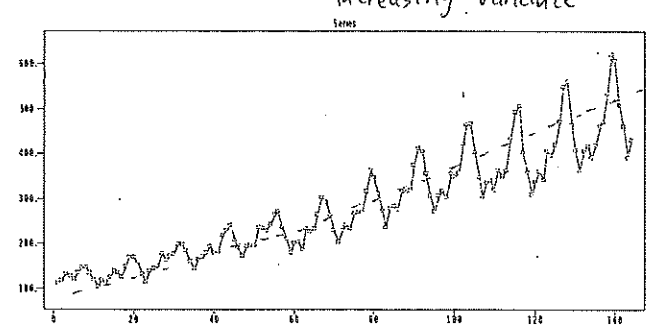
Level of Lake Huron
("lake" in itsmr, LakeHuron-570 in R base)
- Annual levels (feet): 1875-1972.
- Linearly (?) decreasing trend.



- Decreasing trend?

International Airline Passengers ("airpass" in itsmr)
- Monthly totals: January 1949 to December 1960;
- Strong seasonal effect: high in summer, low in winter.
- Linearly (?) increasing trend.
- Increasing variability.

period 12 (monthly)
increasing variance



- linear increasing trend? ; - increasing variability.
- strong cycle, strong trend!

01/16

Time Series - Preamble

, Sample : $y_1, \dots, y_n$ from Y

↳ Aim : make inference on $Y \sim ?\ (\mu, \sigma^2)$

let $\underset{\sim}{y} = (y_1, \dots, y_n)'$.

(1) Math-Stat : $y_1, \dots, y_n \overset{iid}{\sim} f(y; \overset{\text{location}}{\mu}, \overset{\text{scale}}{\sigma})$

A simple location-scale model is : $y_i = \mu_i + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} (0, \sigma^2)$

$$\Rightarrow \underset{\sim}{y} \sim (\underset{\sim}{\mu}, \sigma^2 I_n)$$

↳ No attempt is made to "understand" behavior of Y.

$$\Rightarrow \underset{\sim}{y} = \underset{\sim}{\mu} + \underset{\sim}{\varepsilon} \quad ; \quad \underset{\sim}{\varepsilon} \sim (\underset{\sim}{0}, \sigma^2 I_n)$$

(2) Regression : $y_1, \dots, y_n$ independent, but not identical (μ depends on X_i)

$$y_i = \underset{\sim}{x}_i' \underset{\sim}{\beta} + \varepsilon_i \quad ; \quad \varepsilon_i \overset{iid}{\sim} (0, \sigma^2)$$

$$\Rightarrow \underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \quad ; \quad \underset{\sim}{\varepsilon} \sim (\underset{\sim}{0}, \sigma^2 I_n) \quad \Rightarrow \quad \underset{\sim}{y} \sim (X\underset{\sim}{\beta}, \sigma^2 I_n)$$

↳ Try to "understand" what factors affect behavior of Y through its mean (conditioned on X fixed), $E(Y \mid X)$.

(3) Mixed Models / Linear Models : (LMM)

$y_1, \dots, y_n$ neither independent nor identically distributed.

$$y_i = \underset{\sim}{x}_i' \underset{\sim}{\beta} + \underset{\sim}{z}_i' \underset{\sim}{u} + \varepsilon_i \quad ; \quad \varepsilon_i \overset{iid}{\sim} (0, \sigma^2)$$

$$\Rightarrow \underset{\sim}{y} = \underbrace{X\underset{\sim}{\beta}}_{\text{fixed effects}} + \underbrace{Z\overset{\text{vector of rv}}{\underset{\sim}{u}}}_{\text{random effects}} + \underset{\sim}{\varepsilon} \quad ; \quad \underbrace{\underset{\sim}{\varepsilon} \sim (\underset{\sim}{0}, \sigma^2 I_n) \quad ; \quad \underset{\sim}{u} \sim (\underset{\sim}{0}, R)}_{\text{indep.}}$$

$\Rightarrow \quad \underset{\sim}{y} \sim (X\underset{\sim}{\beta}, \Sigma)$ where $\Sigma = ZRZ' + \sigma^2 I_n$. (random effects $\nearrow^0$ expectation.)

Try to understand factors affecting behavior of Y through its mean and variance. (conditional on fixed X).

(*) Points:

- In all above models, we are able to work toward iid residuals $\{\varepsilon_i\}$, i.e. factors "causing" the dependence structure in $\underset{\sim}{y}$ are extracted out in the terms to the left of ε_i term.
- The residuals contain behavior of Y that cannot be explained further (residual = left over).
- In data obs. over time & space, obtaining iid residuals is (nearly) impossible! (It's difficult to account for all the covariates that cause dependence...).

(4) Time Series Data : In simplest linear stochastic processes (ARMA models), $y_1, \dots, y_n$ are identical, but not independent : $y_1, \dots, y_n \sim \cancel{i}id$ (not indep.).
marginal dist^{n} are same (not indep.)

↳ Addition of a time-varying mean makes $y_1, \dots, y_n$ neither independent nor identical.

$$y_t = \underset{\sim}{X_t}'\underset{\sim}{\beta} + \varepsilon_t \quad ; \quad \varepsilon_t \sim \underbrace{ARMA(p,q)}_{\text{correlation structure}} \text{ model.}$$

$$\underset{\sim}{\varepsilon} \sim (\underset{\sim}{0}, S)$$

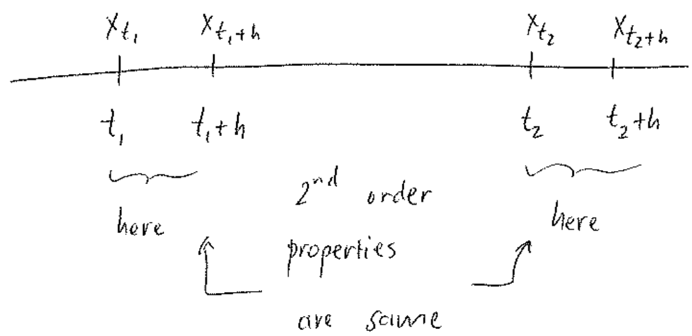
$$\Rightarrow \quad \underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \Rightarrow \underset{\sim}{y} \sim (X\underset{\sim}{\beta}, \Sigma), \quad \Sigma = S.$$

↳ complex, anything we like

(generalization case on (3)).

Because Σ has $\frac{n^2-n}{2} > n$ terms, simplifying assumptions need to be made to make inference with just one sample of size n.

a) Stationarity : The mean and correlation function of process (2^{nd} order properties) are invariant to time shifts :



$\boxed{01/18}$ **CH. 1. Stationary Processes.**

§1.1. Stochastic Processes :

A collection of r.v. $\{X(t) : t \in T\}$, T is an index set, is called a stochastic process. Other names :

- time series : if T is time.
- $X(t)$ is a random function.

Ex/ of indexed sets T :

- $\{1, 2, \dots\}$
- $\{0, \pm 1, \dots\}$

  } mostly these for the class.
- $(0, \infty)$
- $(-\infty, \infty)$

- Time series : Called continuous or discrete based on indexed set.

  Typical notation : - $X(t)$ denotes a continuous t.s. (T = continuous)
  - X_t " discrete " (T = discrete) ← focus!

- Notes : - We will focus primarily on discrete t.s.

$$\{X_t : t = 0, \pm 1, \pm 2, \dots\}$$

- We often abbreviate to $\{X_t\}$ or X_t
- X_t will always be a quantitative variable (for this course).

• Recall: Random variable X sample space

- function whose values depends on $\omega \in \Omega$
- observed value of X is x, is the value of X at the observed ω, i.e.

$$x = X(\omega).$$

• Defn: Realization (sample function, sample path) is the set of real-valued outcomes $\{x_t, t \in T\}$ for a fixed $\omega \in \Omega$, i.e. $x_t = X_t(\omega)$. We often use the notation X_t and x_t to denote a (discrete) ts and its realization.

X_t = theoretical object

x_t = observed value

• Note: If realization is typically of finite duration, i.e. of $T = \{0, \pm 1, \pm 2, \dots\}$ then a realization may be $\{x_t : t = 1, 2, \dots, n\}$.

clouds $\{X_t\}$, $t = 0, \pm 1, \dots$ (theory)

Modelling.

ground $\{x_t\}$, $t = 1, \dots, n$ (observed)

Ex/ Refer to the 5 examples in handout (BD book).

§1.2. Objectives of TSA: (time series analysis)

1) Model any deterministic trend and/or seasonality that may be present, and remove these from the data.

2) Choose from among a family of probability models that best models the residuals from step 1.

3) Estimate parameters of chosen models. ⋮ Most work happen in steps 2-4.

4) Check for the goodness-of-fit of the model.

5) Resulting model: • " provides a compact description of the data "

• " can be used for in-sample inference", eg $<\begin{matrix} CI \\ HT \end{matrix}$ on model para.

• " can also be used for out-of-sample inference (forecasting) "

• " can be used as inputs ~~as~~ into more complex models "

§1.3. Stationary Processes:

• Stationarity is a simplified assumption(s) on which many TSA techniques depend.

• Statistical properties do not change over time.

• Assessing its plausibility for real data is often difficult... $\begin{matrix} X_t \\ \updownarrow \\ x_t \end{matrix}$

(stationarity is a model property)

• Defn: (Strict Stationarity)

Joint distⁿ $(X_{t_1}, \dots, X_{t_k})$ (random vector variables) is same as that of $(X_{t_1+h}, \dots, X_{t_k+h})$, for all $t_1, \dots, t_k$ and any h.

For a ts $\{X_t\}$ with finite variance, define:

i) Mean fn: $\mu(t) = E(X_t)$

ii) Covariance fn: $\gamma(t+h, t) = Cov(X_t, X_{t+h})$

• Defn: (Weak Stationarity) The following 2 properties must hold:

i) $\mu(t) \equiv \mu$, does not depend on t

ii) $\gamma(t+h, t) \equiv \gamma(h)$, depends only on h, for any h.

01/23

- $\{X_t\}$ has :

  - mean function $EX_t = \mu_t \equiv \mu$
  - covariance function $Cov(X_t, X_{t+h}) = \gamma_X(t+h, t) \equiv \gamma(h)$

  } independent of t (under stationarity)

  ↳ anything with trend is not stationary, also same with cycles.

◈ Stationarity means weak stationarity.

- If $\{X_t\}$ is stationary, define its autocovariance function (ACVF) at lag h as:

$$\gamma(h) \equiv \gamma(h,0) = \gamma(t+h, t)$$

and likewise, its autocorrelation function (ACF) as:

$$\rho(h) \equiv \frac{\gamma(h)}{\gamma(0)} = Corr(X_{t+h}, X_t)$$

- ACF is a true correlation function in that $\rho(0) = 1$ and $-1 \leq \rho(h) \leq 1$.
- ACVF has the following properties:

  1) $\gamma(0) = Var(X_t) = \sigma^2 > 0$.
  2) $|\gamma(h)| \leq \gamma(0), \ \forall h$.
  3) $\gamma(h) = \gamma(-h), \ \forall h$.

Ex 1 5 series in handout :

a) US pop : not stationary : increasing mean (trend)

b) Airline : not stationary : increasing mean & variance (trend), periodicity.

c) Death : not stationary : periodicity.

→ Periods : changing correlation signs (+ then - even for same h values) ...

d) Lake

e) Dow-Jones

} erratic : not very clear (?)

Defn : White Noise. This is the simplest stationary process.

- $\{X_t\}$ is a white noise process (WN) if $EX_t = 0$ and $\gamma(h) = \sigma^2$ if $h=0$, and $\gamma(h) = 0$ otherwise : $\gamma(h) = \begin{cases} \sigma^2 & h=0 \\ 0 & \text{otherwise} \end{cases}$. $\to$ Var $= \sigma^2$, Cov $= 0$

Note : X_t's are uncorrelated (serially uncorrelated).

- $\{X_t\}$ is IID noise if it is WN, but with the X_t's being independent.

Notation : $\{X_t\} \sim WN(0, \sigma^2)$; $\{X_t\} \sim IID(0, \sigma^2)$

- Note : IID $\Rightarrow$ WN, but not the reverse, unless the process $\{X_t\}$ is Gaussian.

We also define the partial autocorrelation function (PACF) of $\{X_t\}$ at lag h, $\alpha(h)$, to be the correlation between X_1 and X_{1+h}, adjusted for the intervening observations $X_2, \dots, X_h$. Note : $\alpha(0) = 1$.

- The ACF & PACF will play a key role in ARMA model identification.

Review of Covariances :

Defn : For rv X and Y,

$$Cov(X,Y) = E\left[(X-EX)(Y-EY)\right] = EXY - EXEY.$$

Fact : 1) If X and Y are independent, $EXY = EXEY \Rightarrow Cov(X,Y) = 0$.

Converse is not true, unless (X,Y) are jointly Gaussian.

2) $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ sequences of rv

$a_0, a_1, \dots, a_n$ and $b_0, b_1, \dots, b_m$ Sequences of constants

Then : $$Cov\left(a_0 + \sum_{i=1}^{n} a_i X_i,\ b_0 + \sum b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j\, Cov(X_i, Y_j)$$

Ex/
- $Cov(a_0, Y_1) = 0$
- $Cov(a+bX, c+dY) = bd\, Cov(X,Y)$

- $Cov(a_1X_1 + a_2X_2,\ b_1Y_1 + b_2Y_2) = a_1b_1 Cov(X_1, Y_1) + a_1b_2 Cov(X_1, Y_2) + a_2b_1 Cov(X_2, Y_1)$
$$+ a_2b_2 Cov(X_2, Y_2)$$

- Fact: $\gamma(\cdot)$ is the ACVF of a stationary ts iff :

1) $\gamma(\cdot)$ is an even function, and
2) $\gamma(\cdot)$ is a non-negative definite function, i.e. $\sum_{i,j=1}^{n} a_i \gamma(i-j) a_j \geq 0$ $\forall$ n and $a_1, \dots, a_n$.

Ex/ Is $\gamma(h) = 1 + \cos\left(\frac{\pi h}{2}\right) + \sin\left(\frac{\pi h}{4}\right)$ an ACVF ?

No, $\gamma(h) \neq \gamma(-h)$ $\rightarrow$ not an even function.

Ex/ Is $\gamma(h) = \cos(\omega h)$ an ACVF ?

Yes. Verification of non-negative property is not easy (difficult !).

$\longrightarrow$ Easier to check that $\cos(\omega h)$ is the ACVF of :

EA, EB

$X_t = A\cos(\omega t) + B\sin(\omega t)$, A & B are uncorrelated $(0, 1)$ rv.

Var A, Var B

- $EX_t = 0$
- $\gamma(h) = EX_t X_{t+h} = E\left[A \cdot \cos(\omega t + \omega h) + B\sin(\omega t + \omega h)\right]\left[A\cos(\omega t) + B\sin(\omega t)\right]$

$= \cos(\omega t + \omega h)\cos(\omega t) + \sin(\omega t + \omega h)\sin(\omega t) = \cos(\omega h)$

$\Rightarrow$ $\{X_t\}$ is stationary. $\Rightarrow$ $\gamma(h) = \cos(\omega h)$ is the ACVF.

Ex/ Is $\gamma(h) = \begin{cases} 1 & h = 0 \\ 0.4 & h = \pm 1 \\ 0 & o.w \end{cases}$ an ACVF ? $\leftarrow$ Even ✓ Check for nnd !

Consider $X_t = Z_t + \theta Z_{t-1}$, $\{Z_t\} \sim WN(0, \sigma^2)$

$\hookrightarrow$ MA(1) model.

- $EX_t = EZ_t + \theta EZ_{t-1} = 0.$

- $\gamma(h) = E X_t X_{t+h} = \begin{cases} E(Z_t + \theta Z_{t-1})^2 & h=0 \\ E(Z_{t+1}+\theta Z_t)(Z_t + \theta Z_{t-1}) & h = \pm 1 \\ 0 & \text{o.w} \end{cases}$ $\leftarrow$ $E Z_i Z_j = 0 \quad \forall i \neq j$

$$= \begin{cases} \sigma^2 (1+\theta^2) & h=0 \\ \theta\sigma^2 & h=\pm 1 \\ 0 & \text{o.w} \end{cases}$$

Solving : $\left.\begin{array}{l} \sigma^2(1+\theta^2) = 1 \\ \theta\sigma^2 = \rho \text{ (general)} \\ \quad\ \ \| \\ \quad\ 0.4 \text{ (this case)} \end{array}\right\} \Rightarrow \begin{array}{l} \theta = \dfrac{1 \pm \sqrt{1-4\rho^2}}{2\rho} \\ \sigma^2 = \dfrac{1}{1+\theta^2} \end{array}$

$\Rightarrow$ real solution for $|\rho| \le \frac{1}{2}$ $\Rightarrow$ $\gamma(h)$ is ACVF of MA(1) with $\rho = 0.4$.

$\boxed{01/25}$ Prediction of a Stationary Process.

let $\{X_t\}$ be a stationary process (w.s) with $EX_t = \mu$, $Var X_t = \sigma^2$, ACF $\rho(h)$.

Question : How do we predict X_{t+h} given that we have observed X_t ? (let $\hat{X}_{t+h}$ be predictor).

Want / BP : $\begin{cases} \text{i) } E\hat{X}_{t+h} = \mu & \text{(ubiased), and} \\ \text{ii) } E(X_{t+h} - \hat{X}_{t+h})^2 \text{ should be small.} & \text{(MSE)} \end{cases}$

If $\{X_t\}$ is Gaussian, it can be shown that the best MSE predictor of X_{t+h} given X_t is given by $\hat{X}_{t+h}$ is $\underbrace{E(X_{t+h} \mid X_t)}_{\substack{\text{(BP)} \\ \hookrightarrow \text{ best MSE predictor in general}}} = \underbrace{\mu + \rho(h)(X_t - \mu)}_{\substack{\text{(BLP)} \\ \hookrightarrow \text{ under the Gaussianity assumption}}}$.

with MSE : $E(X_{t+h} - \hat{X}_{t+h})^2 = \underbrace{\sigma^2(1-\rho^2(h))}_{\hookrightarrow \text{ if Gaussian}}$

So: 1) For a Gaussian ts, the "best predictor" (BP) coincides with the "best linear predictor" (BLP).

2) Also, for causal & invertible ARMA processes, the BP and BLP on the infinite past coincide! (later).

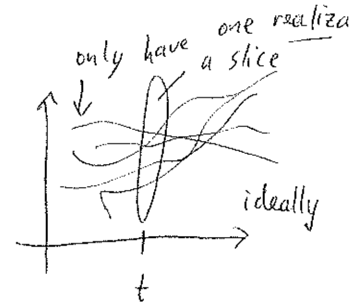
$$\ldots, X_{-1}, X_0, X_1, \ldots, X_n \rightarrow \text{ want } X_{n+h}.$$

→ For these reasons, we will focus attention only on BLP.

(Don't know dist^{n} of $\underbrace{E(X_{t+h} \mid X_t)}_{(BP)}$ to do conditional expectation).

§1.4. Sample Estimates.

. Observe : $\{x_1, \ldots, x_n\} \sim \overset{rv}{X_t}$, a stationary process with mean μ and ACVF $\gamma(h)$ and ACF $\rho(h)$.



◇ Mean: Estimate μ with the sample mean:

$$\bar{x}_n = \frac{1}{n} \sum_{t=1}^{n} x_t$$

hope : longitudinal = cross-sectional ↳ under ergodicity.

* Thm: (Ergodicity of $\bar{X}_n$)

$\bar{X}_n$ is ergodic for μ (converges in mean-square case):

iff $\lim_{n\to\infty} E(\bar{X}_n - \mu)^2 = 0$.

Fact: If $\lim_{h\to\infty} \gamma(h) = 0$ (or $\lim_{h\to\infty} \rho(h) = 0$), then $\bar{X}_n$ is ergodic to μ. ← used to verify

* Thm: (CLT for $\bar{X}_n$):

asymptotically normal

$$\bar{X}_n \sim AN\left(\mu, \frac{v}{n}\right) \iff \left[\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, v)\right]$$

$$\iff \frac{\sqrt{n}(\bar{X}_n - \mu)}{v} \sim N(0,1).$$

where $v = \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma(h)$

Note : If $\{X_t\} \sim WN(0,\sigma^2)$ $\left(X_t \overset{iid}{\sim} (0,\sigma^2) \dots \right) \Rightarrow \nu = \sigma^2 \Rightarrow \gamma(h) = \begin{cases} \sigma^2 & h=0 \\ 0 & \text{o.w.} \end{cases}$

(*) $(1-\alpha)100\%$ CI for μ:

$(1-\frac{\alpha}{2})$ quantile from $N(0,1)$

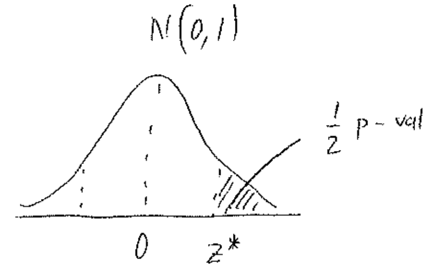
$$1-\alpha = P\left(-z_{1-\alpha/2} \le \frac{\bar{X}_n - \mu}{\sqrt{\nu/n}} \le z_{1-\alpha/2} \right)$$

$$= P\left(\bar{X}_n - z_{1-\alpha/2}\sqrt{\frac{\nu}{n}} \le \mu \le \bar{X}_n + z_{1-\alpha/2}\sqrt{\frac{\nu}{n}} \right)$$

$\hookrightarrow$ endpoints of CI

(*) α-level test of $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$.

TS : $z^* = \dfrac{\bar{X}_n - \mu_0}{\sqrt{\nu/n}} \sim N(0,1)$ under H_0.



- Reject H_0 if $|z| > z_{1-\frac{\alpha}{2}}$
- p-val $= 2P(Z > |z^*|) \Rightarrow$ reject H_0 if p-val $< \alpha$.

◇ ACVF : Estimate $\gamma(h)$ with :

$$\hat{\gamma}(h) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}) & , \quad 0 \le h \le n-1. \\ 0 & , \quad h \geq n \end{cases}$$

[Analog of estimating $Cov(X,Y)$, $\{(X_1,Y_1), \dots, (X_n,Y_n)\}$]

$\hookrightarrow \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

◇ ACF : Since $\rho(h) = \dfrac{\gamma(h)}{\gamma(0)} \Rightarrow \hat{\rho}(h) = \dfrac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$.

* Thm : (CLT for $\hat{\rho}(h)$)

$$\hat{\rho}(h) \sim AN\left(\rho(h), \frac{1}{n} w_{hh}\right)$$

where w_{hh} is given by Barlett's formula :

$$w_{hh} = \sum_{k=1}^{\infty} \left[\rho(k+h) + \rho(k-h) - 2\rho(h)\rho(k) \right]^2$$

• Replace unknown quantities with consistent sample estimates. (converging $n \to \infty$)

$$\left[95\% \text{ CI for } \rho(h) : \quad \hat{\rho}(h) \pm 1.96 \sqrt{\frac{w_{hh}}{n}} \right]$$

replace $\rho(h)$ with $\hat{\rho}(h)$.

$\boxed{01/28}$ last time ...

analog of consistency for t.s.

Defn : Ergodicity of $\bar{X}_n$:

$\bar{X}_n$ is ergodic iff : $\lim_{n \to \infty} E(\bar{X}_n - \mu)^2 = 0$

$\bar{X}_n \to \mu$ in mean square

$\Downarrow$

Fact : If $\lim_{h \to \infty} \gamma(h) = 0$ (or $\lim_{h \to \infty} \rho(h) = 0$), then $\bar{X}_n$ is ergodic for μ.

• ACVF / ACF Estimation : $\hat{\gamma}(h) = \ldots \Rightarrow \hat{\rho}(h) = \dfrac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$; $-1 \le \hat{\rho}(h) \le 1$

Thm : CLT for $\bar{X}_n$ (last time).

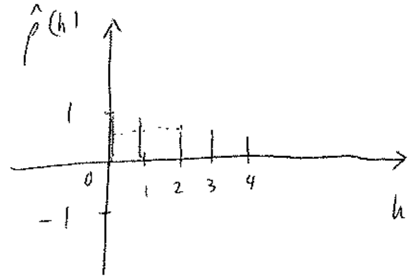
last time

Thm : CLT for $\hat{\rho}(h)$: $\hat{\rho}(h) \sim AN\left(\rho(h), \frac{1}{n} w_{hh}\right)$

$\Rightarrow$ 95% CI of $\rho(h)$: $\hat{\rho}(h) \pm 1.96 \sqrt{\dfrac{w_{hh}}{n}}$.

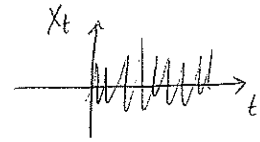
Fact : Ergodicity conditions for $\hat{\gamma}(h)$ are more complicated (Prietsley, 1981). The class of ARMA processes to be studied can be shown to be ergodic for $\bar{X}_n$ and $\hat{\gamma}(h)$.

<u>Defn</u>: A plot of $\hat{\rho}(h)$ vs. h is called an ACF plot or correlogram.



$\hat{\rho}(2) = 0.7 \Rightarrow corr(X_t, X_{t+2}) = 0.7$, for all $t \Rightarrow$ under stationarity.

↳ not: very slowly decaying: correlated.



$\Rightarrow$ gives valuable information about $\rho(h)$.

<u>Ex</u>/ (White Noise) If $X_t \sim iid(0, \sigma^2) \Rightarrow \gamma(h) = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \neq 0 \end{cases} \Rightarrow \rho(h) = \begin{cases} 1, & h = 0 \\ 0, & h \neq 0 \end{cases}$

- $v = \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma(h) = \gamma(0) = \sigma^2$ ← CLT of $\bar{X}_n$.

$\Rightarrow$ 95% CI for μ: $\bar{x}_n \pm 1.96\sqrt{\frac{v}{n}} = \bar{x}_n \pm 1.96 \frac{\sigma}{\sqrt{n}}$ ← usual CI, same as classic CLT (since $X_t \sim iid$).

- Now get CI for $\rho(h)$:

$$W_{hh} = \sum_{k=1}^{\infty} \Big[\underbrace{\rho(k+h)}_{k=h} - \underbrace{\rho(k-h)}_{k=h} - 2\rho(h)\rho(k)\Big]^2 = \Big[\rho(2h) - \underbrace{\rho(0)}_{1} - 2\rho^2(h)\Big]^2 = 1; \quad h \geq 1.$$

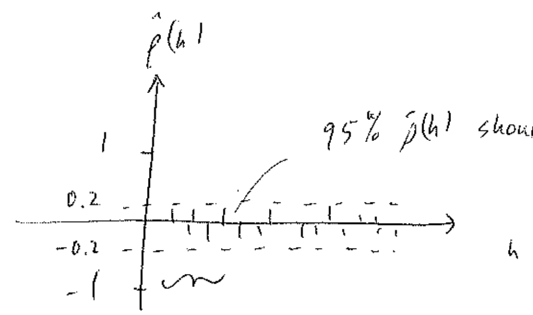
$$\Rightarrow \quad \hat{\rho}(h) \sim AN\left(\rho(h), \frac{1}{n}\right) \quad \forall\, h \geq 1.$$

$\Rightarrow$ 95% CI for $\rho(h)$: $\hat{\rho}(h) \pm 1.96\frac{1}{\sqrt{n}}$. $\Rightarrow$ approx. 95% of $\hat{\rho}(h)$ should fall within the bounds: $\pm \frac{1.96}{\sqrt{n}}$.

<u>Ex</u>/ Simulate 100 obs. from WN (0,1):

$$\pm \frac{1.96}{\sqrt{100}} \approx \pm 0.2.$$

↳ 95% $\hat{\rho}(h)$ should be within.



95% $\hat{\rho}(h)$ should be inside the significant bounds

interesting at the beginning.

stick out: correlated, cannot be WN.

<u>Note</u> : Process is trivially ergodic in $\bar{X}_n$ since $\lim_{h\to\infty} \rho(h) = 0$ ← actually after first lag.

<u>Ex</u> / MA(1) Model : $X_t = \theta Z_{t-1} + Z_t$, $\{Z_t\} \sim WN(0, \sigma^2)$

$$\Rightarrow \quad \gamma(h) = \begin{cases} \sigma^2(1+\theta^2) & , h = 0 \\ \theta\sigma^2 & , h = \pm 1 \\ 0 & , \text{o.w.} \end{cases} \quad \Rightarrow \quad \rho(h) = \begin{cases} 1 & , h = 0 \\ \dfrac{\theta}{1+\theta^2} & , h = \pm 1 \\ 0 & , \text{o.w.} \end{cases}$$

$$W_{11} = \sum_{k=1}^{\infty} \left[\rho(k+1) - \rho(k-1) - 2\rho(1)\rho(k)\right]^2 \qquad \text{for } k = 1, 2:$$

$$= \underbrace{\left[\rho(2) - \rho(0) - 2\rho^2(1)\right]^2}_{k=1} + \underbrace{\left[\rho(3) - \rho(1) - 2\rho(1)\rho(2)\right]^2}_{k=2}$$

$$= 1 - 3\rho^2(1) + 4\rho^4(1).$$

$$W_{22} = \left[\rho(3) + \rho(1)\right]^2 + \left[\rho(4) + \rho(0)\right]^2 + \left[\rho(5) + \rho(1)\right]^2 = 1 + 2\rho^2(1).$$

Also, $W_{hh} = 1 + 2\rho^2(1)$, $h \geq 2$.

Thus : 95% CI for $\rho(h)$: $\hat{\rho}(h) \pm 1.96 \sqrt{\dfrac{1 + 2\hat{\rho}^2(1)}{n}}$ $(h \geq 2)$

In practice, $\rho(1)$ is not known, so we use $\hat{\rho}(1)$ inside radical.

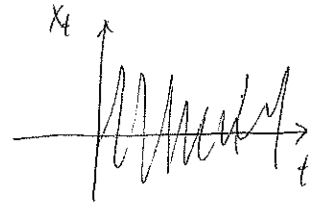
<u>Note</u> : Process ergodic in $\bar{X}_n$ since $\lim_{h\to\infty} \gamma(h) = 0$ (after 2 lags).

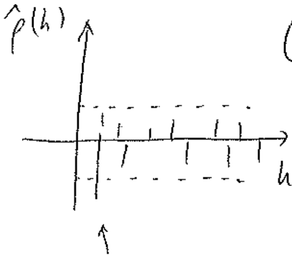
<u>Ex</u> / Simulate 200 obs. from MA(1) : $X_t = Z_t - 0.8 Z_{t-1}$, $Z_t \sim WN(0,1)$.

$$\rho(1) = \frac{\theta}{1+\theta^2} = \frac{-0.8}{1+0.64} = -0.4878 \leftarrow \hat{\rho}(1) = -0.4638.$$

$\gamma(0) = \sigma^2(1+\theta^2) = \sigma^2(1+0.64) = 1.64\,\sigma^2.$ ⑧

Realizations indistinguishable from WN until we look at ACF plot.





$\hat{\rho}(1) = -0.4638$

$$95\%\ CI \text{ for } \rho(1): \quad \hat{\rho}(1) \pm 1.96\sqrt{\frac{1-3\rho(1)^2-4\rho(1)^4}{n}}$$

$\hat{\rho}(1) = -0.4638$

n: 200

$$= -0.4638 \pm 0.0992 = (-0.5870, -0.3646)$$

capture true value $\rho(1) = -0.4878$.

95% CI for $\rho(2)$: 2 ways:

use $\rho(1)$

1) $\hat{\rho}(2) \pm 1.96\sqrt{\dfrac{1+2\rho(1)^2}{n}} = (-0.2617, 0.1745)$

$\hat{\rho}(2)$: -0.0436

use $\hat{\rho}(1)$

2) $-0.0436 \pm 1.96\sqrt{\dfrac{1+2(-0.4638)^2}{200}} = (-0.2093, 0.1221)$

$\boxed{1/30}$ AR(1) Model example.

<u>Ex</u>/ AR(1) Model. A stationary solution of equations: $X_t = \phi X_{t-1} + Z_t$, $\{Z_t\} \sim WN(0,\sigma^2)$

<u>Claim</u>: $EX_t = 0$. Check: $EX_t = \phi EX_{t-1} + EZ_t \Rightarrow \mu = \phi\mu + 0 \Rightarrow \mu = 0.$

↳ stationarity assumption ↳ $\mu(1-\phi) = 0$

$\Rightarrow EX_t = 0$ if $\{X_t\}$ is stationary.

<u>Can show</u>:

- $\gamma(h) = \dfrac{\sigma^2\phi^{|h|}}{1-\phi^2} \Rightarrow \rho(h) = \phi^{|h|}$, if $|\phi| < 1$

  ↳ condition for stationary.

- $\alpha(h) = \begin{cases} 1, & h=0 \\ \phi, & h=\pm 1 \\ 0, & |h|>1 \end{cases}$

  ↓

  PACF

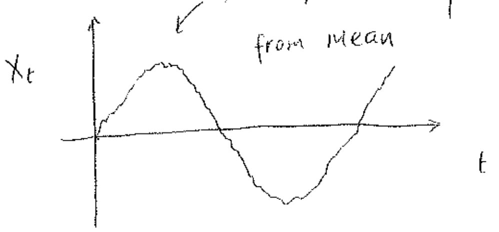
- $W_{hh} = \dfrac{(1-\phi^{2h})(1+\phi^2)}{1-\phi^2} - 2h\phi^{2h}$ if $|\phi|<1$.

Simulate $n=100$ from AR(1) with $\phi = \pm 0.9$: $\swarrow \rho(h) = \phi^{|h|}$

(both highly correlated)

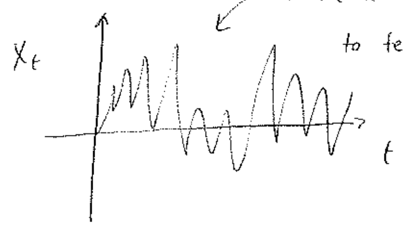
$\phi = 0.9$

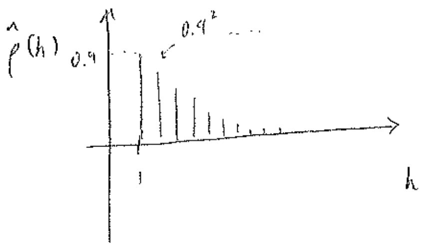
smooth, not too far from mean

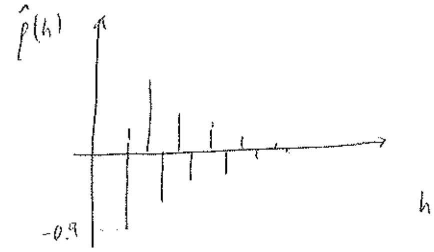


$\phi = -0.9$

white-noise like, hard to tell without ACF.



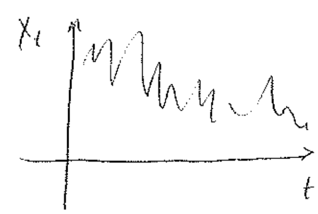




→ Figure 2.2 (B&D) gives $\rho(h)$ and $\hat\rho(h)$ for an AR(1) model: ($n=98$)

$$\underset{2:n}{X_t} = 0.791 \underset{1:(n-1)}{X_{t-1}} + Z_t, \quad Z_t \sim WN(0,\sigma^2)$$

fitted to the lake Huron data. (obtained via LS),

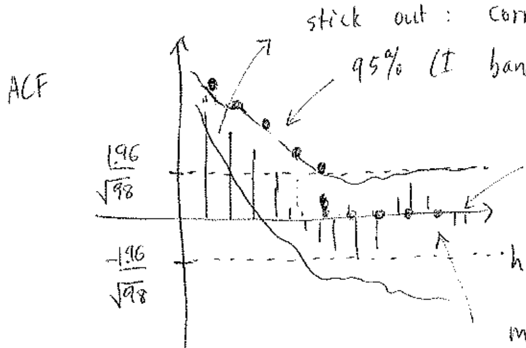


along with 95% confidence bounds computed from:

$$\hat\rho(h) \pm 1.96\sqrt{\frac{W_{hh}}{n}}$$

stick out: correlations

95% CI bands



sample ACF ($\hat\rho(h)$)

model ACF ($\rho(h)$)

(some outside, did not show)

↓

CI bounds for WN!

Note: $\frac{5}{40}$ (12.5%) of $\rho(h)$ fell outside bounds, suggests the model is inadequate!

- In practice, we just use WN CI bounds: $\pm \dfrac{1.96}{\sqrt{n}}$.

Note: $\gamma(h)$ of AR(1):

conv. in MS if $|\phi|<1$.

$$X_t = \phi X_{t-1} + Z_t = \phi(\phi X_{t-2} + Z_{t-1}) + Z_t = \dots = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

$$X_t X_{t+h} = \left(\sum_{i=0}^{\infty} \phi^i Z_{t-i}\right)\left(\sum_{j=0}^{\infty} \phi^j Z_{t+h-j}\right) = \sum_{i=1}^{\infty} \phi^i \phi^{i+h} Z_{t-i}^2 + \underbrace{\text{cross terms}}_{0}$$

$$t-i = t+h-j \Rightarrow j = i+h.$$

$$\Rightarrow \quad E(X_t X_{t+h}) = \gamma(h) = \sigma^2 \phi^h \sum_{i=1}^{\infty} \phi^{2i} = \frac{\sigma^2 \phi^h}{1-\phi^2}, \quad \text{if } |\phi|<1.$$

- Duality btw model & sample world:

Population Model world: $\{X_t\} \to \mu, \gamma(h), \rho(h), \alpha(h)$ — PMW

Random sample world: $\{X_1, \dots, X_n\}, \bar{X}_n, \hat{\gamma}(h), \hat{\rho}(h), \hat{\alpha}(h)$ — RSW

Observed Sample world: $\{x_1, \dots, x_n\} \to \bar{x}, \hat{\gamma}(h), \hat{\rho}(h), \hat{\alpha}(h)$ — OSW

- With theory developed in RSW, we use information gathered in OSW to make inference about PMW.

CH 2: Estimation & Removal of Trends & Seasonality.

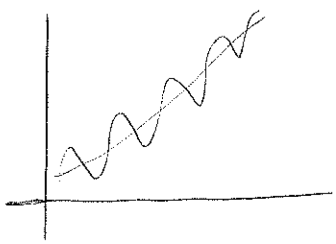
- Classical decomposition of a t.s: (capital: random things).

$$X_t = m_t + s_t + Y_t$$

- m_t = trend component (deterministic, not random, changes slowly with time).
- s_t = seasonal component (deterministic, periodic of period d.).
- Y_t = noise component (random, stationary).

Aim : Extract (model ?) components m_t and s_t, and hope that the residuals Y_t will be stationary. $Y_t = X_t - m_t - s_t$.

- We may need preliminary transformation if the amplitude of series appears to be changing over time (eg. log transform the Air Pass data stabilizes the variance).



$\leftarrow$ variance is stabilized ($X_t = m_t + s_t + Y_t$ now)

- Box-Cox transformation.

$\boxed{02/01}$ Classical Decomposition.

$$X_t = \underset{\text{trend}}{\underset{\downarrow}{m_t}} + \underset{\text{seasonality}}{\underset{\downarrow}{s_t}} + \underset{\text{noise}}{\underset{\searrow}{Y_t}}$$

Note : Modeling m_t and s_t maybe the most important thing !

(*) $E(Y_t) = 0$

Goal : Model & extract m_t and s_t, leaving residuals Y_t.

§2.1. Estimation & Elimination of Trend.

Nonseasonal model with trend : $X_t = m_t + Y_t$, $E(Y_t) = 0$.

a) Moving average smoothing :

Consider MA filter : $$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t+j} = \sum_{j=-q}^{q} \theta_j X_{t+j}$$

If m_t approx. linear on $[t-q, t+q]$, we get :

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} (m_{t+j} + Y_{t+j}) = \underbrace{\frac{1}{2q+1} \sum_{j=-q}^{q} m_{t+j}}_{m_t} + \underbrace{\frac{1}{2q+1} \sum_{j=-q}^{q} Y_{t+j}}_{0 = E(X_t)} \approx m_t.$$

$\downarrow$ linear : $m_t = a_0 + a_1 t$

$\hookrightarrow$ average out

- Process akin to appling linear filter:

$$\{X_t\} \longrightarrow \boxed{\text{MA filter}} \longrightarrow \hat{m}_t = \sum \theta_j X_{t+j}$$

Choosing q too small (large) results in a series that is too rough (smooth).

- If trend is a polynomial with degree ≥ 2, other choices for the $\{\theta_j\}$ can also be used which are optimal (hwk 3).

b) Exponential Smoothing:

Set $\hat{m}_1 = X_1$, and $\hat{m}_t = \alpha X_t + (1-\alpha)\hat{m}_{t-1}$, $t = 2, \ldots, n$

↳ $\alpha = 1 \Rightarrow$ unchange $\leftarrow$ correspond to $q = 0$
$\hat{m}_t = \hat{m}_{t-1}$ if $\alpha = 0$ $\leftarrow$ $q = \infty$.

} Max / min smoothing obtained with $\alpha = 0$ ($\alpha = 1$).

c) Spectral Smoothing:

Smoothes by eliminating a fraction, $1-f$, of the high frequency components of a Fourier series expansion of X_t (later). Max (min) smoothing obtained with $f = 0$ (or $f = 1$).

d) Polynomial Fitting:

Propose an appropriate polynomial for m_t: $m_t = a_0 + a_1 t + \ldots + a_k t^k$, and estimate coefficients with least squares.

e) Differencing k times to eliminate trend:

Define the backwards shift operator B as:

$$B X_t = X_{t-1}$$

Instead of removing the trend by smoothing, we can remove it by differencing, e.g.

$$(1-B) X_t = X_t - BX_t = X_t - X_{t-1} \quad \text{(differencing at lag 1)}.$$

$(1-B)^2 X_t = (1-2B+B^2) X_t = X_t - 2X_{t-1} + X_{t-2}$ $\qquad$ ※ $B^k X_t = X_{t-k}$.

↳ differencing twice at lag 1.

Can show that a polynomial trend of degree k will be reduced to a constant by differencing k times, i.e. by applying operator: $(1-B)^k X_t$.

Given a t.s. X_t we could proceed by differencing repeatedly (at lag 1), until the resulting series looks stationary.

$$(1-B) X_t \to \text{stationary} < \begin{array}{l} \text{Yes} \to \text{stop} \\ \text{No} \to (1-B)^2 X_t < \begin{array}{l} Y \\ N \end{array} \quad \cdots \end{array}$$

§2.2. Estimation / Removal of Seasonality.

Seasonal with seasonality only:

$$X_t = s_t + Y_t, \quad EY_t = 0, \quad s_{t+d} = s_t$$

$$\Rightarrow \sum_{t=1}^{d} s_t = 0.$$

a) Classical Decomposition:

The s_t, $t = 1, \ldots, d$ are estimated using the respective deviations from mean for each season. Deseasonalize by subtracting off the estimated s_t: $X_t - \hat{s}_t = Y_t$.

b) Harmonic Regression:

Use Fourier series representation to model:

$$s_t = \alpha_0 + \sum_{k=1}^{h} \left[\alpha_k \cos(\lambda_k t) + \beta_k \sin(\lambda_k t) \right]$$

where α_k and β_k are unknown coefficients and the λ_k are fixed frequencies, usually multiples of $\frac{2\pi}{d}$.

If n is a multiple of d, use $\lambda_k = \frac{2\pi m_k}{d}$, $m_k = \frac{nk}{d}$ (Fourier index)

will eliminate seasonal components with period $\frac{d}{k}$ (usually choose 1^{st} h harmonics, $k = 1, \dots h$).

c) Differencing: at lag d to remove period d:

Since $(1-B^d) s_t = s_t - s_{t-d} = 0$ ($\because s_t = s_{t+d}$). Differencing at lag d will eliminate a seasonality of period d.

Note: We can plot a periodogram (a spectral density estimator) to identify the dominant periodicities present in $\{X_t\}$ (later).

$\boxed{02/04}$

$$X_t = m_t + s_t + Y_t$$

§ 2.3. Estimation of Trend & Seasonality.

Combine methods in § 2.1 & § 2.2. The most common procedures are:

a) Classical decomposition:

Use smoothing like 2.1 (a) and 2.2 (a) and 2.1 (d) in order respectively.

Ex/ "death" data with $d=12$ plus quadratic trend.

b) Differencing:

Use 2.2(c) and 2.1 (e), in turn.

Ex/ "death" data, apply operator $\underbrace{(1-B)}_{\text{trend}}\underbrace{(1-B^{12})}_{\text{seasonal}} X_t \Rightarrow$ stationary, hope WN.

↳ go to remove obvious feature first: either trend or seasonality.

c) Smoothing (S & S) (model fitting)

A variety of techniques can also be used.

i) Moving average smoothers : the simple MA smoother :

$$w_t = \sum_{j=-q}^{q} \theta_j X_{t+j} \quad , \quad \theta_j = \frac{1}{2q+1}$$ ← default choice (average points)

can also be used to approx. both trends & seasonality at ≠ time scales, by judicious choice of q (and also possibly θ_j).

Ex/ smooth the weekly cardiovascular mortality series.

- $q = 2$ (5-pt MA), to approx. seasonality.
- $q = 26$ (53-pt MA), brings out the trend (yearly).

ii) Combination of polynomial & harmonic regressions :

$$X_t = m_t + s_t + Y_t$$

$$m_t = a_0 + a_1 t + \dots + a_p t^p$$

$$s_t = \alpha_0 + \sum_{k=1}^{h} \left[\alpha_k \cos\left(\frac{2\pi k}{d} t\right) + \beta_k \sin\left(\frac{2\pi k}{d} t\right) \right]$$

Estimate parameters via least-squares (similar to classical decomp.)

Ex/ smooth cardio data with $p = 3$, $h = 1$, $d = 52$.

iii) Kernel smoothing :

"Clever" MA smoother, uses a kernel function, $\theta_t(j; b)$ instead of constant θ_j.
Amount of smoothing is controlled by the bandwidth parameter b.

Ex/ Smooth cardio data using the Naradaya-Watson normal kernel :

$$\theta_t(j; b) = \frac{\phi\left(\frac{t-j}{b}\right)}{\sum_{i=1}^{n} \phi\left(\frac{t-i}{b}\right)} \quad , \quad \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad , \quad N(0,1) \text{ pdf.}$$

We then have: $W_t = \sum_{j=1}^{n} \theta_t(j; b) X_{t+j}$ ← on continuous scale

Ex/ smooth cardio data using the above, with $b = 2$ (for trend) and $b = \frac{10}{52}$ (for seasonality)

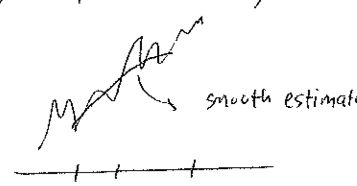
iv) <u>Lowess</u>:

A complex method using combination of locally weighted least-squares and nearest neighbors regression. The larger fraction of nearest neighbors that are included, the smoother the result.

Ex/ Smooth cardio data with a 2% and 67% fraction of nearest neighbors.
2% ↓ seasonality
67% ↓ trend

v) <u>Smoothing splines</u>:

Divide time $1, \dots, n$ into k intervals $[t_0 = 1, t_1), [t_1 + 1, t_2), \dots, [t_{k-1} + 1, t_k = n)$, where $t_0, \dots, t_k$ are called "knots".



Fit a polynomial regression over each interval (typically $p = 3 \Rightarrow$ cubic splines), and introduce a smoothing parameter λ.

Parameters are estimated by: $\text{argmin} \left\{ \sum_{t=1}^{n} (X_t - m_t)^2 + \lambda \int (m_t'')^2 \, dt \right\}$,

$m_t = \beta_0^{(i)} + \beta_1^{(i)} t + \dots + \beta_p^{(i)} t^p$, $\beta_j^{(i)}$ = coeff. over i^{th} interval, $i = 1, \dots, k$

Total # parameters: $k(p+1)$.

Ex/ Smooth cardio data here with $\underset{\text{(rough)}}{\lambda = 0}$ and $\underset{\text{(smooth)}}{\lambda = 1}$.
$\lambda = 0$: seasonality, $\lambda = 1$: trend

$\lambda = 0$ ↓ just LE over an interval $\Rightarrow$ rough fit

$\lambda = 1$: $\int (m_t'')^2 \, dt$ = roughness measure $\Rightarrow \sum_{t=1}^{n} \int (m_t'')^2 \, dt$ penalizes roughness of curve $\Rightarrow$ a smoother one.

§ 2.4 Testing the Estimated Noise Sequence

Once all apparent deterministic components have been removed, we want to model the resulting (stationary) residuals $\{Y_t\}$.

Remark: We typically only "visually" assess Y_t for seasonality. (There are "unit-root" tests for checking if the underlying ARMA model has unit roots in the AR poly., but this is just one kind of non-stationarity ...).

If there is no evidence of dependence on the residuals (in Y_t), then we can treat them as IID, and no further modeling is needed.

$\boxed{02/06}$

Here we want to test $H_0: \{Y_t\} \sim \text{iid}$.

There are many methods for doing this.

a) Sample ACF: Check to see if $|\hat{\rho}(h)| < \frac{1.96}{\sqrt{n}}$ ↙ large n: $\hat{\rho}(h) \sim N\left(0, \frac{1}{n}\right)$.

and reject H_0 if more than 5% fall outside the bounds; $\pm \frac{1.96}{\sqrt{n}}$.

This is based on asymptotic normality results (under H_0):

$$\hat{\rho}(h) \sim AN\left(0, \frac{1}{n}\right) \Rightarrow \sqrt{n}\,\hat{\rho}(h) \sim N(0,1)$$

Note: One should be suspicious of outlying $\hat{\rho}(h)$ values, eg. $|\hat{\rho}(h)| > \frac{3}{\sqrt{n}}$, particularly for small h. (Strong correlations at small lags).

b) Nonparametric Test of Randomness:

There are a variety of procedures for testing if a sequence of numbers is truly random or unpredictable (WN). These are based on several principles that WN should adhere to, e.g. series should have approx. the right # of turning pts, (max & min), changes of sign (after mean correction), pairs of points where $Y_{t+h} > Y_t$, (rank based tests), etc... In most cases, it's possible to arrive at a p-value (usually based on an asymptotic normality results). Implementation in R:

- library("randtests").
- via function test in library("itsmr").

c) Order of minimum AIC / BIC ARMA Model:

Under H_0 the order of min AIC / BIC ARMA model should be ARMA ($p=0, q=0$).

d) Portmanteau Tests:

Based on fact that under H_0: $\sqrt{n}\,\hat{\rho}(h) \overset{iid}{\sim} N(0,1)$, for any sequence of integers $h=1,\dots,n$ so that:

$$Q(m) = n \sum_{h=1}^{m} \hat{\rho}^2(h) \sim \chi^2(m-g)$$

g = # of model coefficients (ARMA, ARIMA, SARIMA, regression with ARMA errors, ...) that were fitted to the data before arriving at $\{Y_t\}$.

Later we use g. For now, $g=0$ since no model has yet been fitted.

- There are several refinements of the test, the most popular is the Ljung-Box test, implemented in R, library("portes"), with the stated default:

LjungBox obj, lags = seq (5, 30, 5), order = 0, season = 1, squared.residuals = F).

where
- obj = the ts data (or a fitted model ts object).
- lags = values of m to compute $Q(m)$.
- order = value of g for df of asymptotic χ^2 dist^{n}. (automatically obtained from obj. if it's a ts model object).

- Season: period of seasonality (if we have reasons to believe seasonality was not property modeled / removed.
- sq. residuals: boolean, asking if series should be squared before doing test (for detecting if sq. series is autocorrelated, usually only of interest for financial).

If we reject H_0, there is evidence of correlation in $\{Y_t\}$. Rest of course devotes to.

CH 3. Autoregressive Moving Average (ARMA) Models.

Defn: The ts $\{X_t\}$ is a linear process if it has the following representation:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} , \quad Z_t \sim WN(0, \sigma^2).$$

where $\forall t$, $\{\psi_j\}$ is a seq. of constants with $\sum_{-\infty}^{\infty} |\psi_j| < \infty$. Can write compactly:

$$X_t = \psi(B) Z_t , \quad \psi(B) = \sum_{-\infty}^{\infty} \psi_j B^j .$$

Check: $X_t = \psi(B) Z_t = \sum_{-\infty}^{\infty} \psi_j B^j Z_t = \sum_{-\infty}^{\infty} \psi_j Z_{t-j}$.

linear processes provide a general framework for studying stationary processes. In fact, every stat. processes is either linear, or can be made linear by subtracting a deterministic component. (Wold Decomposition Thm).

The Wold Decomposition Theorem:

If $\{X_t\}$ is a nondeterministic stationary ts, then

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} + V_t , \quad \text{where}$$

- $\psi_0 = 1$ and $\sum_{-\infty}^{\infty} \psi_j^2 < \infty$
- $\{Z_t\} \sim WN(0, \sigma^2)$
- $Cov(Z_s, V_t) = 0 \quad \forall s,t$
- $\{V_t\}$ is deterministic (can be perfectly predicted from its past).

This result justifies modeling any stationary ts $\{X_t\}$ as a linear process!

2/8 Wold Decomposition Theory

Any stationary $\{X_t\}$ can be written as: $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t$

$\sum_{j=0}^{\infty} \psi_j^2 < \infty$; V_t: deterministic

Defn: linear process: $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

This justifies the ts modeling as linear process.

⊛ Properties of a linear process:

- Weakly stationary (strict stationary of $Z_t \sim$ iid).
- $EX_t = 0$
- $\gamma(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2$, $h \geq 0$

Ex/ MA(1): $X_t = \theta Z_{t-1} + Z_t \Leftarrow V_t = 0$, $\psi_0 = 1$, $\psi_1 = \theta$

↳ already in the Wold decomposition, $V_t = 0$, $\psi = \begin{cases} 1 & j=0 \\ \theta & j=1 \\ 0 & \text{o.w.} \end{cases}$

$\Rightarrow \gamma(0) = \sum_{j=0}^{\infty} \psi_j \psi_j \sigma^2 = \sigma^2(\underbrace{1^2}_{j=0} + \underbrace{\theta^2}_{j=1}) = \sigma^2(1+\theta^2)$

$\gamma(1) = \sum_{j=0}^{\infty} \psi_j \psi_{j+1} \sigma^2 = \theta\sigma^2$ $\Rightarrow \rho(h) = \begin{cases} 1 & h=0 \\ \dfrac{\theta}{1+\theta^2} & h=1 \\ 0 & \text{o.w.} \end{cases}$

$\gamma(h) = 0 \quad h \geq 2$

Ex/ AR(1): $X_t - \phi X_{t-1} = Z_t$, $|\phi| < 1$

Stationary, can be brought back to Wold decomposition.

Consider $X_t = \sum \phi^j X_{t-j}$ process. (?)

Now show this is a solution of the AR(1) eqn. This is a linear process

$\Rightarrow$ hence it is weakly stationary.

$$X_t - \phi X_{t-1} = \sum_{j=0}^{\infty} \phi^j Z_{t-j} - \phi \sum_{j=0}^{\infty} \phi^j Z_{t-j} = Z_t + \left(\sum_{j=0}^{\infty} \phi^j Z_{t-j} - \sum_{j=0}^{\infty} \phi^j Z_{t-j} \right) = Z_t .$$

Note : $\psi_j = \begin{cases} \phi^j & j \geq 0 \\ 0 & j < 0 \end{cases}$

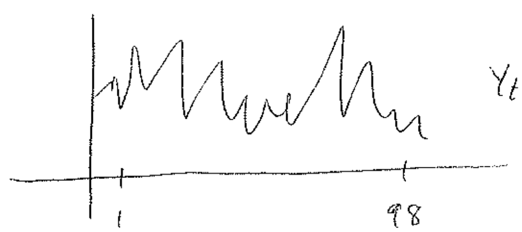
$$\Rightarrow \gamma(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2 = \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \sigma^2 = \sum_{j=0}^{\infty} \phi^j \phi^{j+h} \sigma^2 = \sum_{j=0}^{\infty} \phi^{2j+h} \sigma^2$$

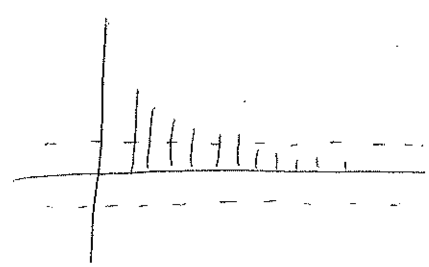
$$= \sigma^2 \phi^h \sum_{j=0}^{\infty} (\phi^2)^j = \sigma^2 \frac{\phi^h}{1-\phi^2}$$

$$\Rightarrow \rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h , \quad h \geq 0 \quad \Rightarrow \rho(h) = \phi^{|h|} \quad \forall h .$$

$$\Rightarrow \text{Also : } X_t = (1-\phi B)^{-1} Z_t = \sum_{j=0}^{\infty} (\phi B)^j Z_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

Ex/ Lake Heron





- Sample ACF / PACF suggests not WN. In fact, it is not even stationary.

It has slowly decaying trend.

- Fit LS lines to it : $Y_t = \beta_0 + \beta_1 t + X_t$, $\{X_t\} \sim WN(0, 1.251)$

$\hat{\beta}_0 = 10.202$

$\hat{\beta}_1 = -0.0242$

- 2 main features in the residuals of $\{X_t\}$: ① no trend ② smoothness (long stretches of obs. with same sign).

    ↳ very unlikely if data were WN.

- Smoothness suggests some form of dependence. Such dependence can be used to forecast. If WN model were a good fit, then the "best" predictor of X_{99} would be 0. However, plot of X_t suggests $X_{99} > 0$.

⊕ How do we quantify dependence ?

<u>Ans</u> : ACVF / ACF.

⊕ How do we construct models for forecasting that incorporates dependence ?

<u>Ans</u> : Stationary processes, esp. ARMA models, are a flexible family that exhibits a variety of correlation structures.

For AR(1), $\dfrac{\rho(h+1)}{\rho(h)} = \dfrac{\phi^{h+1}}{\phi^h} = \phi$.

The rough geometric decay of $\hat{\rho}(h)$, $h = 1, 2, 3, \ldots$ with $\dfrac{\hat{\rho}(h+1)}{\hat{\rho}(h)} \simeq 0.7$

$\Rightarrow$ suggest AR(1) with $\phi = 0.7$.

$$X_t = 0.7 X_{t-1} + Z_t \quad , \quad \gamma(0) = \frac{\sigma^2}{1-\phi^2} \Rightarrow \hat{\sigma}^2 \approx \gamma(0)(1 - 0.7^2)$$

X_t	X_{t-1}
X_2	X_1
X_3	X_2
$\vdots$	$\vdots$
X_n	X_{n-1}

Fit LS regression, we get : $X_t = 0.791 X_{t-1} + Z_t$

$$Z_t \sim WN(0, .713^2)$$

This gives a complete (rudimental) model for Y_t.

Better methods later ...

(*) ARMA models / processes : an important class of linear ts models. They provide a flexible parametric structure to approx. behavior of stationary process and lead to a prediction theory that is relatively simple and elegant. WLOG, assume $\{X_t\}$ has zero mean, since if $\{Y_t\}$ has μ as mean, $X_t = Y_t - \mu \Rightarrow 0$ mean.

This approach differs from S&S, who treat μ as another para. to be simultaneously estimated with the remaining parameters.

$\boxed{02/11}$ §3.1 The Autoregressive (AR) Process [Original $Y_t \rightarrow X_t = Y_t - \bar{x}$]

One of the most intuitive ways to construct a model for stationary $\{X_t\}$, with $EX_t = 0$. The AR (p) model is :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t \quad , \quad Z_t \sim WN\ (0, \sigma^2)$$

Main facts :

- Similar to linear regression, except we regress the series on its past values : $X_{t-1}, \dots, X_{t-p}$.
- Defining the AR polynomial :

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \quad , \quad z \in \mathbb{C}$$

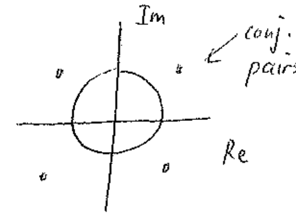
then we can write the AR(p) model as :

$$\phi(B) X_t = Z_t \quad , \quad \{Z_t\} \sim WN\ (0, \sigma^2)$$

$$\downarrow$$

$$(1 - \phi_1 B - \dots - \phi_p B^p)\ X_t = X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}$$

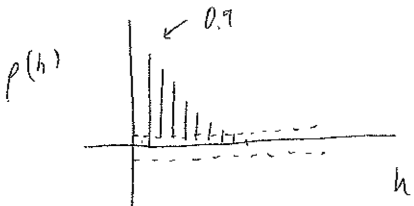
- To ensure $\{X_t\}$ satisfying the AR(p) eqn. is stationary & depends only on the past (is <u>causal</u>), all roots of $\phi(z)$ must be greater than 1 in magnitude. (all roots must lie outside of unit circle in complex plane).



- For an AR(p), the PACF at lag p, $\alpha(p) = \phi_p$, and $\alpha(h) = 0$ $\forall\ h > p$.
- The ACF of an AR(p) decays exponentially.

<u>Ex</u>/ Simulate from AR(1) with $\phi = 0.9$ and $\phi = -0.9$.

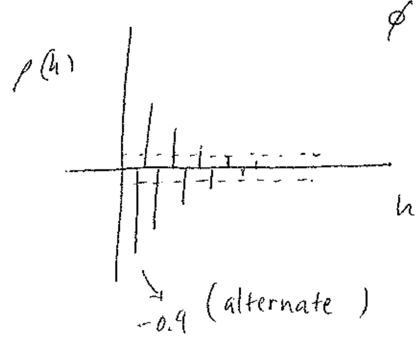
$\phi = 0.9$



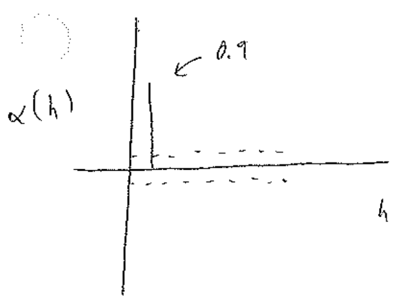
(decay exp.)

$\rho(h) = \phi^h$

$\phi = -0.9$



(alternate)

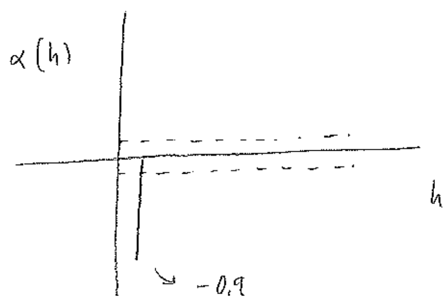


$$\alpha(h) = \begin{cases} \phi, & h = 1 \\ 0, & h > 1 \end{cases}$$

$\Downarrow$

PACF of AR(1)

$\equiv$ ACF of MA(1)



<u>Ex</u>/ Causality for AR(1): $X_t = \phi X_{t-1} + Z_t$

Iterate: $X_t = \phi(\phi X_{t-2} + Z_{t-1}) + Z_t$ ... In looking for a stationary s/n:

(k times) $\vdots$

$$= \sum_{j=0}^{\infty} \phi^j Z_{t-j} + \phi^{k+1} X_{t-k-1}$$

Can be shown $\{X_t\}$ converges in mean-square sense to:

$\sum_{j=0}^{\infty} \phi^j Z_{t-j}$, if $|\phi| < 1$. condition for causality.

AR poly: $\phi(z) = 1 - \phi z \Rightarrow$ root $= z = \frac{1}{\phi} \Rightarrow |\phi| < 1$ for roots to be outside unit circle.

- If $|\phi| > 1$: can write model as:

$$X_t = \phi X_{t-1} + Z_t \Rightarrow X_t = -\frac{1}{\phi} Z_{t+1} + \frac{1}{\phi} X_{t+1}$$

$\downarrow$

isolate X_{t-1} then change index

Iterate as before, can show with similar arguments that:

$$X_t = -\sum_{j=1}^{\infty} \phi^{-j} Z_{t+j}$$ (but need to know all future values $\Rightarrow$ unrealistic).

is the unique stationary solution of the AR(1) eqn.

$\Rightarrow$ This is unatural, since $\{X_t\}$ is future-dependent! It can be shown that every AR(1) with $|\phi| > 1$ can be re-expressed as AR(1) with $|\phi| < 1$, and a new WN sequence. Therefore, nothing is lost if we ignore AR(1) with $|\phi| > 1$.

↳ This way ensures the solution is future-independent or causal.

Thm: For an AR(p) to be causal, all roots of $\phi(z)$ must lie outside the unit circle (greater than 1 in magnitude).

Thm: For an AR(p) to be stationary, no roots of $\phi(z)$ can equal 1 in magnitude. (Roots must not lie on unit circle).

Summary:

Roots	Model
outside UC	causal sln (realistic & useful)
on UC	non-stationary
inside UC	non-causal sln (not realistic)

§ 3.2 The MA (q) : Moving Average of order q.

Analogously to the AR, the MA (q), regresses on the lagged values of $\{Z_t\}$:

$$X_t = \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} + Z_t, \quad \{Z_t\} \sim WN\ (0, \sigma^2)$$

Main Facts :

- Defining the MA polynomial :

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$$

can write MA (q) as: $X_t = \theta(B) Z_t$

- For parameter identifiability reasons, and in analogy with concept of causality of AR, we require all roots of $\theta(z)$ to be greater than 1 in magnitude. The resulting process is said to be invertible. (analogous to causal AR).

Ex/ Consider 2 MA (1) models :

$$X_t = \theta Z_{t-1} + Z_t \quad (1)$$

$$X_t = \frac{1}{\theta} Z_{t-1} + Z_t \quad (2)$$

$$\Rightarrow \rho_1(h) = \begin{cases} \dfrac{\theta}{1+\theta^2} & h = 1 \\ 0 & \text{o.w. } (h \geq 2) \end{cases}$$

$\updownarrow$ same

$$\rho_2(h) = \begin{cases} \dfrac{\theta}{1+\theta^2} & h = 1 \\ 0 & h \geq 2 \end{cases}$$

identifiability problem! (indistinguishable from 1^{st} and 2^{nd} order moments).

But lead to $\neq$ realization.

Soln : restrict $|\theta| < 1 \Rightarrow$ roots of MA (1) polynomial : $z = \dfrac{1}{\theta} > 1$.

- The PACF of MA(q) decays exponentially.
- The ACF of an MA(q) is zero beyond lag q.

Note: Have duality on $\rho(h)$ / $\alpha(h)$ for AR(p) / MA(q).

	AR(p)	MA(q)
$\rho(h)$	decays exp.	zero for $h > q$
$\alpha(h)$	zero for $h > p$	decays exp.

$\boxed{02/13}$ AR(p) & MA(q)

§3.3 The Autoregressive Moving Average (ARMA) Process

Idea: combine AR(p) and MA(q) to form ARMA(p,q):

$$\underbrace{X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}}_{\phi(B) X_t} = \underbrace{Z_t + \theta Z_{t-1} + \dots + \theta_q Z_{t-q}}_{\theta(B) Z_t} \quad ; \quad W_t \sim WN(0, \sigma^2)$$

Main Facts

↳ causal if all roots outside UC ↳ invertible if all roots outside UC

- By defn. we require $\{X_t\}$ to be stationary. Using compact AR & MA polynomials notation, we can write:

$$\phi(B) X_t = \theta(B) Z_t, \quad Z_t \sim WN(0, \sigma^2)$$

where $\phi(z)$ and $\theta(z)$ have no common factors.

- For causality & invertibility, we require as before all roots of $\phi(z)$ and $\theta(z)$ be greater than 1 in magnitude.
- ARMA(p,0) = AR(p)

  ARMA(0,q) = MA(q)
- ACF & PACF both decay exponentially, in general.

Note: Although we can model almost any correlation structure with an AR (p), when p is large, a relatively low order ARMA (p, q) can suffice to approximate the correlation sturucture (parsimonious → speed sufficient & easy to estimate parameters).

Notes on § 3.1 – § 3.3 .

Defn: (Causality) Solution $\{X_t\}$ is causal if we can write :

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} = \psi(B) Z_t \quad , \text{ with } \sum_{j=0}^{\infty} |\psi_j| < \infty \qquad (MA(\infty) \text{ representation of } X_t).$$

Fact: $\{X_t\}$ is causal $\iff$ $\phi(z) \neq 0$ for $|z| \leq 1$. (all roots of $\phi(z)$ outside U.C.).

Since $X_t = \psi(B) Z_t$, if it's causal , and $\phi(B) X_t = \theta(B) Z_t$, since $X_t \sim ARMA(p,q)$

$$\Rightarrow \boxed{\psi(B) = \frac{\theta(B)}{\phi(B)}} \longrightarrow \text{ use to compute } \{\psi_j\} .$$

Defn: (Invertibility) $\{X_t\}$ is invertible if we can write :

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = \pi(B) X_t \quad , \text{ with } \sum_{j=0}^{\infty} |\pi_j| < \infty \qquad (AR(\infty) \text{ representation of } X_t)$$

Fact: $\{X_t\}$ invertible $\iff$ $\theta(z) \neq 0$ for $|z| \leq 1$ (all roots of $\theta(z)$ outside U.C.)

Since $Z_t = \pi(B) X_t$, if it's invertible , and $\phi(B) X_t = \theta(B) Z_t$, since $X_t \sim ARMA(p,q)$

$$\Rightarrow \boxed{\pi(B) = \frac{\phi(B)}{\theta(B)}} \rightarrow \text{ use to compute } \pi_j .$$

Thm: Every noncausal and/or noninvertible ARMA (p, q) process can be reexpressed as a causal & invertible process, so no generality is lost by focusing only on these processes.

(*) The ACVF of an ARMA (p,q):

$\{X_t\}$ a causal ARMA (p,q): $\phi(B)X_t = \theta(B)Z_t$, $Z_t \sim WN(0,\sigma^2)$

$$\Rightarrow \quad X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad \psi(B) = \frac{\theta(B)}{\phi(B)}$$

First Method: Since $\{X_t\}$ is a linear process, $\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$ $\forall h$.

Compute the ψ_j's directly from $\psi(z) = \dfrac{\theta(z)}{\phi(z)}$ or cross-multiply to get:

$$(1-\phi_1 z - \dots - \phi_p z^p)(\psi_0 + \psi_1 z + \dots) = 1 + \theta_1 z + \dots + \theta_q z^q,$$

and equate coefficients of z^j to get a recursive formula:

$$(*) \quad \psi_j = \theta_j + \sum_{k=1}^{p} \phi_k \psi_{j-k}$$

with boundary conditions:
$$\begin{cases} j = 0, 1, \dots \\ \theta_0 = 1, \ \theta_j = 0 \ \forall j > q \\ \text{and } \psi_j = 0, \ j < 0 \end{cases}$$

Ex/ ARMA $(1,1)$: $X_t - 0.5X_{t-1} = Z_t + 0.4Z_{t-1}$

$$\left.\begin{array}{l} \phi(z) = 1 - 0.5z = 0 \Rightarrow z = 2 \\ \theta(z) = 1 + 0.4z = 0 \Rightarrow z = -2.5 \end{array}\right\} \text{outside U.C.} \Rightarrow \begin{array}{l} \text{causal} \\ \text{invertible} \end{array}$$

For a general ARMA $(1,1)$:

$$\psi(z) = \frac{\theta(z)}{\phi(z)} = \frac{1+\theta z}{1-\phi z} = (1+\theta z)\sum_{j=0}^{\infty} \phi^j z^j$$

$$= 1 + (\theta+\phi)z + \phi(\theta+\phi)z^2 + \dots$$

$$\Rightarrow \quad \psi_0 = 1, \quad \psi_1 = \theta + \phi, \quad \dots, \quad \psi_j = \phi^{j-1}(\theta+\phi) \quad j \geq 1.$$

$\Rightarrow \quad \gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}$. Cases :

- $\gamma(0) = \sigma^2 \left[1 + (\theta+\phi)^2 + \phi^2 (\theta+\phi)^2 + \phi^4 (\theta+\phi)^2 + \dots \right]$

$$= \sigma^2 \left[1 + (\theta+\phi)^2 \sum_{j=0}^{\infty} (\phi^2)^j \right] = \sigma^2 \left(1 + \frac{(\theta+\phi)^2}{1-\phi^2} \right)$$

- $\gamma(1) = \sigma^2 \left[(\theta+\phi) + (\theta+\phi)^2 \phi \sum_{j=0}^{\infty} \phi^{2j} \right] = \sigma^2 \left[(\theta+\phi) + \frac{(\theta+\phi)^2 \phi}{1-\phi^2} \right]$

and $\gamma(h) = \phi^{h-1} \gamma(1)$, $\quad h \geq 2$.

Also : $\quad \pi_j = (-\theta)^{j-1} (-\phi - \theta)$, $\quad j > 0$.

Second Method : Solve Yule - Walker eqn. (YW)

Write : $\quad X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta Z_{t-1} + \dots + \theta_q Z_{t-q}$

Take expectation of both sides with X_{t-k} to get :

(1) $\quad \gamma(k) - \phi \gamma(k-1) - \dots - \phi_p \gamma(k-p) = \sigma^2 \sum_{j=0}^{\infty} \theta_{k+j} \psi_j$, $\quad 0 \leq k < m \quad \leftarrow \{\alpha_1, \dots, \alpha_p\}$

(2) $\quad \gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) = 0$, $\quad k \geq m$

$\hookrightarrow$ linear difference eqn.

where $m = \max(p, q+1)$

Solve for $\gamma(0), \dots, \gamma(p)$, then use recursions in (2) to get $\gamma(p+1), \gamma(p+2), \dots$

$\boxed{2/15}$ Computing ACVF

- Method 1 : $\quad \psi(z) = \frac{\theta(z)}{\phi(z)} \Rightarrow$ equate coefficients $\Rightarrow \gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$

- Method 2 : Yule - Walker eqn.

Alternatively, general solution is :

$\gamma(h) = \alpha_1 \xi_1^{-k} + \dots + \alpha_p \xi_p^{-k}$, $k \geq m-p$, $m = \max(p, q+1)$

↳ ARMA (p,q). $\Rightarrow$ $\{\xi_1, \dots \xi_p\}$ are the p distinct roots of $\phi(z) = 0$,

and $\{\alpha_1, \dots, \alpha_p\}$ are arbitrary constants to be determined by (1) of YW.

Ex/ AR(2) : $X_t - 0.7 X_{t-1} + 0.1 X_{t-2} = Z_t$, $Z_t \sim WN(0, \sigma^2)$

$\Rightarrow$ $\phi(z) = 1 - 0.7z + 0.1z^2 = (1-0.5z)(1-0.2z)$ $\Rightarrow$ $z_1 = 2$, $z_2 = 5$ $\Rightarrow$ causal.

- Invertible : no root, but fn is AR form $\Rightarrow$ yes.

  ↳ AR(p) are always invertible !

$$Z_t = X_t - 0.7 X_{t-1} + 0.1 X_{t-2} = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad \Rightarrow \pi_j = \begin{cases} 1 & , j=0 \\ -0.7 & , j=1 \\ 0.1 & , j=2 \\ 0 & , \text{o.w.} \end{cases}$$

(*) eqn : $\theta_j = \psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k}$ ← Method 1.

From here, $\psi_0 = 1$

$$\theta_1 = 0 = \psi_1 - \phi_1 \psi_0 \Rightarrow \psi_1 = \theta_1 + \phi_1 = 0 + 0.7 = 0.7$$

$$\theta_2 = 0 = \psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0 \Rightarrow \psi_2 = \phi_1 \psi_1 + \phi_2 \psi_0 = 0.7^2 - 0.1 = 0.39$$

$\vdots$

$\Rightarrow$ $\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h} = \dots$ decays very fast !

The PACF.

- Fact : PACF of stationary process $\{X_t\}$ is given by the function $\alpha(h)$, defined as

$$\alpha(h) = \begin{cases} 1 & , h = 0 \\ \phi_{hh} & , h \geq 1 \end{cases}$$

where ϕ_{hh} is coefficient of X_1 in predicting X_{h+1} in terms of $\{X_2, \dots, X_h\}$ i.e.

prediction operator $\rightarrow$ $P_n X_{n+1} = \phi_{n1} X_n + \phi_{n2} X_{n-1} + \dots + \phi_{nn} X_1 \equiv P(X_{n+1} \mid X_1, \dots, X_n) \leftarrow$ {BLP}

Predictor

= Best (min MSE) Linear (linear function of values basing prediction on) Predictor of X_{n+1} in terms of $\{X_1, \dots, X_n\}$.

= BLP of X_{n+1} in terms of $\{X_1, \dots, X_n\}$.

The $\{\phi_{n1}, \dots, \phi_{nn}\}$ are obtained from

$$\underset{n\times 1}{\begin{pmatrix} \phi_{n1} \\ \vdots \\ \phi_{nn} \end{pmatrix}} = \underset{n\times n}{\begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ & \gamma(0) & \ddots & \vdots \\ \underset{\text{symmetric}}{*} & & \ddots & \gamma(1) \\ & & & \gamma(0) \end{pmatrix}^{-1}} \underset{n\times 1}{\begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{pmatrix}} \equiv \Gamma_n^{-1} \underset{\sim}{\gamma}_n$$

$$\equiv \Gamma_n^{-1} \underset{\sim}{\gamma}_p \left(\frac{\gamma(0)}{\gamma(0)}\right) \leftarrow \text{help to change ACVF} \rightarrow \text{ACF}$$

$$\Rightarrow \begin{pmatrix} \phi_{n1} \\ \vdots \\ \phi_{nn} \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \cdots & \rho(n-1) \\ & 1 & \ddots & \vdots \\ * & & \ddots & \rho(1) \\ & & & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(n) \end{pmatrix} \equiv R_n^{-1} \underset{\sim}{\rho}_n \qquad (*)$$

- Interpretation: $\alpha(n)$ is the correlation between X_1 and X_{n+1} adjusted for $\{X_2, \dots, X_n\}$

(correlation btw the prediction errors):

$$X_{n+1} - P(X_{n+1} \mid X_2, \dots, X_n) \ \underbrace{\&}_{\text{corr. btw.} \equiv \alpha(n)} \ X_1 - P(X_1 \mid X_2, \dots, X_n) \Big).$$

↳ easy definition.

- Fact :

Fact : For an AR(p), $\alpha(p) = \phi_p$ and $\alpha(h) = 0 \quad \forall \; h > p$.

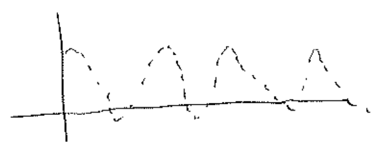
Use $\hat{\rho}(\cdot)$ in ⊛ to estimate sample PACF $\hat{\alpha}(\cdot)$.

Use significant bounds $\pm \frac{1.96}{\sqrt{n}}$ to detect if $\hat{\alpha}(h)$ is significant.

Thm : If $\{X_t\}$ is an AR(p) with $Z_t \sim \text{iid}(0, \sigma^2)$ then for $h > p$ we have asymptotically :
$$\hat{\alpha}(h) = \hat{\phi}_{hh} \sim N\left(0, \tfrac{1}{n}\right).$$

Ex/ Sunspots in "itsmr" ← subset of "sunspot.year" in R.

- Number of spots on Sun recorded 1770 - 1869 $(n = 100)$: $\{Y_1, \ldots, Y_{100}\}$.
- Series looks stationary (assumed)



← look like cycle but not well-defined.

↳ $S_t = 0, \; m_t = 0$.

- Sample estimates : $\bar{y} = 46.93$, $\hat{\gamma}(h) = \begin{cases} 1382.2, & h = 0 \\ 1114.4, & h = 1 \\ 591.73, & h = 2 \end{cases}$

$$\Rightarrow \hat{\rho}(h) = \begin{cases} 0.8062, & h = 1 \\ 0.4281, & h = 2 \end{cases} \qquad \pm \frac{1.96}{\sqrt{n}} \approx \pm \frac{2}{\sqrt{n}} = \pm 0.2$$

→ significant.

- Sample PACF : $\hat{\alpha}(1) = \phi_{11} = \hat{\rho}(1) = 0.8062$. → significant

$\hat{\alpha}(2) = \phi_{22} = ?$

↓

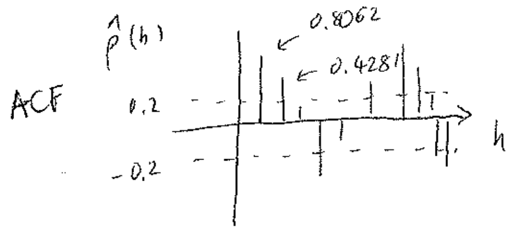
Model (⊛) : $\phi_{11} = 1^{-1} \cdot \rho(1) = \rho(1)$ ↓

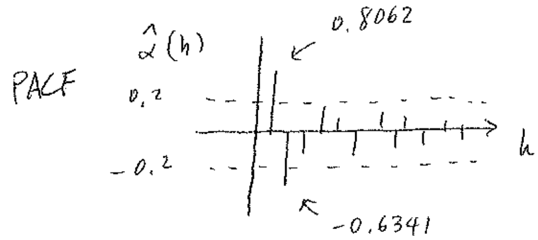
$$\begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix} = \frac{1}{1 - \rho^2(1)} \begin{pmatrix} 1 & -\rho(1) \\ -\rho(1) & 1 \end{pmatrix} \begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix}$$

$$\Rightarrow \phi_{22} = \frac{1}{1 - \rho^2(1)} \left(-\rho^2(1) + \rho^2(2) \right) \Rightarrow \hat{\alpha}(2) = \frac{-\hat{\rho}^2(1) + \hat{\rho}^2(2)}{1 - \hat{\rho}^2(1)} = -0.6341$$

→ significant

$\boxed{2/18}$





$\Downarrow$ ($\hat{\rho}(h)$ & $\hat{\alpha}(h)$)

Suggests AR (2) is a plausible model.

Duality :

	AR(p)	MA(q)
ACF	decays exp.	zero for $h > q$
PACF	zero for $h > p$	decays exp.

- How to estimate parameters ? ($\mu, \phi_1, \phi_2, \sigma^2$)

$$\underbrace{Y_t - \mu}_{X_t} = \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{t-2} - \mu) + Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2).$$

or :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2)$$

- Obviously use $\hat{\mu} = \bar{y} = 46.93$

Recall YW eqn for AR(2) :

k=0 : X_t

$$X_t X_t - \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t = Z_t X_t$$

$$E(X_t X_t - \phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t) = E(Z_t X_t)$$

$$\gamma(0) - \phi_1 \gamma(1) + \phi_2 \gamma(2) = \sigma^2$$

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

$$X_t Z_t = Z_t^2 + Z_t Z_{t-j}$$

$$\hookrightarrow E(Z_t X_t) = E Z_t^2 = \sigma^2$$

k=1 : X_{t-1}

$$E(X_t X_{t-1} - \phi_1 X_{t-1} X_{t-1} + \phi_2 X_{t-2} X_{t-1}) = E(Z_t X_{t-1})$$

$$\gamma(1) - \phi_1 \gamma(0) + \phi_2 \gamma(1) = 0$$

⋮

k=2 : X_{t-2} "

YW eqn :

$$\gamma(0) - \phi_1 \gamma(1) + \phi_2 \gamma(2) = \sigma^2$$

$$\gamma(1) - \phi_1 \gamma(0) - \phi_2 \gamma(1) = 0$$

$$\gamma(2) - \phi_1 \gamma(1) - \phi_2 \gamma(0) = 0$$

Usually, solve for $\gamma(\cdot)$ given the model parameters, but we can also reverse this by using $\hat{\gamma}(\cdot)$ and solving the parameters.

① & ③ $$\begin{pmatrix}\gamma(1)\\ \gamma(2)\end{pmatrix} = \begin{pmatrix}\gamma(0) & \gamma(1)\\ \gamma(1) & \gamma(0)\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}$$

$$\Rightarrow \begin{pmatrix}\hat\phi_1\\ \hat\phi_2\end{pmatrix} = \begin{pmatrix}\hat\gamma(0) & \hat\gamma(1)\\ \hat\gamma(1) & \hat\gamma(0)\end{pmatrix}^{-1}\begin{pmatrix}\hat\gamma(1)\\ \hat\gamma(2)\end{pmatrix} = \begin{pmatrix}1 & \hat\rho(1)\\ \hat\rho(1) & 1\end{pmatrix}^{-1}\begin{pmatrix}\hat\rho(1)\\ \hat\rho(2)\end{pmatrix} = \begin{pmatrix}1.318\\ -0.634\end{pmatrix}$$

Plug these values into ① to get:

$$\hat\sigma^2 = \hat\gamma(0) - \hat\phi_1\hat\gamma(1) - \hat\phi_2\hat\gamma(2) = 289.2$$

By doing this, we are forcing the fitted model ACVF ~~by~~ to agree with sample ACVF at first 3 lags i.e. $\gamma(h) = \hat\gamma(h)$, $h = 0, 1, 2$. (MOME).

Fitted model: $X_t - 1.318X_{t-1} + 0.634X_{t-2} = Z_t$, $\{Z_t\} \sim WN(0, 289.2)$

Final model: $(Y_t - 46.83) - 1.318(Y_{t-1} - 46.83) + 0.634(Y_{t-2} - 46.83) = Z_t$

$$\Rightarrow Y_t - 1.318Y_{t-1} + 0.634Y_{t-2} = 14.83 + Z_t, \quad \{Z_t\} \sim WN(0, 289.2)$$

↳ not the mean!

- Model & sample PACF:

$\alpha(1) = \phi_{11} = (1)^{-1}\rho(1) = \rho(1) \Rightarrow \hat\alpha(1) = \hat\rho(1) = 0.8062 = \alpha(1)$

$\alpha(2) = \phi_{22}$; where $\begin{pmatrix}\phi_{21}\\ \phi_{22}\end{pmatrix} = \begin{pmatrix}1 & \rho(1)\\ \rho(1) & 1\end{pmatrix}^{-1}\begin{pmatrix}\rho(1)\\ \rho(2)\end{pmatrix} \Rightarrow \phi_{22} = \dfrac{\hat\rho(2) - \hat\rho(1)}{1 - \hat\rho^2(1)} = 0.634 = \alpha(2) = \hat\alpha(2)$.

Thus, $$\alpha(h) = \begin{cases} 0.8062 = \hat\alpha(1) & h = 1\\ -0.634 = \hat\alpha(2) & h = 2\\ 0 & h \geq 3\end{cases}$$ ← $\boxed{\text{For } AR(p),\ \alpha(h) = 0\ \forall\ h > p.}$

↑ use defn.

- Model ACVF: Using YW eqn, we get:

$$\left.\begin{aligned} k=0:&\quad \gamma(0)-1.318\,\gamma(1)+0.634\,\gamma(2)=289.2\\ k=1:&\quad \gamma(1)-1.318\,\gamma(0)+0.634\,\gamma(1)=0\\ k=2:&\quad \gamma(2)-1.318\,\gamma(1)+0.634\,\gamma(0)=0 \end{aligned}\right\} (*) \Rightarrow \gamma(h)=\hat{\gamma}(h), \quad h=0,1,2,\ldots$$

Obtain the remaining ACVF's iteratively:

$$\gamma(k)-\phi_1\gamma(k-1)-\phi_2\gamma(k-2)=0 \qquad k\geq 1$$

$$\Rightarrow \gamma(h)=1.318\,\gamma(h-1)-0.634\,\gamma(h-2), \quad h\geq 3$$

Could use general solution: (2^{nd} method) $\leftarrow$ complicated!

$$\gamma(h)=\alpha_1\xi_1^{-h}+\alpha_2\xi_2^{-h} \qquad h\geq 0, \text{ where}$$

$\xi_1=re^{i\theta}$, $\xi_2=re^{-i\theta}$ and solve for α_1 and α_2 from YW eqn (*)

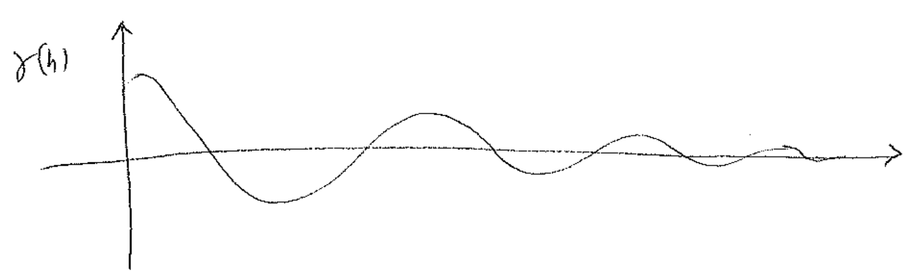
Get eventually:

$$\gamma(h)=\frac{\sigma^2 r^{4-h}\sin(h\theta+\psi)}{(r^2-1)(r^4-2r^2\cos 2\theta+1)\sin\theta} \qquad h\geq 0 \qquad (**)$$

where $\tan\psi=\dfrac{r^2+1}{r^2-1}\tan\theta$.

$\gamma(\cdot)$ for an AR(2) with complex conjugate roots is therefore a damped sinusoid fn with damping factor of $\frac{1}{r}$ and period $\frac{2\pi}{\theta}$.

If roots are close to unit circle, then r is close to 1, as in this case. The damping is slow and $\gamma(h)$ behaves almost like a sine wave!

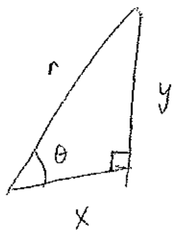


- Causality: Fitted AR polynomial:

$\phi(z) = 1 - 1.318z + 0.634z^2$ has roots:

$$z = \frac{1.318 \pm \sqrt{1.318^2 - 4 \cdot 0.634}}{2 \cdot 0.634} = 1.039 \pm 0.7045i = x+iy$$

In polar form:



$x + iy = re^{i\theta}$

$x - iy = re^{-i\theta}$

$r = \sqrt{x^2+y^2} \approx 1.26$

$\tan\theta = \frac{y}{x} = 0.6789 \Rightarrow \theta \approx 34.2° \Rightarrow \psi \approx 71.5°.$

Model is causal, but close to non-causal.

Note: Get remaining $\gamma(h)$ from (**)

$\boxed{02/20}$ §3.4 Forecasting: Obs: $\{X_1, \dots, X_n\}$

Suppose $X_t \sim AR(1)$. How should we predict X_{t+1}?

Model: $X_t = \phi X_{t-1} + Z_t$

mean 0 of Z_t

$\Rightarrow \hat{X}_{n+1} = \phi X_n + 0 = \phi X_n$? (should we do it this way?)

But what if $X_t \sim MA(1)$? $X_t = \theta Z_{t-1} + Z_t$

$\Rightarrow \hat{X}_{n+1} = 0 + 0 = 0$?

(*) Proper approach: Decision Theory. Specify a loss function.

Want forecast to satisfy?

1) Be "unbiased" $\Rightarrow$ $E(\hat{X}_{n+1} - X_n) = 0$.

2) Have minimum mean square error MSE: $\min E(\hat{X}_{n+1} - X_{n+1})^2$.

In fact, the best predictor of X_{n+h} based on $\{X_1, \dots, X_n\}$ in MSE sense is:

$$\hat{X}_{n+h} = \mathbb{E}(X_{n+h} \mid X_1, \dots, X_n)$$

↳ it's also unbiased. This is BP or BUP. ← always unbiased

Usually difficult to compute in practice (an analytical sln is required), so we focus on simple problem: require 1) and 2) criteria and the additional:

3) $\hat{X}_{n+h}$ should be linear (a linear combination of $X_1, \dots, X_n$ - the observed values).

↳ BLP. (or BLUP). ← unbiased included.

Outline:

a) BLP Theory for a general collection of rv's.

b) " stationary processes.

c) " ARMAs: - on finite past $\{X_1, \dots, X_n\}$

- on infinite past $\{X_n, X_{n-1}, \dots, X_1, X_0, X_{-1}, \dots\}$.

- BLP General Theory:

$\{Y, W_1, \dots, W_n\}$ collection of rv's with finite 2^{nd} moments.

Define: $\mu = EY$, $\mu_i = EW_i$, $\operatorname{Var} Y = \sigma^2$, $\underset{\sim}{W} = \begin{pmatrix} W_n \\ \vdots \\ W_1 \end{pmatrix}$, $\mu_W = E\underset{\sim}{W} = \begin{pmatrix} \mu_n \\ \vdots \\ \mu_1 \end{pmatrix}$,

$$\operatorname{Var} \underset{\sim}{W} = \Gamma = E\left[(\underset{\sim}{W} - \underset{\sim}{\mu_W})(\underset{\sim}{W} - \mu_W)'\right]$$

$$\underset{\sim}{\gamma} = \begin{pmatrix} \operatorname{Cov}(Y, W_n) \\ \vdots \\ \operatorname{Cov}(Y, W_1) \end{pmatrix} = \operatorname{Cov}(Y, \underset{\sim}{W}) = E\left[(Y - \mu)(\underset{\sim}{W} - \underset{\sim}{\mu_W})\right]$$

Thm : The BLP of Y in terms of $\{1, W_n, \ldots, W_1\}$

$$P(Y|\underset{\sim}{W}) = \mu + a'(W - \mu_{\underset{\sim}{w}}), \quad \underset{\sim}{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \text{ where } \underset{\sim}{a} \text{ is any solution of } \boxed{\underset{\sim}{\Gamma}\,\underset{\sim}{a} = \underset{\sim}{\gamma}}$$

The MSE of $P(Y|\underset{\sim}{W})$ is :

$$E\left(Y - P(Y|\underset{\sim}{W})\right)^2 = \sigma^2 - \underset{\sim}{a}'\underset{\sim}{\gamma}$$

Proof : Let $P(Y|\underset{\sim}{W}) = a_0 + \underset{\sim}{a}'\underset{\sim}{W}$ and write :

$$Y - P(Y|W) = (Y-\mu) + (\mu - a_0) + a'(W - \mu_w) - a'\mu_w$$

$$\equiv Y - a_0 - a'W \Rightarrow \text{square \& take expectation leads to this to minimize:}$$

$$f(a_0, \underset{\sim}{a}) = E\left[Y - P(Y|W)\right]^2 = \sigma^2 + (\mu - a_0)^2 - 2(\mu - a_0)a'\omega - 2a'\gamma + (a'u_w)^2 + a'\Gamma a.$$

- $\dfrac{\partial f}{\partial a_0} = -2(\mu - a_0) + 2a'\mu_w = 0 \Rightarrow \hat{a}_0 = \mu - a'\mu_w.$
- Plug optimal $\hat{a}_0$ into $f(a_0, a)$: $f(\hat{a}_0, a) = -2a'\gamma + a'\Gamma a$ (ignore const. by themselves).
- $\dfrac{\partial f(\hat{a}_0, a)}{\partial a} = -2\gamma + 2\Gamma a = 0 \Rightarrow \Gamma a = \gamma$

Note : Vector / matrix calculus results :

$$\frac{\partial (c'x)}{\partial x} = c, \quad \frac{\partial (x'Cx)}{\partial x} = (C + C')x.$$

Note : $P(Y|W) = \mu + \sum_{i=1}^{n} a_i (W_{n+1-i} - \mu_{n+1-i})$

Ex/ Estimate a missing value in $AR(1)$: $X_t - \phi X_{t-1} = Z_t$

Observe $\{X_1, X_3\}$. What is the min MSE linear estimator of X_2? Find vector $\underset{\sim}{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ such that $a_1 X_1 + a_2 X_3 = P(X_2 | X_1, X_3)$ is BLP of X_2 based on $\{X_1, X_3\}$.

Now: $\underset{\sim}{W} = \begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$, $Y = X_2$.

For $AR(1)$, $\gamma(h) = \dfrac{\sigma^2 \phi^{|h|}}{1-\phi^2} \Rightarrow \rho(h) = \phi^{|h|}$.

$$\Gamma = \begin{pmatrix} Cov(X_1, X_1) & Cov(X_1, X_3) \\ Cov(X_3, X_1) & Cov(X_3, X_3) \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(2) \\ \gamma(2) & \gamma(0) \end{pmatrix} = \frac{\sigma^2}{1-\phi^2} \begin{pmatrix} 1 & \phi^2 \\ \phi^2 & 1 \end{pmatrix}$$

$$\gamma = \begin{pmatrix} Cov(X_2, X_1) \\ Cov(X_2, X_3) \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(1) \end{pmatrix} = \frac{\sigma^2}{1-\phi^2} \begin{pmatrix} \phi \\ \phi \end{pmatrix}$$

$$\underset{\sim}{a} = \Gamma^{-1}\gamma = \begin{pmatrix} 1 & \phi^2 \\ \phi^2 & 1 \end{pmatrix} \begin{pmatrix} \phi \\ \phi \end{pmatrix} = \frac{1}{1-\phi^4} \begin{pmatrix} 1 & -\phi^2 \\ -\phi^2 & 1 \end{pmatrix} \begin{pmatrix} \phi \\ \phi \end{pmatrix} = \frac{1}{1+\phi^2} \begin{pmatrix} \phi \\ \phi \end{pmatrix}$$

$$MSE = \gamma(0) - (a_1, a_2) \begin{pmatrix} \gamma(1) \\ \gamma(1) \end{pmatrix} = \frac{\sigma^2}{1+\phi^2}$$

$$\Rightarrow P(X_2 | X_1, X_3) = \frac{\phi}{1+\phi^2}(X_1 + X_3) \quad \text{with } MSE = \frac{\sigma^2}{1+\phi^2}.$$

BLP: Stationary Processes

Apply general BLP theory to $\{Y_t\} \sim$ stationary with ACVF $\gamma(\cdot)$ and mean μ.

The BLP of Y_{n+h} based on $\{Y_1, \dots, Y_n\}$ is :

$$P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (Y_{n+1-i} - \mu) \quad \text{where the } a_i \text{ satisfy :}$$

$$\underset{\sim}{a_n} = \begin{pmatrix} a_1 \\ \vdots \\ \vdots \\ a_n \end{pmatrix} = \Gamma_n^{-1} \underset{\sim}{\gamma_n}(h) \quad , \quad \text{with MSE :}$$

$$E\left(P_n X_{n+h} - X_{n+h}\right)^2 = \gamma(0) - a_n' \underset{\sim}{\gamma_n}(h) \equiv \sigma_n^2(h) \quad , \quad \text{where :}$$

$$\underset{\sim}{\gamma_n}(h) = \begin{pmatrix} \gamma(h) \\ \vdots \\ \gamma(h+n-1) \end{pmatrix} , \quad \text{and} \quad \Gamma_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ & \ddots & \ddots & \gamma(n-2) \\ & & \ddots & \vdots \\ & (*) & & \gamma(1) \\ & & & \gamma(0) \end{pmatrix} .$$

Toeplitz matrix

$\boxed{02/22}$ BLP for stationary process.

$\{Y_t\}$ is stationary with mean μ and $\gamma(\cdot)$.

$(*)$ $$\underbrace{P_n Y_{n+h}}_{} = \mu + \sum_{i=1}^{n} a_i (Y_{n+1-i} - \mu)$$

BLP of Y_{n+h} based on $\{Y_1, \dots, Y_n\}$.

$$\underset{\sim}{a_n} = \begin{pmatrix} a_1 \\ \vdots \\ \vdots \\ a_n \end{pmatrix} = \underset{\sim}{\Gamma_n^{-1}} \underset{\sim}{\gamma_n}(h) \quad , \quad \text{where } \underset{\sim}{\gamma_n}(h) = \begin{pmatrix} \gamma(h) \\ \vdots \\ \gamma(h+n-1) \end{pmatrix} , \quad \text{and}$$

$$\underset{\sim}{\Gamma_n}(ij) = \left[\gamma(i-j)\right]_{i,j=0}^{n-1}$$

$(*)$ $$MSE = E\Big(Y_{n+h} - \underbrace{P_n Y_{n+h}}_{\text{predictor}}\Big)^2 = \gamma(0) - a' \underset{\sim}{\gamma_n}(h)$$

Ex/ $Y_t \sim WN\ 0(\mu, \sigma^2) \Rightarrow P_n Y_{n+h} = \mu + 0 = 0$

$\hookrightarrow$ since $\Gamma_n = \mathrm{diag}\{\sigma^2\}$, $\gamma(h) = 0' \Rightarrow a_n' = 0$.

$\Rightarrow$ MSE $= \sigma^2$ since $\gamma_n(h) = 0 \rightarrow$ mean is best predictor.

Note : P_n is BLP operator on the linear combination of observations $\{1, Y_1, \ldots, Y_n\}$

Properties of P_n :

- $E(P_n Y_{n+h}) = E(Y_{n+h})$ ($\Rightarrow$ unbiased predictor, ie. BLP = BLUP).
- $P_n Y_t = Y_t$ for $1 \le t \le n$.
- Prediction eqn : $P_n Y_{n+h}$ is unique solution of :

$$E\left[(Y_{n+h} - P_n Y_{n+h})\, Y_t\right] = 0 \qquad \forall\ t = 1, \ldots, n$$

with $a_0 = \mu\left(1 - \sum_{i=1}^{n} a_i\right)$.

$\rightarrow$ WLOG, focus on zero-mean processes X_t ($\mu = 0$).

In this case, the BLP of X_{n+1} is :

$$P_n X_{n+1} = a_1 X_n + \ldots + a_n X_1$$
$$= \phi_{n,1} X_n + \ldots + \phi_{n,n} X_1 \qquad \text{(notation)}$$

✳ Can be expressed in 1-step prediction error form :

$$P_n X_{n+1} = \theta_{n,1} \underbrace{(X_n - P_{n-1} X_n)}_{\text{1-step prediction error}} + \ldots + \theta_{n,n}(X_1 - \underbrace{P_0 X_1}_{=0})$$

(or μ in general)

Recursive algorithm can be used to find $P_0 X_1, \ldots, P_n X_{n+1}$:

- The Durbin - Levinson algorithm : finds the $\phi_{n,j}$'s and is better suited for AR (p) models.

- The Innovations Algorithm : finds the $\theta_{n,j}$'s , and is better suited for MA(q) & ARMA (p, q) models.

(*) For an ARMA(p,q), the BLP of X_{n+h} based on $X_1, \dots, X_n$ is :

$$P_n X_{n+h} = \underbrace{\sum_{i=1}^{p} \phi_i P_n X_{n+h-i}}_{AR(p) \rightarrow \text{recursive}} + \sum_{j=h}^{q} \theta_{n+h-i,j} \underbrace{\left(X_{n+h-j} - P_{n+h-j-1} X_{n+h-j} \right)}_{\text{one-step prediction error}}$$

The corresponding MSE , $\sigma_n^2(h)$ is obtained from the Innovations Algorithm , but is complicated. Can use large sample approximation :

$$\sigma_n^2(h) \approx \tilde{\sigma}_n^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$$

Also, $\tilde{P}_n X_{n+h}$ = BLP of X_{n+h} based on $X_n, X_{n-1}, \dots$.

$$= - \sum_{j=0}^{\infty} \pi_j P_n X_{n+h-j}$$ ← will prove next.

BLP operator on the infinite past.

Prediction on Infinite Past (causal & invertible ARMA) .

Notation : $\tilde{P}_n X_{n+h} = \{ \text{BLP of } X_{n+h} \text{ based on } X_n, X_{n-1}, \dots \}$

ARMA (p,q) : $\phi(B) X_t = \theta(B) Z_t$, $\{Z_t\} \overset{iid}{\sim} (0, \sigma^2)$

Causal : $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$

Invertible : $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ (need later)

$$\Rightarrow \quad X_{n+h} = \sum_{j=0}^{\infty} \psi_j Z_{n+h-j}$$

Write $\tilde{P}_n X_{n+h} = a_0 X_n + a_1 X_{n-1} + a_2 X_{n-2} + \dots$

Write : $\tilde{P}_n X_{n+h} = a_0 X_n + a_1 X_{n-1} + a_2 X_{n-2} + \dots$

$$= a_0 \sum_{j=0}^{\infty} \psi_j Z_{n-j} + a_1 \sum_{j=0}^{\infty} \psi_j Z_{n-1-j} + \dots$$

$$= \psi_h^* Z_n + \psi_{h+1}^* Z_{n-1} + \dots$$

where the ψ_j^* are to be determined.

MSE of this forecast is :

$$\tilde{\sigma}_n^2(h) = E\left(X_{n+h} - \tilde{P}_n X_{n+h}\right)^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 + \sigma^2 \sum_{j=0}^{\infty} \left(\psi_{h+j} - \psi_{h+j}^*\right)^2$$

This is clearly minimized when $\psi_{h+j}^* = \psi_{h+j}$, $j \geq 0$. Hence :

$$\tilde{P}_n X_{n+h} = \psi_h Z_n + \psi_{h+1} Z_{n-1} + \dots = \sum_{j=h}^{\infty} \psi_j Z_{n+h-j} \qquad (*)$$

From the invertibility, can equivalently write :

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = X_t + \sum_{j=1}^{\infty} \pi_j X_{t-j}$$

defn of invertibility

$$\tilde{P}_n X_{n+h} = - \sum_{j=1}^{\infty} \pi_j \tilde{P}_n X_{n+h-j}$$

Proof : Apply $\tilde{P}_n$ to both sides of : $Z_{n+h} = X_{n+h} + \sum_{j=1}^{\infty} \pi_j X_{n+h-j}$

with MSE : $\tilde{\sigma}_n^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$.

(Properties of $\tilde{P}_n$ are similar to P_n) .

Exam 1 : Coverage up to § 3.4, (exclude 3.4).

1 sheet is allowed.

$\boxed{02/25}$ Prediction on Infinite Past (continue) - causal & invertible ARMA.

$$\tilde{P}_n X_{n+h} = \sum_{j=h}^{\infty} \psi_j Z_{n+h-j} = - \sum_{j=1}^{\infty} \pi_j P_n X_{n+h-j} \quad \text{with MSE} \quad \tilde{\sigma}_n^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 .$$

$\swarrow X_t = \sum \psi_j Z_{t-j}$

Now, from causality, note that: $\quad Z_t \sim iid.$

$$E\left(Z_{n+j} \mid \underbrace{X_n, X_{n-1}, \dots}_{\mathcal{F}_n = f(x_n, x_{n-1}, \dots)}\right) = E\left(Z_{n+j} \mid \underbrace{Z_n, Z_{n-1}, \dots}_{\mathcal{F}_n = f(z_n, z_{n-1}, \dots)}\right) = \begin{cases} 0 & j > 0 \\ Z_{n+j} & j \le 0 \end{cases}$$

Thus, since $\quad X_{n+h} = \psi_0 Z_{n+h} + \dots + \psi_{h-1} Z_{n+1} + \psi_h Z_n + \psi_{h+1} Z_{n-1} + \dots$

the BP on infinite past is

$$BP = \mathbb{E}\left(X_{n+h} \mid X_n, X_{n-1}, \dots\right) = \sum_{j=0}^{h-1} \psi_j \, \mathbb{E}\left(Z_{n+h-j} \mid X_n, X_{n-1}, \dots\right) + \sum_{j=h}^{\infty} \psi_j \, \mathbb{E}\left(Z_{n+h-j} \mid X_n, X_{n-1}, \dots\right)$$

$$= \sum_{j=0}^{h-1} \psi_j (0) + \sum_{j=h}^{\infty} \psi_j Z_{n+h-j} = \sum_{j=h}^{\infty} \psi_j Z_{n+h-j} = \tilde{P}_n X_{n+h} = BLP.$$

<u>Thm</u>: For ARMA (causal & invertible), the BP and BLP on infinite past coincide.

<u>Note</u>: i) $\tilde{P}_n X_{n+h} = \sum_{h}^{\infty} \psi_j Z_{n+h-j} \to 0 \quad$ as $\quad h \to \infty$.

ii) $\tilde{\sigma}_n^2(h) = \sigma^2 \sum_{0}^{h-1} \psi_j^2 \to \sigma^2 \sum_{0}^{\infty} \psi_j \quad$ as $\quad h \to \infty$.

So the long (range) run forecast quickly settle on $EX_t = 0$ with constant confidence bands, since the ψ_j's decay exponentially fast.

- Recall BLP on finite past for ARMA (p,q):

$$P_n X_{n+h} = \sum_{i=1}^{p} \phi_i P_n X_{n+h-i} + \sum_{j=h}^{q} \theta_{n+h-1,j}$$

↖ 1-step prediction errors

<u>Exs?</u> $\quad$ (*) $P_n X_t = X_t$, $t \le n$.

- <u>In case of AR(p)</u>: $P_n X_{n+h} = \sum_{i=1}^{p} \phi_i P_n X_{n+h-i}$

$$\Rightarrow \underset{\substack{\uparrow \\ h=1}}{P_n X_{n+1}} = \phi_1 \underbrace{P_n X_n}_{X_n} + \phi_2 \underbrace{P_n X_{n-1}}_{X_{n-1}} + \dots + \phi_p P_n X_{n+1-p} = \phi_1 X_n + \dots + \phi_p X_{n+1-p}. \quad (\text{1-step})$$

$$P_n X_{n+2} = \phi_1 \underbrace{P_n X_{n+1}}_{\text{1-step}} + \phi_2 \underbrace{P_n X_n}_{X_n} + \dots + \phi_p \underbrace{P_n X_{n+2-p}}_{X_{n+2-p}} \qquad \text{(2-step)}$$

<u>Subcase</u>: AR(1): $P_n X_{n+1} = \phi X_n$

$$P_n X_{n+2} = \phi P_n X_{n+1} = \phi^2 X_n$$

$$\vdots$$

$$\boxed{P_n X_{n+h} = \phi^h X_n}$$

$\psi_j = \phi^j, \quad j \geq 0 \quad (AR(1))$

MSE: $\sigma_n^2(h)$ is too complex, use $\tilde{\sigma}_n^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$ instead.

Since $\pi_j = \begin{cases} 1 & j=0 \\ -\phi & j=1 \\ 0 & j \geq 2 \end{cases}$, the large sample approx. gives:

$$\tilde{P}_n X_{n+h} = \phi^h X_n \quad \text{with MSE:} \quad \tilde{\sigma}_n^2(h) = \sigma^2 \sum_{j=0}^{h-1} \phi^{2j}$$

- <u>In case of MA(q):</u> $P_n X_{n+h} = \sum_{j=h}^{q} \theta_{n+h-1,j} (X_{n+h-j} - P_{n+h-j-1} X_{n+h-j})$

$\Rightarrow \quad P_n X_{n+1} = \theta_{n,1} (X_n - P_{n-1} X_n) + \dots + \theta_{n,q} (X_{n+1-q} - P_{n-q} X_{n+1-q})$

<u>Subcase</u>: MA(1): $P_n X_{n+1} = \theta_{n,1} (X_n - P_{n-1} X_n)$

Apply innovations algorithm to find: $v_0, \theta_{11}, v_1, \theta_{21}, v_2, \dots, \theta_{n1}, v_n, \dots$

n	$P_n X_{n+1} = \theta_{n1} (X_n - P_{n-1} X_n)$	MSE
0	$P_0 X_1 = 0$	v_0
1	$P_1 X_2 = \theta_{11} (X_1 - 0) = \theta_{11} X_1$	v_1
$\vdots$	$\vdots$	$\vdots$
n	$P_n X_{n+1} = \theta_{n1} (X_n - P_{n-1} X_n)$	v_n

This is a complicated recursive calculation. Using the large-sample approx.:

$$\tilde{P}_n X_{n+1} = -\sum_{j=1}^{\infty} \pi_j \tilde{P}_n X_{n+1-j} = -(\pi_1 X_n + \pi_2 X_{n-1} + \dots + \pi_n X_1 + \dots)$$

It remains to compute $\{\pi_j\}$ for an MA(1):

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \Rightarrow Z_t = \pi(B) X_t .$$

But: $X_t = \theta(B) Z_t = \theta(B)\pi(B) X_t \Rightarrow \boxed{\pi(B) = \frac{1}{\theta(B)}}$

$|\theta|<1.$

$X_t = Z_t + \theta Z_{t-1} = (1+\theta B) Z_t \Rightarrow \theta(B) = 1+\theta B. \Rightarrow \pi(B) = \frac{1}{1+\theta B}$: infinite sum

Can show: $\pi_j = (-\theta)^j \qquad j = 0, 1, 2, \dots$

go on!

Thus: $\tilde{P}_n X_{n+1} = -\pi_1 X_n - \pi_2 X_{n-1} - \dots - \pi_n X_1 \dots \approx \theta X_n - \theta^2 X_{n-1} + \dots + (-1)^{n+1}\theta^n X_1$

$= \tilde{x}^n_{n+1} \quad$ (S&S)

with MSE $= \tilde{\sigma}_n^2(1) = \sigma^2$.

<u>Notation comparison</u>: B&D and S&S:

- BLP of X_{n+h} based on $\{X_n, \dots, X_1\}$

$$\left\{\begin{array}{l} P_n X_{n+h} \\ \sigma_n^2(h) \end{array}\right. \text{(BD)} \iff \left\{\begin{array}{l} x^n_{n+h} \\ P^n_{n+h} \end{array}\right. \text{(SS)}$$

- BLP of X_{n+h} based on $\{X_n, \dots, X_1, X_0, X_{-1}, \dots\}$ infinite past

$$\left\{\begin{array}{l} \tilde{P}_n X_{n+h} \\ \tilde{\sigma}_n^2(h) \end{array}\right. \text{(BD)} \iff \left\{\begin{array}{l} \tilde{x}_{n+h} \\ P^n_{n+h} \end{array}\right. \text{(Ss)} \iff \left\{\begin{array}{l} \tilde{x}^n_{n+h} \\ P^n_{n+h} \end{array}\right. \text{(SS)}$$

untruncated

truncated version of $\tilde{x}_{n+h}$ by discarding $\{X_0, X_{-1}, \dots\}$

↳ BD do not have truncated stuff.

Prediction Bounds

(AR(p) : $P_n X_{n+h} = \sum_{i=1}^{p} P_n X_{n+h-i} \phi_i$) ← prediction theory

- Under the assumption of Gaussian WN, 95% prediction bounds can be constructed for a predictor, since in this case :

$$X_{n+h} - \tilde{P}_n X_{n+h} \sim N\left(0, \tilde{\sigma}_n^2(h)\right)$$

This gives the PI for X_{n+h} : $\tilde{P}_n X_{n+h} \pm 1.96\, \tilde{\sigma}_n(h)$.

Ex1 Forecast the next 3 obs. for Sunspot data (earlier fitted an AR(2)).

Fitted model: $X_t - 1.318 X_{t-1} + 0.634 X_{t-2} = Z_t$, $\{Z_t\} \sim WN\ (0, 289.2)$

$$Y_t = X_t + 46.93$$

$P_n X_{n+h} = \sum_{i=1}^{p} \phi_i P_n X_{n+h-i}$

- $P_n X_{n+1} = \phi_1 X_n + \phi_2 X_{n-1} = \phi_1 (Y_n - \mu) + \phi_2 (Y_{n-1} - \mu)$

$= 1.318\ (74 - 46.93) - 0.634\ (37 - 46.93) = 41.97$

$\Rightarrow P_n Y_{n+1} = \mu + P_n X_{n+1} = 88.90$

- $P_n X_{n+2} = \phi_1 P_n X_{n+1} + \phi_2 X_n = 1.318 \cdot 41.97 - 0.634\ (-46.93 + 74) = 38.15$

$\Rightarrow P_n Y_{n+2} = \mu + P_n X_{n+2} = 85.08$

- $P_n X_{n+3} = \phi_1 P_n X_{n+2} + \phi_2 P_n X_{n+1} = \phi_1 P_n X_{n+2} + \phi_2 P_n X_{n+1}$

$= 1.318\ (38.15) - 0.634\ (41.97) = 23.67$

$\Rightarrow P_n Y_{n+3} = \mu + P_n X_{n+3} = 70.60$.

The exact MSE's are given by :

$$\sigma_n^2(h) = \gamma(0) - (\phi_1, \phi_2) \begin{pmatrix} \gamma(h) \\ \gamma(h+1) \end{pmatrix}$$

Recall : $\gamma(0) = 1382.2$ $\quad \gamma(4) = -277.27$

$\gamma(1) = 1114.2$ $\quad$ (thru YW eqn)

$\gamma(2) = 591.73$

$\gamma(3) = 74.92$

$$\Rightarrow \quad \sigma_n^2(h) = \begin{cases} 288.8 & h=1 \\ 649.9 & h=2 \\ 1107.7 & h=3 \end{cases} \qquad \leftarrow \text{more accurate (finite, observed values).}$$

- Compare the MSE's with asymptotic MSE's ($n = \infty$):

$$\tilde{\sigma}_n^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \qquad \psi_j = \begin{cases} 1 & j=0 \\ 1.318 & j=1 \\ 1.1031 & j=2 \end{cases}, \qquad \sigma^2 = 289.2$$

$$= \begin{cases} 289.2 & h=1 \\ 791.6 & h=2 \\ 1143.5 & h=3 \end{cases} \qquad n=100.$$

- Finally, 95% PI for $P_n Y_{n+h}$:

$h=1$: $\quad 88.90 \pm 1.96\sqrt{288.8}$

$h=2$: $\quad 85.08 \pm 1.96\sqrt{649.9}$

$h=3$: $\quad 70.60 \pm 1.96\sqrt{1107.7}$

$\boxed{03/04}$

§3.5 Parameter Estimation.

The determination of an appropriate ARMA (p,q) model to represent an observed stationary time series, involves 3 interrelated problems:

- Choice of p and q (order selection or model identification problem).
- Estimation of $\mu, \sigma^2, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$. (parameter estimation).
- Checking whiteness of residuals (test for goodness of fit).

We will first discuss the parameter estimation problem, by assuming p & q are known. The 1st step is to subtract mean μ (estimate it with $\bar{y}$), so that it's appropriate to fit zero-mean ARMA to the mean-corrected data:

$$\{ x_1 = y_1 - \bar{y}, \ldots, x_n = y_n - \bar{y} \}$$

Note : μ can also be incorporated into the likelihood function as an extra para.

(this is default approach of most softwares).

We will discuss 2 main methods of estimation : maximum likelihood & quasi-maximum likelihood.

- Method 1 : (MLE)

When p & q are known, good estimators can be found by assuming the data $\underset{\sim}{x}_n$ is a realization of a Gaussian process:

$$\underset{\sim}{x}_n = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \sim N(\underset{\sim}{0}, \Gamma_n) \quad \text{where} \quad \underset{\sim}{\Gamma}_n(i,j) = \underset{\substack{\downarrow \\ \gamma(|i-j|)}}{\gamma(i-j)}$$

$(i,j)^{th}$ ele.

Now maximize the likelihood w.r.t :

$$\underset{\sim}{\phi} = (\phi_1, \ldots, \phi_p)', \quad \underset{\sim}{\theta} = (\theta_1, \ldots, \theta_q)', \text{ and } \sigma^2.$$

The Gaussian likelihood for a causal / invertible ARMA(p,q) is :

$$L(\underset{\sim}{\phi}, \underset{\sim}{\theta}, \sigma^2) = (2\pi)^{-\frac{n}{2}} \det(\Gamma_n)^{-1/2} \exp\left\{ -\frac{1}{2} x_n' \Gamma_n^{-1} x_n \right\}$$

Using the Innovations algorithm ; this can also be written in terms of the one-step predictors and their MSEs as :

$$L(\underset{\sim}{\phi}, \underset{\sim}{\theta}, \sigma^2) = (2\pi)^{-\frac{n}{2}} (\underbrace{v_0 \cdots v_{n-1}}_{\text{product}})^{-\frac{1}{2}} \exp\left\{ \frac{-1}{2} \sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{v_{j-1}} \right\}$$

$$\left(\prod_{j=1}^{n} v_{j-1} \right)^{-\frac{1}{2}}$$

where $\hat{x}_j = P_{j-1} x_j$ and $v_{j-1} = E(x_j - \hat{x}_j)^2$ observed value

$\updownarrow$

$\hat{X}_j = P_{j-1} X_j$ & $v_{j-1} = E(X_j - \hat{X}_j)^2$ rv

This form is more suitable for computation.

Except in simple special cases, MLE can only be performed numerically, starting with good initial estimates. We therefore will examine preliminary estimation methods which can be used as (recursive) precursors to MLE.

(a) Yule - Walker Estimation (best for AR):

For AR(p): we have YW eqns: $\underset{\sim}{\phi} = \underset{\sim}{\Gamma}_p^{-1} \underset{\sim}{\gamma}_p$ and $\sigma^2 = \gamma(0) - \underset{\sim}{\phi}' \underset{\sim}{\gamma}_p$

Replace $\gamma(h) \to \hat{\gamma}(h)$ to get YW estimators (MOME's).

The Durbin-Levinson Algorithm can be used for this.

The YW estimator has the large sample distribution: $\hat{\underset{\sim}{\phi}} \sim N\left(\underset{\sim}{\phi}, \frac{\sigma^2}{n} \Gamma_p^{-1}\right)$.

This is the same as that of the MLE. By replacing $\underset{\sim}{\Gamma}_p$ by $\hat{\Gamma}_p$, we can construct approx. 95% confidence bounds for ϕ_i: $\hat{\phi}_i \pm 1.96 \sqrt{\left[\frac{\hat{\sigma}^2}{n} \hat{\Gamma}_p^{-1}\right]_{ii}}$.

(b) The Burg- Algorithm: (for AR)

This algorithm is similar to YW, the estimates are obtained using a modified version of Durbin-Levinson algorithm. The estimates here have same large sample dist^{n} as YW est.

(c) The Innovations Algorithm (for MA):

Analogous asymptotic normality results hold for the Innovations Estimates, which allows for the construction of confidence bounds.

For a general ARMA (p, q), some form of least-squares estimation is often used.

We'll discuss 3 versions of least-squares: (d_1, d_2, d_3),

(d1): LS version 1:

Note: Z_t term in ARMA (p,q) is like a residual:

$$Z_t = X_t - \phi X_{t-1} - \dots - \phi_p X_{t-p} - \theta_1 Z_{t-1} - \dots - \theta_q Z_{t-q}$$

It makes sense to choose $(\sigma^2, \underset{\sim}{\phi}, \underset{\sim}{\theta})$ to minimize:

$$(\tilde{\underset{\sim}{\phi}}, \tilde{\underset{\sim}{\theta}}) = \text{argmin } S(\underset{\sim}{\phi}, \underset{\sim}{\theta}), \text{ where } S(\underset{\sim}{\phi}, \underset{\sim}{\theta}) = \sum_{t=1}^{n} Z_t^2(\underset{\sim}{\phi}, \underset{\sim}{\theta} \mid x_{1-p}, \dots, x_0, z_{1-q}, \dots, z_0)$$

$\boxed{03/06}$ §3.5 Con't : Parameter Estimation

Method 1: MLE (Gaussian likelihood)

Preliminary Estimation Methods:

a) YW
b) Burg
$\}$ (AR)

c) Innovations $\}$ (MA)

d) least-squares (ARMA): 3 versions:

→ (d1). $Z_t = X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} - \theta_1 Z_{t-1} - \dots - \theta_q Z_{t-q}$

Choose $(\tilde{\underset{\sim}{\phi}}, \tilde{\underset{\sim}{\theta}}) = \text{argmin } S(\underset{\sim}{\phi}, \underset{\sim}{\theta})$, where

minimize SS of residuals

$$S(\underset{\sim}{\phi}, \underset{\sim}{\theta}) = \sum_{t=1}^{n} z_t^2(\underset{\sim}{\phi}, \underset{\sim}{\theta} \mid x_{1-p}, \dots, x_0, z_{1-q}, \dots, z_0.)$$

where we can use these natural values for pre-sample:

$$\left.\begin{aligned} x_{1-p} = \dots = x_0 = E X_t = 0 \\ z_{1-q} = \dots = z_0 = E Z_t = 0 \end{aligned}\right\} \text{ and } \tilde{\sigma}^2 = \frac{S(\tilde{\underset{\sim}{\phi}}, \tilde{\underset{\sim}{\theta}})}{n} \approx \text{sample variance of } Z_t.$$

Called conditional LSEs (or CLSEs):

Ex/ ARMA (1,1): $S(\underset{\sim}{\phi}, \underset{\sim}{\theta}) = \sum_{t=1}^{n} z_t^2$, $z_t = x_t - \phi x_{t-1} - \theta z_{t-1}$

Start with $x_0 = 0 = z_0$:

$$z_1 = x_1 - \phi x_0 - \theta z_0 = x_1$$

$$z_2 = x_2 - \phi x_1 - \theta z_1 = x_2 - \phi x_1 - \theta x_1 = x_2 - (\phi + \theta) x_1$$

$$\vdots$$

$$z_n = x_n - \phi x_{n-1} - \theta z_{n-1} = x_n - (\phi + \theta) x_{n-1}$$

This is similar to an approx. Gaussian likelihood. For the ARMA (1,1):

$Z_1, \dots, Z_n \sim$ iid $N(0, \sigma^2)$ $\Rightarrow$ likelihood : iid z_t

$$L^*(z_1, \dots, z_n \mid \phi, \theta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ \frac{-1}{2\sigma^2} \sum_{t=1}^{n} z_t^2 \right\}$$

$$\Rightarrow \log L^*(z_1, \dots, z_n \mid \phi, \theta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{S(\phi, \theta)}{2\sigma^2}$$

<u>Note</u> : This is similar to but not exactly the same as $L(x_1, \dots, x_n)$ - the Gaussian likelihood for $(x_1, \dots, x_n)$ given earlier.

CLSEs.

Most softwares use these conditional least squares estimators as default as a starting point for full MLE.

→ (d2) Another version of CLSE is unconditional LSE (ULSE), uses 1-step prediction to define :

$$S(\underset{\sim}{\phi}, \underset{\sim}{\theta}) = \sum_{t=1}^{n} \frac{(X_t - \hat{X}_t)^2}{\nu_{t-1}/\sigma^2} \qquad \begin{cases} \hat{X}_t = P_{t-1} X_t & \text{(one-step ahead)} \\ \nu_{t-1} = E(\hat{X}_t - X_t)^2 \end{cases}$$

and $$\tilde{\sigma}^2 = \frac{S(\tilde{\underset{\sim}{\phi}}, \tilde{\underset{\sim}{\theta}})}{n-(p+q)}$$

ULSE's $(\tilde{\underset{\sim}{\phi}}, \tilde{\underset{\sim}{\theta}})$ minimize $S(\underset{\sim}{\phi}, \underset{\sim}{\theta})$.

→ (d3) Another version of LSE is the Hannon - Zissanen (B & D) is same as d1, but unobserved pre-sampled Z_t's are estimated by fitting a high-order AR to the data first.

(same as before but : $z_t = x_t - \phi x_{t-1} - \theta \hat{z}_{t-1}$)

$x_0 = 0$, $z_0 = \hat{z}_0$

fit from AR.

Method 2 : (QMLE) or CMLE

An alternative to maximize the full or exact likelihood of $\underset{\sim}{x}_n$, is to maximize a conditional or quasi-likelihood (QMLE) (CMLE).

Ex1 ARMA(p,q) : with noise pdf $f_Z(z)$, max the conditional log-likelihood :

$$CL(\underset{\sim}{\phi}, \underset{\sim}{\theta}) = \sum_{t=p+1}^{n} \log f_Z(z_t) \quad \text{depends on } x_1, \dots, x_p,$$

where z_t values are obtained from ARMA model :

$$z_t = x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p} - \theta_1 z_{t-1} - \dots - \theta_q z_{t-q}.$$

(set initial values equal to $0 \leftarrow$ of z_t).

$$L = f(x_1, \dots, x_n)$$
$$CL = f(z_{p+1}, \dots, z_n) = \prod_{p+1}^{n} f_Z(z_t)$$

This allows greater flexibility in specifying other-than-Gaussian likelihoods, but the asymptotic dist^{n} of QMLEs is more complicated.

⊛ Asymptotic Properties of MLEs:

If $\{X_t\}$ is a zero-mean causal / invertible ARMA(p,q) :

$$\phi(B) X_t = \theta(B) Z_t \quad , \quad \underbrace{Z_t \sim \text{iid } (0, \sigma^2)}_{\text{Note !}}$$

If $\underset{\sim}{\hat\beta} = (\underset{\sim}{\hat\phi}', \underset{\sim}{\hat\theta}')'$ denotes the MLE's (from Gaussian likelihood), then :

Thm : $\sqrt{n}(\underset{\sim}{\hat\beta} - \underset{\sim}{\beta}) \xrightarrow{d} N\left(\underset{\sim}{0}, \Omega(\underset{\sim}{\beta})\right)$, where the asymptotic covariance matrix $\Omega(\underset{\sim}{\beta})$, the inverse of usual Fisher information matrix, is given by :

Fisher information

$$\Omega(\underset{\frown}{\beta}) = \begin{cases} \sigma^2 \left[E(\underset{\frown}{U}_t \underset{\frown}{U}_t') \right]^{-1}, & \text{if } q = 0 \quad (AR(p)) \\ \sigma^2 \left[E(\underset{\frown}{V}_t \underset{\frown}{V}_t') \right]^{-1}, & \text{if } p = 0 \quad (MA(q)) \\ \sigma^2 \begin{bmatrix} E(\underset{\frown}{U}_t \underset{\frown}{U}_t') & E(\underset{\frown}{U}_t \underset{\frown}{V}_t') \\ E(\underset{\frown}{V}_t \underset{\frown}{U}_t') & E(\underset{\frown}{V}_t \underset{\frown}{V}_t') \end{bmatrix}^{-1}, & \text{if } p \geq 1 \text{ and } q \geq 1 \quad (ARMA) \end{cases}$$

when $\{U_t\}$ and $\{V_t\}$ are respectively the AR process :

$$\underbrace{\phi(B) U_t = Z_t}_{AR(p)} \quad \& \quad \underbrace{\theta(B) V_t = Z_t}_{AR(q)}$$

$$\underset{\frown}{U}_t = \begin{pmatrix} U_t \\ \vdots \\ U_{t-1+p} \end{pmatrix}, \quad V_t = \begin{pmatrix} V_t \\ \vdots \\ V_{t-1+q} \end{pmatrix}$$

$\boxed{03/08}$ <u>Asymptotics of MLEs :</u>

$$\underset{\frown}{\hat{\beta}} = \begin{pmatrix} \underset{\frown}{\hat{\phi}} \\ \underset{\frown}{\hat{\theta}} \end{pmatrix} \sim AN\left(\underset{\frown}{\beta}, \frac{1}{n} \Omega(\underset{\frown}{\beta}) \right)$$

$$\underset{\frown}{U}_t = \begin{pmatrix} U_t \\ \vdots \\ U_{t-1+p} \end{pmatrix}, \quad \underset{\frown}{V}_t = \begin{pmatrix} V_t \\ \vdots \\ V_{t-1+p} \end{pmatrix}$$

$\phi(B) U_t = Z_t \quad AR(p)$

$\theta(B) V_t = Z_t \quad AR(q)$

$$\Omega(\beta) = \begin{cases} \sigma^2 \left[E(\underset{\frown}{U}_t \underset{\frown}{U}_t') \right]^{-1} & AR(p) \\ \sigma^2 \left[E(\underset{\frown}{V}_t \underset{\frown}{V}_t') \right]^{-1} & MA(q) \\ \sigma^2 \begin{bmatrix} E(\underset{\frown}{U}_t \underset{\frown}{U}_t') & E(\underset{\frown}{U}_t \underset{\frown}{V}_t') \\ E(\underset{\frown}{V}_t \underset{\frown}{U}_t') & E(\underset{\frown}{V}_t \underset{\frown}{V}_t') \end{bmatrix}^{-1} & ARMA(p,q) \quad p \geq 1, q \geq 1 \end{cases}$$

Ex/ $AR(p)$, $E\underset{\sim}{U_t}\underset{\sim}{U_t}' = \begin{pmatrix} EU_t^2 & EU_tU_{t+1} & \cdots & EU_tU_{t-1+p} \\ & EU_{t+1}^2 & & \\ & & \ddots & \\ * & & & \ddots \end{pmatrix} = \Gamma_p$

$\Rightarrow \quad \Omega(\underset{\sim}{\phi}) = \sigma^2 \Gamma_p^{-1}$.

- If $p=1$, $\Gamma_1 = \gamma(0) = \dfrac{\sigma^2}{1-\phi^2} \Rightarrow \hat{\phi} \sim AN\left(\phi, \dfrac{\sigma^2}{n}\Gamma_1^{-1}\right) \overset{d}{=} AN\left(\phi, \dfrac{1-\phi^2}{n}\right)$

$\Rightarrow$ 95% CI for ϕ: $\quad \hat{\phi} \pm 1.96\sqrt{\dfrac{1-\hat{\phi}^2}{n}}$.

Ex/ $MA(q)$, $E\underset{\sim}{V_t}\underset{\sim}{V_t}' = \Gamma_q^* = $ ACVF matrix of the $AR(q)$: $\begin{pmatrix} \gamma^*(0) & \cdots & \gamma^*(q-1) \\ \vdots & \ddots & \vdots \end{pmatrix}$

$$V_t + \theta V_{t-1} + \cdots + \theta_q V_{t-q} = Z_t$$

$$\Rightarrow \quad \Omega(\underset{\sim}{\theta}) = \sigma^2 \Gamma_q^{*-1} \Rightarrow \underset{\sim}{\hat{\theta}} \sim AN\left(\underset{\sim}{\theta}, \frac{\sigma^2}{n}\Gamma_q^{*-1}\right)$$

- If $q=1$, $\Gamma_1^* = \gamma^*(0) \leftarrow$ consider $AR(1)$: $V_t + \theta V_{t-1} = V_t - (-\theta V_{t-1}) = Z_t$

$$= \frac{\sigma^2}{1-\theta^2} \Rightarrow \hat{\theta} \sim AN\left(\theta, \frac{1-\theta^2}{n}\right)$$

$\Rightarrow$ 95% CI for θ : $\quad \hat{\theta} \pm 1.96\sqrt{\dfrac{1-\hat{\theta}^2}{n}}$.

Ex/ $ARMA(1,1)$: $\Omega(\phi,\theta)$ is a mix of ACVF of the 2 $AR(1)$'s :

$$U_t - \phi U_{t-1} = Z_t \quad \& \quad V_t + \theta V_{t-1} = Z_t$$

$\phi = -\theta$

$(p=1) \quad E\underset{\sim}{U_t}\underset{\sim}{U_t}' = EU_t^2 = \gamma_u(0) = \dfrac{\sigma^2}{1-\phi^2}$

$(q=1) \quad E\underset{\sim}{V_t}\underset{\sim}{V_t}' = EV_t^2 = \gamma_v(0) = \dfrac{\sigma^2}{1-\theta^2}$

$E\,U_t V_t' = E\,V_t U_t' = \,?$

Appeal : to the causal representation :

$$U_t = \sum_{i=0}^{\infty} \phi^i Z_{t-i} \qquad \& \qquad V_t = \sum_{j=0}^{\infty} (-\theta)^j Z_{t-j}$$

$$\Rightarrow \quad U_t V_t = \sum_{j=0}^{\infty} \phi^j (-\theta)^j Z_{t-j}^2 + \sum_{\substack{i=0 \\ (i \neq j)}}^{\infty} \sum_{j=0}^{\infty} \phi^i (-\theta)^j Z_{t-i} Z_{t-j}$$

$$\Rightarrow \quad E\,U_t V_t = \sigma^2 \sum_{j=0}^{\infty} (-\phi\theta)^j + 0 = \frac{\sigma^2}{1+\phi\theta}$$

Then, 2×2 :

$$\Omega(\phi, \theta) = \sigma^2 \begin{pmatrix} \frac{1}{1-\phi^2} & \frac{1}{1+\phi\theta} \\ \frac{1}{1+\phi\theta} & \frac{1}{1-\theta^2} \end{pmatrix}^{-1}$$

$$\Rightarrow \quad \begin{pmatrix} \hat{\phi} \\ \hat{\theta} \end{pmatrix} \sim AN \left[\begin{pmatrix} \phi \\ \theta \end{pmatrix}, \frac{1+\phi\theta}{n(\phi+\theta)^2} \begin{pmatrix} (1-\phi^2)(1+\phi\theta) & -(1-\theta^2)(1-\phi^2) \\ -(1-\theta^2)(1-\phi^2) & (1-\theta^2)(1+\phi\theta) \end{pmatrix} \right]$$

<u>Notes :</u> for ARMA (p, q) :

$\boxed{1.}$ <u>Approximating $\Omega(\beta) \doteq$</u>

$$\sigma^2 \begin{array}{cc} \Gamma_{pp} & \Gamma_{pq} \end{array}$$
$$\begin{pmatrix} E\,U_t U_t' & E\,U_t V_t' \\ E\,V_t U_t' & E\,V_t V_t' \end{pmatrix}$$
$$\Gamma_{qq}$$

$\{U_t\}$ and $\{V_t\}$ are :

$$\phi(B) U_t = Z_t \quad \& \quad \theta(B) V_t = Z_t \qquad [AR(p) \;\&\; AR(q).]$$

<u>Method 1 :</u>

$$\cdot\ \Gamma_{pp} = \begin{pmatrix} \gamma_u(0) & \cdots & \gamma_u(p-1) \\ \vdots & \ddots & \vdots \\ \gamma_u(p-1) & \cdots & \gamma_u(0) \end{pmatrix} \approx \hat{\Gamma}_{pp} = \text{approx. of } \gamma_u(\cdot) \text{ via YW.}$$

- $\Gamma_{qq} \approx \hat{\Gamma}_{qq}$ (same way)
- $\Gamma_{pq} \approx \hat{\Gamma}_{pq}$ obtained as follows :

Use MA(∞) representation for U_t and V_t (using MLE's $\hat{\underset{\sim}{\phi}}$ and $\hat{\underset{\sim}{\theta}}$) :

not AR(∞) !

$$U_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad \& \quad V_t = \sum_{j=0}^{\infty} \pi_j Z_{t-j} \quad . \quad \text{Then :}$$

$$E\, U_t V_{t-h} = E\Big[\big(1 + \psi_1 Z_{t-1} + \ldots + \psi_h Z_{t-h} + \psi_{h+1} Z_{t-h-1} + \ldots\big) * \big(1 + \pi_1 Z_{t-1} + \ldots + \pi_h Z_{t-h} + \pi_{h+1} Z_{t-h-1} + \ldots\big)\Big]$$

$$= \sigma^2 \sum_{j=0}^{\infty} \pi_j \psi_{h+j}$$

In practice, we truncate the sum once the terms are small enough.

<u>Method 2</u>: Under correct model specification, $\Omega(\beta)$ can be related to the Hessian matrix of the log-likelihood :

$$\frac{\Omega(\beta)}{n} = -H^{-1}(\beta), \qquad H(\beta) = \left(\frac{\partial^2 \ell(\beta)}{\partial \beta_i \partial \beta_j} \right)_{i,j=1}^{p+q}$$

$\ell(\beta) = \log L(\underset{\sim}{\beta}, \hat{\sigma}^2) = \log$ of reduced / profile likelihood.

(Packages tend to use the Hessian method because it is quite easy to approx. H numerically.)

$\boxed{2.}$ <u>Confidence Regions for $\underset{\sim}{\beta}$:</u>

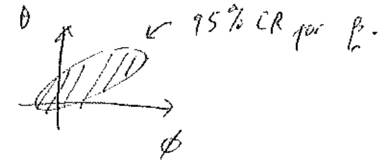
<u>Thm</u> : $\underset{\sim}{y} \sim N(\underset{\sim}{0}, \Sigma) \Rightarrow \underset{\sim}{y}'\Sigma^{-1}\underset{\sim}{y} \sim \chi^2_k$

dim = k

Our case : $\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta} \approx N\left(\underset{\sim}{0}, \frac{\Omega}{n}\right) \Rightarrow (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})' n \Omega^{-1} (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \sim \chi^2_{p+q}$

So an approx. $(1-\alpha)$ confidence region for β is:

$$\left\{ \underset{\sim}{\beta} \in \mathbb{R}^{p+q} : (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})' \, n \, \Omega^{-1} (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \le \frac{1}{n} \chi^2_{p+q}(1-\alpha) \right\}$$



03/18

Recap: CH3: ARMA Models.

§3.1 AR(p)
§3.2 MA(q)
§3.3 ARMA(p, q)
§3.4 forecasting ARMA
} model theory

§3.5 Parameter estimation ← data comes in here

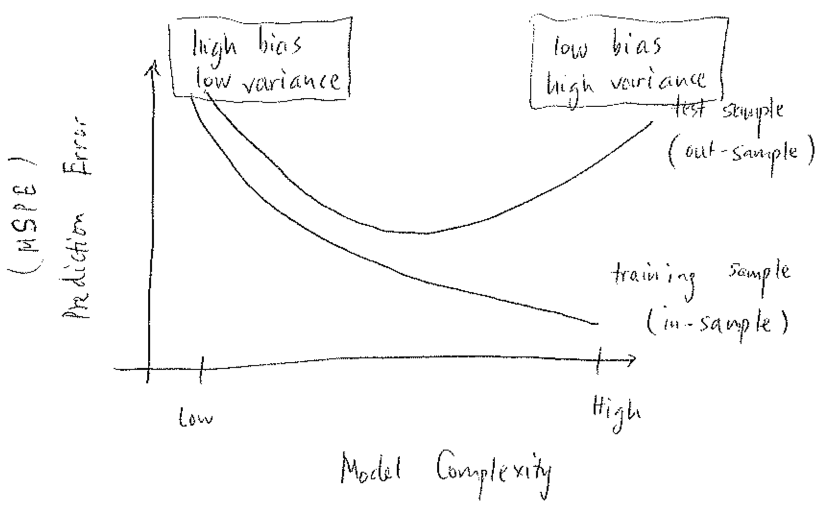
↳ MLE (Gaussian), LSEs, GMLEs

* 3 interrelated prob:

- 3.6 • model selection
- 3.5 • parameter est.
- 3.7 • gof

§3.6 Order Selection / Model Identification.

In real-life there is (usually) no underlying true model. The question then becomes: "how to select an appropriate statistical model (approximating), for a given dataset?"



(Hastie et al) "The element of statistical learning"

↳ 2009, Springer.

⊛ Some Solutions:

Cross validation, AIC/BIC, ...

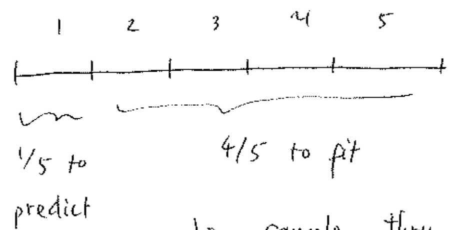
Some solutions :

1) Cross - Validation :

Approximate test error by resampling nonparametrically, don't need a model. In time series, unlike regression, there's the problem that we want the model to do well primarily in out-of-sample inference (forecasting).

↳ Solution is to split data into a test set (use for validation or model selection) and a training set (used for model fitting).

Ex / 5-fold CV



↳ sample thru all combinations : compute errors.

2) Information Criterion :

Approximate test error analytically, needs a model, is a so-called "covariance penalty procedure".

↳ First breakthrough was AIC (Akaike Information Criterion) which relates the Kullback-leibler distance btw the true model & the candidate model to the likelihood and number of model parameters.

We should therefore minimize the distance from truth & select ARMA (p,q) with smallest AIC value :

Akaike (1974)

$$AIC = -2\log L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2) + 2(p+q+1)$$ ← minimize this.

where $L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2)$ is the likelihood evaluated at MLEs. Nowadays, we tend to prefer the bias-corrected version :

Hurch & Tsai (1989)

$$AIC_c = -2\log L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2) + 2\,\frac{(p+q+1)\,n}{n-p-q-2}$$

$AIC_c \to AIC$ as $n \to \infty$.

- In practice, models with AIC value within about 2 units of min. value should also be considered. Final model selection should be based on <u>parsimony</u> (simplicity) & <u>gof</u>.

- There are several others information criterion currently in use: BIC, FPE, SIC, MDL, ... Most popular are AIC & BIC.

$$BIC = -2\log L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2) + (p+q+1)\log n.$$

$\boxed{\cdot}$ <u>The AIC vs. BIC dilemma</u>:

- BIC is <u>consistent</u> if a true model exists.
- AIC is <u>efficient</u> regardless if a true model exists.
- In practice, no true model exists, so efficiency is the only well-defined property...

$\boxed{\cdot}$ <u>Efron & Hastie (2016 p. 227)</u> Computer Age Statistical Inference.

"Covariance penalty estimates, when believable parametric models are available, should be preferred to cross-validation."

$\boxed{03/20}$ <u>Model Selection</u> (con't)

① Cross-validation (later)

② Information Criterion: AIC_c & BIC.

③ Sample ACF / PACF. (rudimentary ~ primitive)

↳ . For AR(p) & MA(q), we know the duality in the behavior of ACF / PACF:

	ACF	PACF
AR(p)	decays exp.	cuts out for $h > p$
MA(q)	cuts off for $h > q$	decays exp.

. But for ARMA (p,q), both ACF & PACF decay exponentially.

§3.7 Testing Goodness of Fit.

If data was truly generated by an ARMA (p,q) whose MLEs are $\{\hat{\phi}, \hat{\theta}, \hat{\sigma}^2\}$ and if we use the fitted model to obtain the 1-step predictors $P_{t-1}X_t$ and the corr. MSEs v_{t-1}, $t=1,\dots,n$ then the scaled residuals:

$$R_t = \frac{X_t - P_{t-1}X_t}{\sqrt{v_{t-1}}} \quad , \quad t=1,\dots,n$$

should be approximatedly dist^d as a WN $(0,1)$ sequence (asymptotic).

We can therefore assess the appropriateness of the fitted model by testing $\{R_t\}$ for WN (see §2.4. of notes).

CH 4: Nonstationary Models & Regression.

Here we examine the problem of finding an appropriate model for data that does not seem to be generated by a stationary time series.

If the data:

i) exhibit no apparent deviation from stationarity, and,

ii) have a rapidly decaying ACVF / ACF, then

we attempt to fit an ARMA model to the mean-corrected data, using the techniques of chapter 3.

If i) and ii) are not satisfied, differencing often achieves this, leading us to consider the class of ARIMA models.

§ 4.1 Autoregressive Integrated Moving Average (ARIMA)

The ARIMA broadens the ARMA class of models to include differencing.

d times

Defn: A process $\{X_t\}$ is said to be an ARIMA (p, d, q) if $(1-B)^d X_t$ is a causal ARMA (p, q). We write the model as:

↳ roots on UC (nonst.)

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t \quad , \quad Z_t \sim WN(0, \sigma^2).$$

Note: Process is stationary iff $d = 0 \rightarrow$ ARMA (p,q). Differencing X_t d times results in an ARMA (p,q) with $\phi(B)$ and $\theta(B)$ as AR & MA poly.

Recall: from CH. 1. that differencing a polynomial of degree $d-1$, d times, will reduce it to zero. We can thus add an arbitrary poly. of degree $d-1$ to $\{X_t\}$ without violating the ARIMA difference eqn:

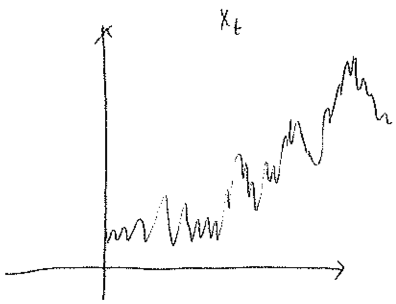
$$m_t = a_0 + a_1 t + \cdots + a_{d-1} t^{d-1}$$

$$(1-B)^d (X_t + m_t) = (1-B)^d X_t + (1-B)^d m_t = (1-B)^d X_t.$$

This means ARIMAs are useful for representing data with trends. Differencing this process will result in a stationary series.

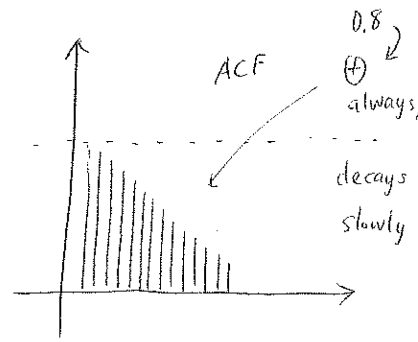
Ex/ Simulate 200 obs. from ARIMA $(1,1,0)$:

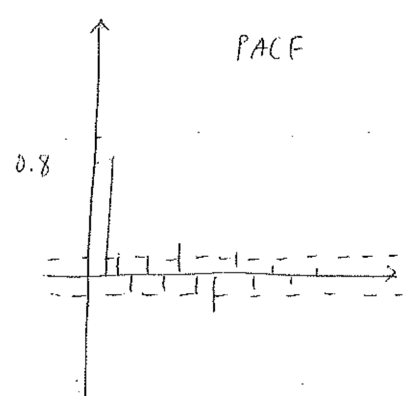
$$(1-0.8B)(1-B) X_t = Z_t \quad ; \quad Z_t \sim WN(0,1)$$



(not mean-reverting)

↳ very smooth (does not change drastically)





The slowly decaying behavior of the ACF is characteristics of ARIMAs. We would thus difference such data successively at lag 1 in the hope that for some d, $(1-B)^d X_t$ will have a rapidly decaying ACF (but be careful not to difference overly, as this can introduce correlation / dependence where none existed before).

Ex/ $X_t = Z_t \rightarrow$ WN. (don't over-difference)

correlation.

$$(1-B)X_t = X_t - X_{t-1} = Z_t - Z_{t-1} \Rightarrow Y_t = Z_t - Z_{t-1} \Rightarrow MA(1)$$

Ex/ Apply $\underbrace{(1-B)X_t}_{Y_t}$ where $\{X_t\}$ is the ARIMA (1,1,0) simulated above. Fit the ML model :

$$(1-0.8B)\underbrace{(1-B)X_t}_{Y_t} = Z_t \sim WN(0,1) \quad \text{(before)}$$

$$(1-0.787)(1-B)X_t = Z_t \; ; \; Z_t \sim WN(0, 1.012) \quad \text{(fitted model)}.$$

↳ ⊛ stable !!!

Now, fit min AIC_c AR model to X_t and we get :

$$(1-0.802B)(1-0.985B)X_t = Z_t \; ; \; Z_t \sim WN(0, 1.010)$$

Note: Coefficients of these 2 are very similar, first model is of course non-stationary. The second model is just barely stationary. In practice, it's better to fit ARIMA to data that appears to be unit-root non-stationary. (Estimation is more stable).

$\boxed{03/22}$ X_t is ARIMA (p,d,q) if $Y_t = (1-B)^d X_t \sim$ ARMA (p,q)

and $\phi(B)X_t = \theta(B)Z_t$.

⊛ Forecasting ARIMA :

→ If we denote : $(1-B)^d X_t = Y_t$, $t = 1, 2, \dots$ $\quad (\forall\, t>0)$

then under the assumption that $(X_{t-d}, \dots, X_0)$ is uncorrelated with Y_t, the BLP of X_{n+h} based on $\{X_1, \dots, X_n\}$ can be correlated to ARMA case as:

$$P_n X_{n+h} = \sum_{i=1}^{p+d} \phi_i^* P_n X_{n+h-i} + \sum_{j=h}^{q} \theta_{n+h-1,j} \left(X_{n+h-j} - P_{n+h-j-1} X_{n+h-j} \right)$$

↳ ARIMA $\underset{p+d}{\sim}$ $\underset{q}{\sim}$

As before, the $\{\theta_{n,j}\}$ are obtained with Innovations Algorirthm and the $\{\phi_i^*\}$ are the coefficients in the transformed AR polynomial:

$$\phi^*(z) = \phi(z)(1-z)^d$$

Similar results hold for MSE. For large n, these can be approx. with:

$$\tilde{\sigma}_n^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^{*2} \quad , \quad \text{where } \psi_j^* \text{ are obtained from:}$$

$$\psi^*(z) = \frac{\theta(z)}{\phi^*(z)}$$

⊛ Summary of ARMA / ARIMA Modeling Procedures:

$\boxed{1.}$ Perform preliminary transformation (if needed) to stabilize variance over time; can be done via the Box-Cox transformations:

$$f_\lambda(x_t) = \begin{cases} (x_t^\lambda - 1)/\lambda, & \text{if } x_t \geq 0 \ \& \ \lambda > 0. \\ \log x_t & \text{if } x_t > 0 \text{ and } \lambda = 0. \end{cases}$$

In practice, $\lambda = 0$ or $\lambda = 0.5$ are often adequate.

$\boxed{2.}$ Detrend / Deseasonalize the data again (if necessary) and hence to make stationarity plausible. (recall: a slowly decaying ACF is indicative of non-stationarity in a form of trends and/or unit roots).

CH.2

↳ Primary tools for achieving this are: classical decomposition & differencing.

$\boxed{3.}$ If the data looks non-stationary without a well-defined trend a seasonality, an alternative is to difference successively at lag $\underline{1}$.

$\boxed{4.}$ Examine sample ACF / PACF to look at the autocorrelation structure, to glean of potential values for p and q. (For an $AR(p)$ / $MA(q)$ we have the usual duality in ACF / PACF, decays exp. & cuts out.)

1 step in software

$\boxed{5.}$ Obtain preliminary estimates of the models for select values of p and q.

$\boxed{6.}$ Start from preliminary estimates, obtain MLE of the promising models found in step 5.

$\boxed{7.}$ From the fitted models (via ML) above, choose the one with smallest AIC$_c$ / BIC, taking into consideration also all models that are close to the minimum (within about 2 units).

$\boxed{8.}$ Can bypass steps 4–7 by using a search program that loops through all ARMA(p,q) up to given maximum ARMA orders p & q. 2 such
↳ $2^p \times 2^q$.

programs are:

- "autofit" fn in package <u>itsmr</u>. (not very good, doesn't start / really use proper step 5 techniques).
- "auto.arima" fn in package <u>forecast</u>. (very sophisticated).

$\boxed{9.}$ Inspection of the standard errors of the coefficients for the promising models in step 8, may reveal that some of them are not significant. If so, <u>subset</u> models can be fitted by constraining these to be zero.

$\boxed{10.}$ Check the candidate models for goodness-of-fit. (Tests for WN applied to the residuals.) ← end of CH. 3.

Ex/ Lake Huron data. (book ?)

$\boxed{3/25}$ § 4.2 SARIMA Models.

Often the dependence on past tends to occur most strongly at multiplies of some seasonal lag s. Eg. monthly (quarterly) economic data usually show a strong yearly component occuring at lags that are multiplies of $s = 12$ ($s = 4$).

However, it may not be reasonable to assume (as in classical decomposition models) the seasonal component as $X_t = m_t + S_t + Y_t$, i.e. S_t repeats itself the same way, cycle after cycle.

Seasonal ARIMA (SARIMA) models are extensions on the ARIMA to account for the seasonal nonstationary behavior of some series. They allow for randomness in the seasonal pattern from one cycle to another.

Defn: The process $\{X_t\}$ is a SARIMA $(p,d,q) \times (P,D,Q)_s$ with period s, if the differenced series:

$$Y_t = (1-B)^d (1-B^s)^D X_t$$

$\phi(B) = 1 - \phi_1 z - \dots - \phi_p z^p$

$\Phi(B^s) = 1 - \Phi_1 z^s - \dots - \Phi_P z^{sP}$

is a causal ARMA defined by:

$$\underbrace{\phi(B)\,\Phi(B^s)}_{\phi^*(B)} Y_t = \underbrace{\theta(B)\,\Theta(B^s)}_{\theta^*(B)} Z_t \quad , Z_t \sim WN(0, \sigma^2)$$

where $\phi(B)$ and $\Phi(B)$ are different AR polynomials of orders p and P, respectively, and $\theta(B)$ and $\Theta(B)$ are different MA polynomials of orders q and Q, respectively.

The above can be written as:

$$\phi^*(B)\, Y_t = \theta^*(B)\, Z_t$$

$p + sP$, $q + sQ$

1) • $\phi^*(B) / \theta^*(B)$ are polys of order $p + sP / q + sQ$.

2) • Provided $p < s$ and $q < s$, the constraints on the coefficients $\phi^*(\cdot)$ and

$\theta^*(\cdot)$ can be expressed as multiplicative relations :

$$\begin{cases} \phi^*_{is+j} = \phi^*_{is}\ \phi_j & \begin{cases} i = 1, 2, \ldots \\ j = 1, \ldots, s-1 \end{cases} \\ \\ \theta^*_{is+j} = \theta^*_{is}\ \theta_j & \begin{cases} i = 1, 2, \ldots \\ j = 1, \ldots, s-1 \end{cases} \end{cases}$$

(*) Motivating Example : Take $s = 12$. Idea is to try to model the monthly behavior (at multiples of lag 12) via the ARMA :

within-month model

$$\Phi(B^s)\, Y_t = \Theta(B^s)\, U_t \ , \qquad U_t \sim WN(0, \sigma^2)$$

$$\overbrace{\underline{\text{Jan}} \quad \cdots \quad \underline{\text{Dec}}}^{12 \text{ series}}$$
$$\vdots \qquad\qquad\quad \vdots$$

The 12 series for this between months-years model (one for each month) are uncorrelated.

To incorporate dependence between them, allow the process $\{U_t\}$ to follow an ARMA (p, q) :

$$\phi(B)\, U_t = \theta(B)\, Z_t \ , \qquad Z_t \sim WN(0, \sigma^2)$$

This is the model for the nonseasonal component (between month model). The multiplicative combination of these two models and allowing for differencing, leads directly to the definition of the SARIMA.

(*) SARIMA Modeling with Guidelines :

- With knowledge of s, select appropriate values of d and D in order to make $Y_t = (1-B)^d (1-B^s)^D X_t$ appear stationary (D is rarely more than 1).

- Choose P & Q so that $\hat{\rho}(hs)$ $h = 1, 2, \ldots$ is compatible with ACF of an ARMA (P, Q) (Usually P & Q are small).

- Choose p and q so that $\hat{\rho}(1), \dots, \hat{\rho}(s-1)$ is compatible with ACF of ARMA (p, q)
- Choice from among competing models should be based on AIC / BIC and goodness of fit tests.

→ A more direct approach to modeling Y_t is to simply fit a high order subset ARMA to it without making the use of the SARIMA multiplicative structure!

<u>Forecasting</u>: Analogous to ARIMA.

Expanding: $Y_t = (1-B)^d (1-B^s)^D X_t$

$$(1-B)^d = \sum_{i=0}^{d} \binom{d}{i} (-B)^i$$

$$(1-B^s)^D = \sum_{j=0}^{D} \binom{D}{j} (-B^s)^j$$

$$\Rightarrow \quad (1-B)^d (1-B^s)^D = 1 - \sum_{j=1}^{d+sD} a_j B^j \qquad \leftarrow \text{some coefficients } \{a_j\}$$

$$\Rightarrow \quad (1-B)^d (1-B^s)^D X_t = X_t - \sum_{j=1}^{d+sD} a_j X_{t-j} = Y_t$$

Setting $t = n+h$ and rearranging:

$$X_{n+h} = Y_{n+h} + \sum_{j=1}^{d+sD} a_j X_{n+h-j} \qquad (*)$$

Under assumptions: $X_{-d-Ds-1}, \dots, X_0$ uncorrelated with Y_t, $t \geq 1$, apply P_n to each side of $(*)$ to get BLP of X_{n+h}:

$$P_n X_{n+h} = P_n Y_{n+h} + \sum_{j=1}^{d+sD} a_j P_n X_{n+h-j} \qquad (**)$$

- BLP of Y_{n+h} is calculated as before, for this causal ARMA $(p+sP, q+sQ)$:

$$\phi^*(B) Y_t = \theta^*(B) Z_t$$

- $P_n X_{n+h}$ can be computed recursively from $(**)$ & $(*)$, keeping in mind that $P_n X_{n+h-j} = X_{n+h-j}$, $j \geq 1$.

For large n, we can use approximations :

$$\tilde{P}_n X_{n+h} = -\sum_{j=1}^{\infty} \pi_j^* \tilde{P}_n X_{n+h-j}$$

$$\tilde{\sigma}_n^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^*$$

where $\{\pi_j^*\}$ & $\{\psi_j^*\}$ are for the ARMA of Y_t.

$\boxed{03/29}$ Forecasting SARIMA.

. "arima" ← intercept = F default

Ex 1 Compute ARIMA BLP by hand :

Model : $\underbrace{(1-B)X_t}_{Y_t} = \mu + \theta Z_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$

$\Rightarrow$ $X_t \sim ARIMA(0,1,1)$ with mean $EY_t = \mu$. ; $Y_t \sim MA(1)$ with mean $EY_t = \mu$.

X_t ? not stationary , $Y_t \sim$ stationary

. Forecast function : $g(h) \equiv P_n X_{n+h}$ satisfies :

$Y_{t_2} = Y_t - \mu = X_t - X_{t-1} - \mu \sim MA(1)$ with mean 0 ← redefine Y_t.

$\Rightarrow$ $\underbrace{P_n Y_{n+h}}_{f(h)} = \underbrace{P_n X_{n+h}}_{g(h)} - \underbrace{P_n X_{n+h-1}}_{g(h-1)} - \underbrace{P_n \mu}_{\mu}$ $\Rightarrow$ $g(h) = f(h) + g(h-1) + \mu$.

Solve recursively :

$g(0) = P_n X_n = X_n$

$f(1) = P_n Y_{n+1} = \hat{Y}_{n+1} = $ BLP of MA(1)

. $h=1$: $g(1) = f(1) + g(0) + \mu$

. $h=2$: $g(2) = f(2) + g(1) + \mu$

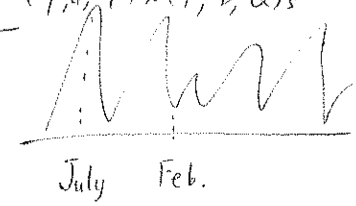
$= P_n Y_{n+2} + (\hat{Y}_{n+1} + X_n + \mu) + \mu = \dots$

$\vdots$

Note : $P_n Y_{n+h}$ = for zero-mean MA(1) is non-trivial to compute by hand, but for large n can approximate with : (e.g. :

$$\tilde{P}_n Y_{n+1} \approx \theta Y_n - \theta^2 Y_{n-1} + \dots + (-1)^{n+1} \theta^n Y_1 .$$

Ex 1 Deaths (USAacDeaths in R).   SARIMA $(p,d,q)\times(P,D,Q)_s$



- Form $Y_t = (1-B)^d (1-B^s)^D X_t$, $s = 12$ (monthly), $D = 1 = d$

  ↑ looks stationary.

- $\hat{\rho}(12)$, $\hat{\rho}(24)$, $\hat{\rho}(36)$, ... suggests an MA(1) for the between-year model ⇒ $P=0$, $Q=1$.

- $\hat{\rho}(1), \hat{\rho}(2), \ldots, \hat{\rho}(11)$ suggests an MA(1) for the btw-month model ⇒ $p=0$, $q=1$.

- Our mean-corrected proposed model for $\tilde{Y}_t = Y_t - \bar{y}$ is :

$$\tilde{Y}_t = (1+\theta_1 B)(1+\Theta_1 B^{12}) Z_t \quad \leftarrow \text{SARIMA } (0,1,1)\times(0,1,1)_{12}$$

$$\sim \text{zero-mean MA}(13) \leftarrow \text{subset MA}(13) \quad (\text{only has } 1^{st}, 12^{th}, 13^{th}).$$

- Make initial guesses for parameters :

$$\theta_1 \approx \rho(1) = \frac{\theta_1}{1+\theta_1^2} \approx \theta_1 \quad \Rightarrow \quad \hat{\theta}_1 = \hat{\rho}(1) = -0.3$$

$$\hat{\Theta} \approx \hat{\rho}(12) = -0.3.$$

Use autofit ↳ subset

Preliminary model : $\tilde{Y}_t = (1-0.3B)(1-0.3B^{12}) Z_t$

- Final model : $Y_t = 28.83 + Z_t - 0.479 Z_{t-1} - 0.591 Z_{t-12} + 0.283 Z_{t-13}$.

$Z_t \sim WN(0, 94251)$; $AIC_c = 855.5$.

- If we fit instead a subset MA(13) automatically (search over all possible subset MA models up to 13).

  → MA(13) with coefficients at lags (2, 3, 8, 10, 11) ← check significance ↳ not sign.

  ↳ constraint these to 0 and refit.

  - subset MA(13) with coefficients at lags {4, 5, 7} → not sign.
  - now set lag 9 = 0.

⇒ final model : $Y_t = 28.83 + Z_t - 0.596 Z_{t-1} - 0.406 Z_{t-6} - 0.686 Z_{t-12} + 0.459 Z_{t-13}$

$Z_t \sim WN(0, 71278)$; $AIC_c = 855.6$

- "R notes":
  - "arima" fn (tseries package)
  - "sarima" fn (astsa package) $\leftarrow$ improved front-end & diagnostics.

§ 4.3 Unit Roots in ARMA Models.

Important implications for modeling arise when an ARMA has a root on (or near) the unit circle:

- A unit root in AR polynomial suggests the data should be differenced before fitting a model (like ARMAs). $\rightarrow$ end up with ARIMA / SARIMA model.
- A unit root in MA polynomial suggests data were over-differenced, e.g.

$$\phi(B) X_t = \theta(B) Z_t \quad \sim \text{ARMA}(p,q)$$

$$\Rightarrow \quad Y_t \equiv (1-B) X_t \sim \text{ARMA}(p, q+1)$$

Proof: $X_t = \dfrac{\theta(B)}{\phi(B)} Z_t \Rightarrow Y_t = (1-B) X_t = \dfrac{\theta(B)}{\phi(B)} (1-B) Z_t$

$$\Rightarrow \phi(B) Y_t = \theta(B)(1-B) Z_t.$$

$\boxed{03/29}$ § 4.5 Unit Roots in AR Polynomials.

Let $X_1, \dots, X_n$ be observations from AR(1), $|\phi_1| < 1$ and $\mu = EX_t$:

$$X_t - \mu = \phi_1 (X_{t-1} - \mu) + Z_t, \qquad Z_t \sim WN(0, \sigma^2)$$

Test: $H_0: \phi_1 = 1$ vs. $H_1: |\phi_1| < 1$

Rewrite model in "error-correction" form:

$$\nabla X_t = X_t - X_{t-1} = \phi_0^* + \phi_1^* X_{t-1} + Z_t$$

• $\phi_0^* = \mu(1-\phi_1)$; • $\phi_1^* = \phi_1 - 1$.

Now regress ∇X_t on $(1, X_{t-1})$ and get OLS estimates $(\hat{\phi}_0^*, \hat{\phi}_1^*)$.

$\boxed{\cdot}$ Dickey-Fuller (1979) derived the limit distribution of the standardized $\hat{\phi}_1^*$ from which a test of : $H_0 : \phi_1^* = 0$ can be obtained.

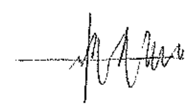
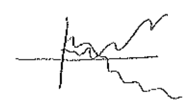
Can be extended to an $AR(p)$ with time treads:

$$X_t - \mu = \beta_0 + \beta_1 t + \phi_1 (X_{t-1} - \mu) + \dots + \phi_p (X_{t-p} - \mu) + Z_t$$

Results in the Augmented - Dickey Fuller test (ADF):

$$H_0 : \underbrace{\phi(z) = \phi^*(z)(1 - \delta z) \quad \text{with} \quad \delta = 1.}_{\phi(z) \text{ has a unit root (at least one).}}$$

Two possible alternatives are (H_1):

- Stationary : $|\delta| < 1 \iff$ causal configuration for $(1 - \delta z)$.  conv.
- explosive : $|\delta| > 1 \iff$ noncausal configuration for $(1 - \delta z)$.  trail off

Notes :

• ADF test extends to an ARMA (p, q), but the test is sensitive to correct specifications to p and q. (Although one does not have to specify p & q in the packages, just use the default which is to fit & $AR(k)$ with $k = (n-1)^{\frac{1}{3}}$.).

$\boxed{\cdot}$ Alternative tests for unit roots : proposed by Phillips & Perron (PP test) (1992) & KPSS test.
↳ widely used in econometrics, since financial data tend to be "cointegrated" (have unit roots). Caution :

H_0: model has unit root (AR poly.) $\overset{\not\Leftarrow}{\underset{\Longrightarrow}{\not\Longleftrightarrow}}$ H_0: model is non-stationary
↓
comes in many forms
(trend, seasonality, nonconst. σ^2)

In particular, if H_0 is rejected, we cannot claim stationarity (only that it has no unit roots). In R: adf.test, pp.test, kpss.test.

Unit Roots in MA Polynomials:

- Davis & Dunsmuir (1996) for MA(1): $X_t = Z_t + \theta Z_{t-1}$, $Z_t \sim iid(0, \sigma^2)$

$H_0: \theta = -1$ and $H_1: \theta > -1$. 2 tests:

(1) MLE: let $\hat{\theta}$ be MLE under H_0, then:

$$n(\hat{\theta}+1) \xrightarrow{d} C \qquad \left[\begin{array}{c}\text{non-standard dist}^{\underline{n}}\text{-} \\ \text{quantiles tabulated}\end{array}\right]$$

Reject H_0 if $\hat{\theta} > \frac{C_\alpha}{n} - 1$, $C_\alpha = 1-\alpha$ quantile of C.

(2) LRT: $$\hat{\lambda}_n = -2\log\left[\frac{L(\hat{\theta}_0 = -1, \hat{\sigma}_0^2)}{L(\hat{\theta}, \hat{\sigma}^2)}\right] \xrightarrow{d} D \qquad [\text{non-standard}]$$

Reject H_0 if $\hat{\lambda}_n > D_\alpha$, $D_\alpha = 1-\alpha$ quantile of D.

- Recently, extended to MA(q): Davis & Sing (2011). But ARMA(p,q)? nobody did!

§ 4.4 Regression with ARMA Errors.

Here we consider a generalization of the standard linear regression model, allowing for correlated errors. General model:

$$Y_t = \beta_1 X_{t,1} + \dots + \beta_k X_{t,k} + W_t, \quad t = 1, \dots, n.$$

$$\Leftrightarrow \quad \underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{W}, \quad \text{matrix form}$$

- $\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ = vector of responses.

- X = design matrix consisting of explanatory variables (covariates):

$$\underset{\sim}{X_t} = \begin{pmatrix} X_{t,1} \\ \vdots \\ X_{t,k} \end{pmatrix} \qquad (t^{th} \text{ row of } X \text{ is } \underset{\sim}{X_t}').$$

- $\underset{\sim}{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$ = regression parameters.

- $\underset{\sim}{W} = \begin{pmatrix} W_1 \\ \vdots \\ W_n \end{pmatrix}$ = error or residual vector, consisting of observations from the zero-mean ARMA (p,q):

$$\phi(B)\, W_t = \theta(B)\, Z_t \quad , \quad Z_t \sim WN(0, \sigma^2).$$

Note: in standard regression, $\{W_t\} \sim WN(0, \sigma^2)$.

- Have already seen one application of this model for estimating trend. E.g. in a model with quadratic trend we would set:

$$\left.\begin{aligned} X_{t,1} &= 1 \\ X_{t,2} &= t \\ X_{t,3} &= t^2 \end{aligned}\right\} \Rightarrow \quad Y_t = \beta_1 + \beta_2 t + \beta_3 t^2 + W_t$$

Here, each $X_{t,j}$ is a function of t only, but in general they could be any covariates (e.g., temp., pressure, ...)

$\boxed{04/01}$ Regression with ARMA Errors.

$$Y_t = \beta_1 X_{1t} + \dots + \beta_k X_{kt} + W_t \quad , \quad W_t \sim ARMA(p,q) \qquad t = 1, \dots, n$$

random

fixed

$$\Rightarrow \quad \underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{W} \quad \leftarrow \text{matrix form.}$$

How to estimate parameters? $(\underset{\sim}{\beta}, \underset{\sim}{\phi}, \underset{\sim}{\theta}, \sigma^2)$

- The OLSE (ordinary LSE) of $\underset{\sim}{\beta}$ is:

$$\hat{\beta}_{OLSE} = \arg\min\{(Y - X\beta)'(Y - X\beta)\} = (X'X)^{-1}X'Y.$$

$$= MLE \text{ if } \{W_t\} \sim N(0, \sigma^2).$$

use any g-inverce if $X'X$ is singu.

- The OLSE is also the Best (smallest variance) Linear Unbiased ($E\hat{\beta}_{OLSE} = \beta$) Estimator (BLUE) in case of uncorrelated errors $\{W_t\}$. ← By Gauss-Markov Thm.

- In the case of correlated errors (eg. $\{W_t\} \sim ARMA(p,q)$), the OLSE is still linear & unbiased, but no longer the Best!

↳ in this case, the BLUE of β is the generalized LSE:

GLSE coincides MLE of $\hat{\beta}$ if $W_t \sim$ Gaussian.

$$\hat{\beta}_{GLSE} = \arg\min\{(Y - X\beta)'\Gamma_n^{-1}(Y - X\beta)\} = (X'\Gamma_n^{-1}X)^{-1}X'\Gamma_n^{-1}Y \quad \text{where:}$$

$\Gamma_n = E(\underset{\sim}{W}\underset{\sim}{W}') =$ Covariance matrix of $\underset{\sim}{W}$. (From theory of linear models).

- If the ARMA parameters $\{\underset{\sim}{\phi}, \underset{\sim}{\theta}, \sigma^2\}$ were known, we could just use & compute $\hat{\beta}_{GLSE}$ by maximizing the Gaussian likelihood of process:

$$W_t = Y_t - \underset{\sim}{\beta}'\underset{\sim}{x}_t \qquad \underset{\sim}{x}_t = \text{appropriate vector from } X. \quad t = 1, \dots, n$$

- In practice, we don't know ARMA parameters, so we max likelihood over the entire set of parameters: $\{\underset{\sim}{\beta}, \underset{\sim}{\phi}, \underset{\sim}{\theta}, \sigma^2\}$ and the orders p & q.

- Can do this by max. the reduced likelihood $L(\hat{\beta}, \hat{\phi}, \hat{\theta})$ because σ^2 can be profiled out. $\underset{\sim}{W} \sim N(\underset{\sim}{0}, \Gamma_n)$.

$$\underset{\sim}{Y} \sim N(\underset{\sim}{\beta}'\underset{\sim}{x}_t, \Gamma_n).$$

- This suggests the following procedure:

Step 0:

1) $\hat{\beta}^{(0)} = \hat{\beta}_{OLSE} = (X'X)^{-1}X'Y$
2) $W_t^{(0)} = Y_t - \hat{\beta}^{(0)}X$
3) Identify p & q and get MLE's $\hat{\underset{\sim}{\phi}}^{(0)}$ and $\hat{\underset{\sim}{\theta}}^{(0)}$.

Step 1 : 1) $\hat{\beta}^{(1)} = \hat{\beta}_{GLSE} = (X'\Gamma_n^{-1}X)^{-1}X'\Gamma_n^{-1}Y$

$f(\hat{z}^{(0)}, \hat{z}^{(1)}, \hat{\sigma}^2)$

2) $W_t^{(1)} = Y_t - \hat{\beta}^{(1)} x_t$

3) Compute MLEs of $\hat{\underset{\sim}{\phi}}^{(1)}$ and $\hat{\underset{\sim}{\theta}}^{(1)}$.

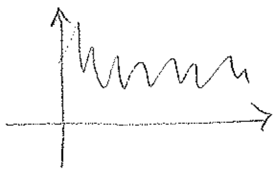
Step 2 : Same as step 1 with new $\hat{\underset{\sim}{\phi}}^{(1)}$, $\hat{\underset{\sim}{\theta}}^{(1)}$.

⋮

Stop : Stop when there is "no change" ← convergent. (in $\hat{\beta}$ from previous step).

↳ usually just a few iterations suffice. Algorithm will conv. under appro. conditions.

Ex 1 Lake Heron.



Investigate if there's a declining trend in lake levels.

✓ Yes!

$Y_t = \beta_0 + \beta_1 t + W_t$.

Use arima fn in R with "xreg" option:

- Need to look at W_t residuals from OLSE fit (MA(2) is identified using AICc).
- Then use "arima" again with chosen p & q.

Final model:

$$Y_t = 10.091 - 0.022t + 1.004 W_t - 0.290 W_{t-2} + Z_t \quad , \quad Z_t \sim WN(0, 0.457).$$

↓ 0.081 = se

under regularly condition   significant, ≠ 0 ⇒ ↓ lake level

95% for β_1 : $-0.022 \pm 1.96(0.081) = (-0.038, -0.006)$

There's enough evidence of a declining trend in lake Heron level. ← large n.

$\boxed{04/03}$ Notes { Regression with ARMA errors }.

- Model : $Y_t = \beta' X_t + W_t$, $W_t \sim ARMA(p,q)$, $Cov(W_t) = \Gamma_n$.
  - $Var(\hat{\beta}_{OLSE}) = (X'X)^{-1} X'\Gamma_n^{-1} X \, (X'X)^{-1} \geq (X'\Gamma_n^{-1} X)^{-1} = Var(\hat{\beta}_{GLSE})$.
  - The algorithm for iteratively finding the MLEs is known as Iteratively ReWeighted Least Squares (IRWLS). Also common in longitudinal & spatial modeling. Is known as Feasible GLS (FGLS) in econometrics.

- Theory extends straightforwardly to nonlinear regression with ARMA errors:

$$Y_t = \underbrace{f(\beta, X_t)}_{\text{any fn of } \beta \text{ and covariates } X_t.} + W_t, \qquad t = 1, \dots, n$$

- Asymptotics: If W_t is a causal & invertible ARMA, $Z_t \sim iid(0, \sigma^2)$, and a few other mild regularity conditions, the entire vector of MLEs $\{\hat{\beta}, \hat{\phi}, \hat{\theta}, \hat{\sigma}^2\}$ is asymptotically normal with a covariance matrix that can be approx. with:

$$2H^{-1}(\hat{\beta}, \hat{\phi}, \hat{\theta}, \hat{\sigma}^2)$$

where $H(\cdot)$ is the Hessian matrix of $-2 \log \text{Likelihood} = n \log \sigma^2 + \frac{1}{\sigma^2} \sum_{t=1}^{n} z_t^2(\beta, \phi, \theta)$

and $z_t(\beta, \phi, \sigma^2)$ are the Z_t's in the ARMA expressed in terms of W_t & lagged values as in OLSE's of ARMA's from before. (Seber & Wild, "Nonlinear Reg." 1988).

- Irregularly spaced data: All theory / methods thus far assumes we observe series at regularly spaced time intervals: $t = 1, \dots, n$.

If the observation times are irregularly spaced, a very natural way to model is to imbed the discrete time series ARMA into a continuous time ARMA or CARMA.

Ex/ Continuous AR(1) $\to$ CAR(1) process: Defined as the stationary solution of the 1st order SDE (Itô diff. eqn):

(*) $dX_t + aX_t = \sigma dB(t) + b\,dt$, $t > 0$ where :

- d : denotes the increments of its argument in time interval $(t, t+dt)$, e.g.

$$dX(t) = X(t+dt) - X(t).$$

- $B(t)$: standard Brownian motion (the continuous time analogue of WN).
- $\{a, b, \sigma\}$: parameters (to be estimated via ML).
- $X(0) \sim$ random variable with finite variance, indep. of $B(t)$, $t > 0$.
- The solution of (*) is :

$$X(t) = e^{-at}X(0) + e^{-at}I(t) + be^{-at}\int_0^t e^{au}\,du \quad , \quad \text{where :}$$

$$I(t) = \sigma\int_0^t e^{au}\,dB(u) \quad \longleftarrow \text{ an } \hat{\text{I}}\text{to integral, satisfying :}$$

- $EI(t) = 0$
- $\text{Cov}\left(I(t+h), I(t)\right) = \sigma^2\int_0^t e^{2au}\,du$ $\qquad \forall\ t \geq 0,\ h \geq 0.$

If $a > 0$, it follows that for $u \leq s \leq t$:

$X(t) \mid X(u) \sim X(u) \mid X(s) \Rightarrow$ process is Markov with conditional mean & variance :

$$E\left[X(t) \mid X(s)\right] = e^{-a(t-s)}X(s) + \frac{b}{a}\left[1 - e^{-a(t+s)}\right]$$

$$\text{Var}\left[X(t) \mid X(s)\right] = \frac{\sigma^2}{2a}\left[1 - e^{-2a(t-s)}\right]$$

This implies that it's straightforward to write down Gaussian likelihood of $\{X(t_1), \cdots, X(t_n)\}$:

$$l(a, b, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\tau_i} \phi\left(\frac{x(t_i) - \mu_i}{\tau_i}\right), \quad \phi(\cdot) = \text{pdf of } N(0,1), \text{ where:}$$

- $\mu_1 = \frac{b}{a}$, $\tau_1^2 = \frac{\sigma^2}{2a}$, and for $i = 2, \dots, n$:

  . $\mu_i = e^{-a(t_i - t_{i-1})} x(t_{i-1}) + \frac{b}{a}\left[1 - e^{-a(t_i - t_{i-1})}\right]$

  . $\tau_i^2 = \frac{\sigma^2}{2a}\left[1 - e^{-2a(t_i - t_{i-1})}\right]$

Also: if $a > 0$ and $X(0) \sim (\mu_1, \tau_1^2)$, it can be shown that $X(t)$ in ⊛ is stationary with:

$$E X(t) = \mu_1 \quad \& \quad Cov\left(X(t+h), X(t)\right) = \tau_1^2 e^{-ah}.$$

This CAR(1) naturally accommodates irregularly spaced data:

- In R: "arima" → equally spaced linear regression with ARIMA / ARMA / SARIMA errors.

"gls" → unequally spaced linear regression with ARMA / ARIMA / SARIMA errors.

"nmle" → nonlinear regression with ts errors ...

. missing data: interpolate.

$\boxed{04/05}$ Test 2: (open notes) & a few sheets (HW).

- Forecasting: $\underbrace{ARMA}_{\text{stationary}} / \underbrace{ARIMA / SARIMA}_{\text{nonstationary}}$.
- Estimation, Model selection
- Regression with ARMA errors (SARIMA / ARIMA).

} CH 3 & 4 (notes).

CH 5. Forecasting Techniques (CH10 BD)

So far: focused on models for stationary & nonstationary series, and the calculation of minimum MSE predictors based on these models.

Now, discuss some forecasting techniques that have less emphasis on an explicit model.

Why is forecasting important ?

Ex 1 . Stocking inventory : need to forecast consumption months ahead.

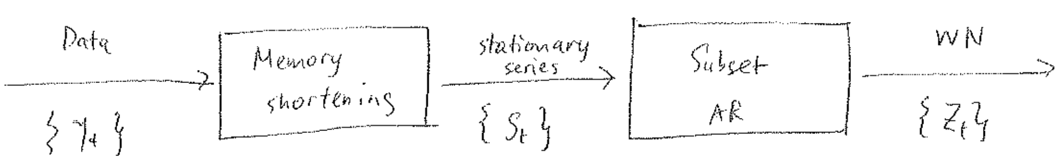
. Telecommunication routing : need to forecast traffic minutes ahead.

The following techniques that have been found in practice to be useful on a variety wide range of real datasets :

§ 5.1 The ARAR Algorithm.

This is an adaptation of the ARARMA of Newton & Parzen (1984). This algorithm has 2 steps :

- Memory shortening : reduces the data to a series that can be reasonably modeled as an ARMA (complicated 5-step procedures...)
- Fitting a subset autoregression : fit a subset AR with lags $\{1, k_1, k_2, k_3\}$, where $1 < k_1 < k_2 < k_3 \le m$, m either 13 or 26, to the memory-shortening data. The lags $\{k_1, k_2, k_3\}$ and corresponding model parameters are estimated by min σ^2 (SSE) or max Gaussian likelihood.



$$\phi(B) X_t = \theta(B) Z_t \Rightarrow X_t = \frac{\theta(B)}{\phi(B)} Z_t .$$

(Min MSE forecasts can be computed based on fitted models.)

§ 5.2 Simple Exponential Smoothing (SES, EWMA $\swarrow$ SS)

Let $\hat{Y}_{t+1} = P_t Y_{t+1} =$ BLP of Y_{t+1} based on $\{Y_t, \dots, Y_t\}$. A very simple & intuitive (ad-hoc) forecasting idea is SES, where for some $0 \le \alpha \le 1$:

$$\hat{Y}_{n+1} = \alpha Y_n + (1-\alpha)\hat{Y}_n$$

$\rightarrow$ the weighted average of current value (Y_n) and the previous forecast ($\hat{Y}_n$). α is the "smoothing" parameters. Easy to show that:

$$\hat{Y}_{n+1} = \alpha Y_n + \alpha(1-\alpha) Y_{n-1} + \alpha(1-\alpha)^2 Y_{n-2} + \dots$$

- Application: for series showing no seasonality or long-term trend, but with a locally constant mean showing "drift" over time.
- Can be shown that SES is optimal (gives min MSE forecasts) for several underlying models, e.g., the ARIMA(0,1,1):

$$X_t = (\alpha - 1) Z_{t-1} + Z_t, \quad X_t = (1-B) Y_t$$

Proof: Ex 3.38 of EWMA (SS) book.

- Estimate α by min sums of squares of 1-step prediction errors:

$$\hat{\alpha} = \operatorname{argmin} \sum_{t=2}^{n} (Y_t - \hat{Y}_t)^2, \quad \text{initial condition: } \hat{Y}_2 = Y_1.$$

- Defining $\hat{a}_{t+1}$ to be estimated level (value of Y) at time $t+1$, rewrite SES as:

$$\hat{a}_{t+1} = \alpha Y_{t+1} + (1-\alpha)\hat{a}_t.$$

And the h-step forecast based on $\{Y_1, \dots, Y_n\}$ is:

$$P_n Y_{n+h} = \hat{a}_n, \quad h = 1, 2, \dots$$

- Flat forecast: cannot cope with series showing trend or seasonality.

§ 5.3. The Holt - Winters (HW) Algorithm :

This extends SES to allow for Series with locally linear trend, but no seasonality

Basic idea : allow for a time-variant trend with following specification :

$$P_t Y_{t+h} = \hat{a}_t + \hat{b}_t h \quad , \quad h = 1, 2, \dots \quad , \quad \text{where :}$$

- $\hat{a}_t$ = estimated level at time t
- $\hat{b}_t$ = estimated slope at time t.

Like SES, take estimated level at $t+1$ to be weighted average of observations and predicted values:

$$\hat{a}_{t+1} = \alpha Y_{t+1} + (1-\alpha) P_t Y_{t+1} \quad \swarrow h=1$$

$$= \alpha Y_{t+1} + (1-\alpha)(\hat{a}_t + \hat{b}_t)$$

Similarly, estimated slope at $t+1$ given by : $\hat{b}_{t+1} = \beta(\hat{a}_{t+1} - \hat{a}_t) + (1-\beta)\hat{b}_t$

With the initial conditions : $\hat{a}_2 = Y_2$ and $\hat{b}_2 = Y_2 - Y_1$ and by choosing α, β to minimize :

$$\sum_{t=3}^{n} (Y_t - P_{t-1} Y_t)^2$$

the recursion can be solved for α & β, so that the forecast fn is :

$$P_n Y_{n+h} = \hat{a}_n + \hat{b}_n h \quad , \quad h = 1, 2, \dots$$

- Holt - Winters is optimal when true model is an ARIMA (0, 2, 2).

$\boxed{04/08}$ CH5. Forecasting Methods. (Algorithms).

5.1. ARAR Algorithm (use for anything)

5.2. SES (no trends, no seasonality) ← optimal ARIMA (0, 1, 1)

5.3. Holt - Winters. (trend, no seasonality) ← optimal ARIMA (0, 2, 2)

§ 5.4. Seasonal Holt - Winters (SHW).

- Generalization of HW to accomodate seasonality, and of course trend (competitor of ARAR - handle all). (period: d) Forecast function:

$$P_t Y_{t+h} = \underbrace{\hat{a}_t}_{\text{level}} + \underbrace{\hat{b}_t}_{\text{slope}} h + \hat{c}_{t+h} \quad , \quad h = 1, 2, \ldots$$

$\hat{c}_{t+h}$: estimated seasonal component at time t : $\hat{c}_t$.

- With same recursions for $\hat{b}_t$ as in HW, modify recursions for $\hat{a}_t$ as:

$$\hat{a}_{t+1} = \alpha \left(Y_{t+1} - \hat{c}_{t+1-d} \right) + (1-\alpha)\left(\hat{a}_t + \hat{b}_t \right).$$

- Add an additional recursion for $\hat{c}_t$:

$$\hat{c}_{t+1} = \gamma \left(Y_{t+1} - \hat{a}_{t+1} \right) + (1-\gamma)\, \hat{c}_{t+1-d}.$$

- With analogous initial conditions as in HW, to start off the recursions, once again choose $\{\alpha, \beta, \gamma\}$ to minimize sum of squares of 1-step prediction errors, we get:

$$P_n Y_{n+h} = \hat{a}_n + \hat{b}_n h + \hat{c}_{n+h} \quad , \quad h = 1, 2, \ldots$$

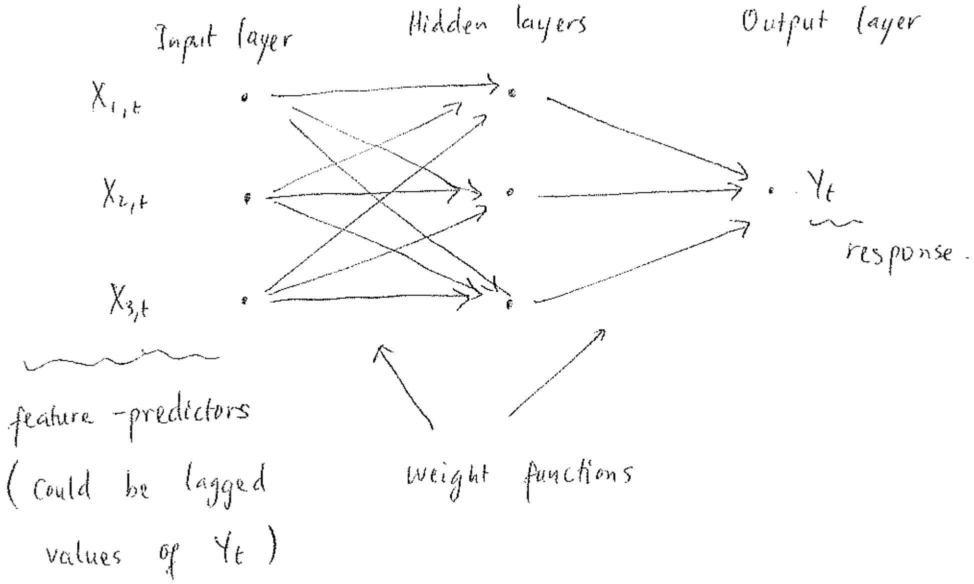
- SHW is optimal when true model is a complicated SARIMA.

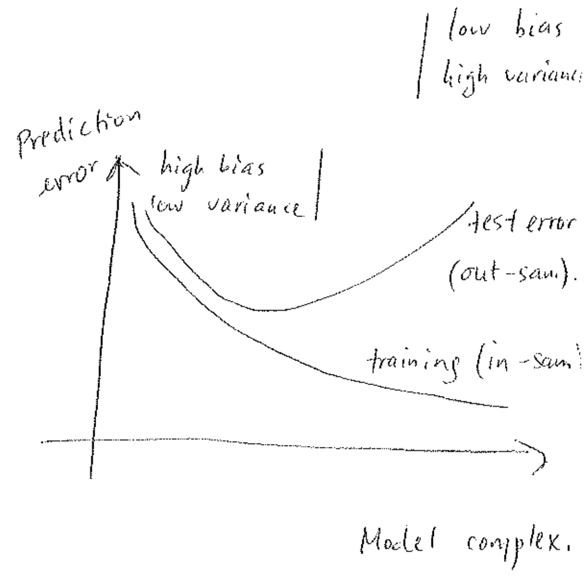
In R:
- Use "Holt Winters" for SES, SHW, HW.
  (see example 3.38 in SS for details).
- Package "itsmr" includes ARAR, HW, SHW.

§ 5.5. Some recent forecasting methods. (Machine - Learning based)

- New approaches to forecasting where relationships between variables are modeled in deep and "layered" hierarchies.
- "Deep learning" (a neural network with many deep layers) has the ability to exploit complex non-linear relationship among variables.

• Neural Network (NNs)





- Recurrent NNs (RNNs): sole objective is to predict ts data, previous information is stored in hidden layers; but not capable of storing too much past info.
- long-short term memory (LSTM): A special type of RNN that solves the memory storage problem in hidden layers (1997 paper in "neural computation").

These "deep architectures" require heavy regularization (shrinkage / thresholding) because of large number of parameters (layers & nodes). → in R: library "keras".

§ 5.6. Choosing a Forecasting Algorithm. Difficult question!

• Real data doesn't really follow any model, so smallest MSE forecasts may not in fact have smallest MSE (we assume fitted model was true model).

• Some general advice can be given:

First, choose a measure of forecast accuracy based on forecast error: $e_t = Y_t - \hat{Y}_t$

Common measures:

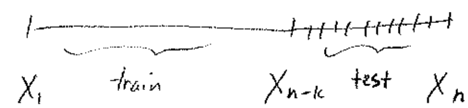
- Mean absolute error (MAE) $= \text{mean}(|e_t|)$.
- Root mean squared error (RMSE) $= \sqrt{\text{mean}(e_t^2)}$.

- For comparing forecasts among different datasets need scale-free measures :

  e.g. percentage error = $\frac{e_t}{Y_t} \times 100\%$

Now form training & test sets from available data to evaluate forecast accuracy. The size of test set (k) is typically about 20% of the sample size (n). But this depends on how large n is and how far ahead you want to predict (h = forecast horizon). The size of test set should ideally be at least as large as h $(k \geq h)$.

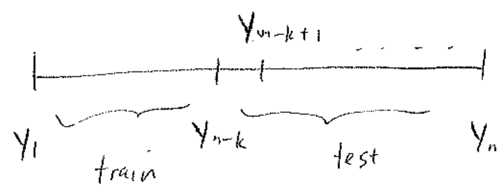
Now proceed as follows :



04/10   Choosing Forecasting Algorithm .

First, form training set (size $= n-k$) & test sets (size $= k$). → total sample size $= n$.

Proceed as follows :



Y : original data

- use Models /methods to forecast observations in test sets.

$$\{ P_{n-k} Y_{n-k+1}, \ldots, P_{n-k} Y_n \}$$

- Use a measure of forecast accuracy , e.g.

$$RMSE = \sqrt{\frac{1}{k} \sum_{i=1}^{k} \left(Y_{n-k+i} - P_{n-k} Y_{n-k+i} \right)^2}$$

and select as " best " the method that gives the smallest MSE .

- Use the best method identified above on the complete dataset: training + test, to obtain the desired out-of-sample forecasts.

A more sophisticated method is to use a <u>rolling forecast</u>.

- Choose a base training set size k, and a forecasting horizon h that you want to train the method to predict well for. Typically $h=1$ (1-step), but with monthly data, you may want $h=12$ (12-step), ...

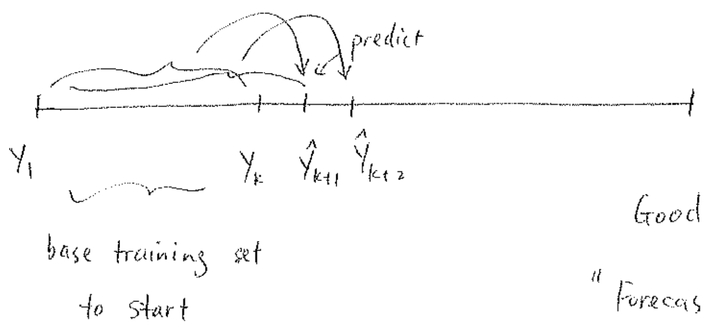
- Select obs at $t = k+i-1+h$ for test set, and use obs at $t = 1, 2, \dots, k-1$, to train the method and then compute forecast error:

$$\hookrightarrow \quad e_i = Y_{(k+i-1)+h} - P_{k+i-1} Y_{(k+i-1)+h} = Y_{(k+i-1)+h} - P_{k+i-1} Y_{(k+i-1)+h}, \quad \text{or } \bar{e}_i$$

- Repeat the above step for $i = 1, \dots, n-k-h+1$.
- The method with smallest value of forecast accuracy measure (e.g. RMSE) for $\{e_i\}$ is best:

$$\bar{e}_i = \sqrt{\frac{1}{h} \sum_{j=1}^{h} \left(Y_{(k+i-1)+j} - P_{k+i-1} Y_{(k+i-1)+j} \right)^2}$$

Illustration:



Ex/ ($h=1$)

Good book for forecasting:

"Forecasting: Principles & Practice" (2013)

$e_1 = Y_{k+1} - \hat{Y}_{k+1}$

$e_2 = Y_{k+2} - \hat{Y}_{k+2}$ $\rightarrow$ $\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2}$.

$\vdots$

Ex/ (USA Acci. Deaths = R.). $n = 78$ seasonality involved

Implement the simple CV procedure (non-rolling). With $k=6$ (train algorithm to predict well 6-step ahead).

Using $\underbrace{\{Y_1, \dots, Y_{72}\}}_{\text{train}}$, predict $\underbrace{\{Y_{73}, \dots, Y_{78}\}}_{\text{test}}$.

Forecasting Method: $\hookrightarrow$

	RMSE
HW	1143
SARIMA (4,2)	583
Subset MA(13)	501
SHW	401
ARAR	253 ✓✓

$\hookrightarrow$ best!

CH6. Spectral Analysis.

linear process: $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$

Stationary process $\{X_t\}$, zero-mean, ACVF $\gamma(h)$.

Idea: Decompose X_t into a sum of sinusoidal components with uncorrelated random coefficients:

$$X_t = \sum_{j=1}^{k} \left[A_j \cos(2\pi w_j t) + B_j \sin(2\pi w_j t) \right] \quad , \text{ where :}$$

. analogous to Fourier series decomposition

- $\{w_j\}$ = distinct frequency / frequencies.
- Frequency = $w \in \left[0, \frac{1}{2}\right]$, measured in number of cycles per unit time. For discrete data, need at least 2 points to determine a cycle, and so highest possible value of w is 0.5 cycles / unit time.
- Period: units of time per cycle $\Rightarrow T = \frac{1}{w}$.
- Coefficients: $\{A_j, B_j\}$, $A_j \sim (0, \sigma_j^2) \sim B_j$, $j = 1, \dots, k$; A_j & B_j: uncorr. (over j)

Thm: Every zero-mean stationary process can be expressed as a superposition of infinitely many uncorrelated sinusoids with frequency $\{w_j\} \in [0, 0.5]$ $(k = \infty)$.

This is analogous to Fourier series decomposition for deterministic functions.

- Can show that:

$$EX_t = 0$$

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

$$\Rightarrow \operatorname{Var}(X_t) = \gamma(0) = \sum_{j=1}^{k} \sigma_j^2 .$$

Note: Some texts with frequency

$$\lambda = \frac{\text{radians}}{\text{time}}$$

$$\Rightarrow \boxed{2\pi w = \lambda} \quad (BD)$$

SS = w; BD = λ (books).

$\boxed{04/15}$

§6.1 The Spectral Density

If $\gamma(\cdot)$ satisfies $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, which includes all ARMAs, then it can be represented as:

$$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i w h} f(w)\, dw \qquad h = 0, \pm 1, \pm 2, \ldots \qquad (*)$$

where $f(w)$ is the spectral density function, satisfying :

i) $f(w) = f(-w)$ i.e. even function

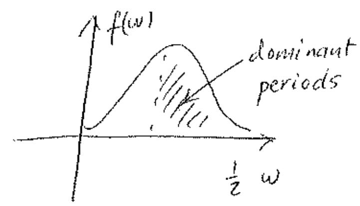
ii) $f(w) \geq 0$

iii) $\gamma(0) = \int_{-1/2}^{1/2} f(w)\, dw$

Thus, $f(w)$ acts like a pdf on $\left[-\frac{1}{2}, \frac{1}{2}\right]$, except that it integrates to $\gamma(0)$, not 1.
Also, because it's even, it's sufficient to plot it on $\left[0, \frac{1}{2}\right]$.

Thm 1 : $f(w)$ is the spectral pdf of a stationary process with ACVF $\gamma(\cdot)$ w/ representation :

$$f(w) = \sum_{h=-\infty}^{\infty} \gamma(h)\, e^{-2\pi i w h}\, I\left(-\frac{1}{2} \leq w \leq \frac{1}{2}\right)$$



iff $\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i w h} f(w)\, dw \quad (*)$ reminder holds. $\qquad (**)$

Remarks :

- Not every stationary process has a spectral pdf, but every stationary process has a spectral cdf :

$$F(w) = \int_{-\frac{1}{2}}^{w} dF(\lambda) = \int_{-1/2}^{w} f(\lambda)\, d\lambda \quad , \text{ if } f(\lambda) \text{ exists.}$$

$F(\cdot)$ behaves like a cdf, except that $F\left(\frac{1}{2}\right) = \gamma(0)$, not 1.

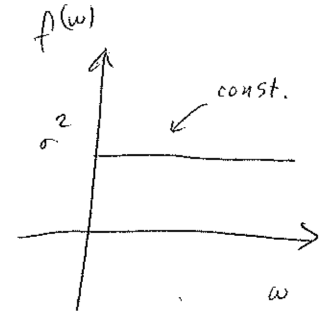
$\Rightarrow$ We can express $(*)$ in terms of spectral cdf : (*) !

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i w h}\, dF(w) = \int_{-1/2}^{1/2} e^{2\pi i w h} f(w)\, dw \quad , \text{ if } f(\cdot) \text{ exists.}$$

. The equivalence btw $\gamma(\cdot)$ and $f(\cdot)$ given by Thm 1 means that stationary processes equivalently be analyzed in the spectral or <u>frequency domain</u>, instead of in the <u>time domain</u> like we have done thus far.

<u>Ex/</u> Spectral pdf of WN : $\gamma(h) = \begin{cases} \sigma^2 & h=0 \\ 0 & \text{o.w.} \end{cases}$

$$X_t \sim WN(0,\sigma^2) \Rightarrow f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i\omega h} = \gamma(0) = \sigma^2 .$$



No dominating frequency (all the same).

Spectral pdf is flat, WN contains all frequencies in equal amounts.

<u>Thm 2</u> : If $\{X_t\}$ is a zero-mean stationary ARMA process, not necessarily causal nor invertible :

$$\phi(B) X_t = \theta(B) Z_t \quad , \quad Z_t \sim WN(0,\sigma^2)$$

then its spectral pdf is given as : fn of $e^{-2\pi i\omega}$

$$f(\omega) = \frac{|\theta(e^{-2\pi i\omega})|^2}{|\phi(e^{-2\pi i\omega})|^2} \sigma^2$$

<u>Remarks</u> :

- Use of following trig identities :

$$e^{i\alpha} = \cos\alpha + i\sin\alpha \quad \begin{cases} \cos\alpha = \dfrac{e^{i\alpha}+e^{-i\alpha}}{2} \\ \sin\alpha = \dfrac{e^{i\alpha}-e^{-i\alpha}}{2i} \end{cases}$$

ARMA spectral pdf is a ratio of trig polynomials.

- Complex conjugacy : $z = x+iy \Rightarrow \bar{z} = x-iy$; $\bar{\bar{z}} = z$, $|z|^2 = z\bar{z} = x^2+y^2$

$$\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2 , \quad \overline{(z_1+z_2)} = \bar{z}_1 + \bar{z}_2$$

- $\overline{1-ae^{ib}} = 1-a\overline{e^{ib}} = 1-ae^{-ib}$. (A)

Ex1 MA(1) : $X_t = \theta Z_{t-1} + Z_t$. Compute $f(\cdot)$ in two ways:

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma^2, & h=0 \\ \theta\sigma^2, & |h|=1 \\ 0, & \text{o.w.} \end{cases}$$

① from (**): $f(w) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i w h}$

$$= \gamma(0) + 2\sum_{h=1}^{\infty} \cos(2\pi w h)\gamma(h) + \underbrace{i \sum_{\substack{h=-\infty \\ h\neq 0}}^{\infty} \sin(2\pi w h)\gamma(h)}_{0}$$

↖ sin(·) is an odd fn.

$$= \sigma^2(1+\theta^2) + 2\theta\sigma^2\cos(2\pi w) = \sigma^2\left(1+2\theta\cos(2\pi w)+\theta^2\right).$$

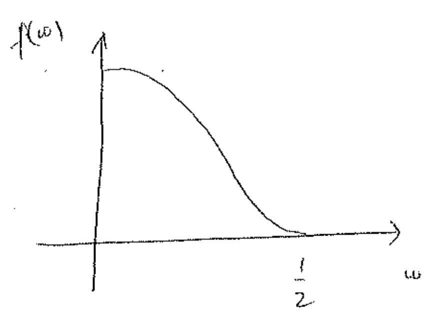
② from thm 2: $f(w) = \dfrac{|\theta(e^{-2\pi i w})|^2}{|\phi(e^{-2\pi i w})|^2}\sigma^2$ $\qquad \theta(z) = 1+\theta z$, $\phi(z) = 1$

$$= \frac{|1+\theta e^{-2\pi i w}|^2}{1}\sigma^2 = (1+\theta e^{-2\pi i w})\overline{(1+\theta e^{-2\pi i w})}\sigma^2$$

$$= (1+\theta e^{-2\pi i w})(1+\theta e^{2\pi i w})\sigma^2 = \Big(1+\theta\underbrace{(e^{-2\pi i w}+e^{+2\pi i w})}_{2\cos(2\pi w)}+\theta^2\Big)\sigma^2$$

$$= \sigma^2\left(1+2\theta\cos(2\pi w)+\theta^2\right).$$

$\theta = 0.9$

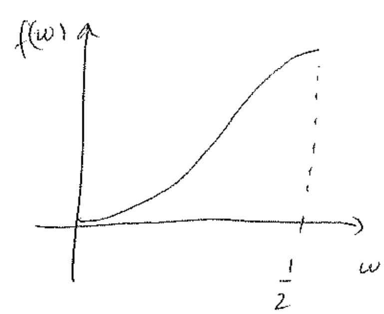


- $\gamma(1)$ large, positive
- series is smooth
- low frequencies dominate

⇒ series is smooth

- Spectrum peaks at 0

$\theta = -0.9$



- $\gamma(1)$ large, negative
- series is rough, fluctuates rapidly
- high frequencies dominate

⇒ series is rough

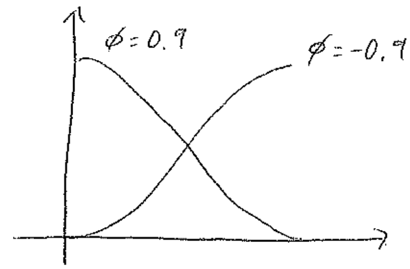
- spectrum peaks at $\frac{1}{2}$.

$\boxed{4/17}$ **Thm 2**: Spectral pdf of ARMA: $f(\omega) = \dfrac{|\theta(e^{-2\pi i \omega})|^2}{|\phi(e^{-2\pi i\omega})|^2}\sigma^2$

Ex/ MA(1). (last time)

Ex/ AR(1): $X_t = \phi X_{t-1} + Z_t \Rightarrow f(\omega) = \dfrac{\sigma^2}{1+\phi^2 - 2\phi\cos(2\pi\omega)}$ (Hwk 4.6 SS)

$\phi = 0.9$	$\phi = -0.9$
$\gamma(1)$ large positive	$\gamma(1)$ large negative
• series smooth	• series rough
• low freq. dominate	• high freq. dominate
• $f(\omega)$ peaks at 0	• $f(\omega)$ peaks at $\frac{1}{2}$



Ex/ YW AR(2) model fitted to the mean corrected sunspots data from Ch. 3.

$$X_t = \underbrace{1.318}_{\phi_1} X_{t-1} \underbrace{-0.6341}_{\phi_2} X_{t-2} + Z_t$$

Spectral pdf: $f(\omega) = \dfrac{\sigma^2}{1+\phi_1^2+\phi_2^2 - 2\phi_1\cos(2\pi\omega) - 2\phi_2\cos(4\pi\omega) + 2\phi_1\phi_2\cos(2\pi\omega)}$

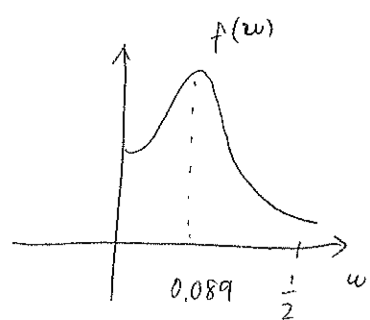
• If $x = \cos(2\pi\omega)$, we can show that for $\phi_2 < 0$, the denominator is minimum at

$\hat{x} = \dfrac{\phi_1\phi_2 - \phi_1}{4\phi_2}$. $\Rightarrow$ $f(\omega)$ is maximum at:

$$\omega = \frac{1}{2\pi}\cos^{-1}(\hat{x}) \approx 0.56 \cdot \frac{1}{2\pi} = 0.089 \Rightarrow T = \frac{1}{\omega} = 11.2$$

$\Rightarrow$ $\approx$ 11-year cycle in the series.

An example of a parametric estimator of $f(\omega)$.



§6.2 The Periodogram.

Defn: Given data $\{x_1, \ldots, x_n\}$ the discrete Fourier transform (DFT) at Fourier fundamental frequency $w_j = \frac{j}{n}$ is defined to be:

$$d(w_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{-2\pi i w_j t}, \qquad j = 0, \ldots, n-1$$

Then, the periodogram is defined as: $I(w_j) = |d(w_j)|^2, \quad j = 0, \ldots, n-1.$

Notes:

- $I(w)$ is symmetric about $w = \frac{1}{2}$, so only plot or complete it for $j = 0, \ldots, \left[\frac{n}{2}\right]$
- It can be shown that for $j \neq 0$:

$$I(w_j) = \sum_{|h|<n} \hat{\gamma}(h) e^{-2\pi i w_j h}, \qquad \hat{\gamma}(\cdot) = \text{Sample ACVF}.$$

Comparing with (**) from Thm 1:

$$f(w) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i w h}$$

We can see from this that $I(w_j)$ is an empirical estimate of $f(w_j)$.

- $I(w_j) \in \mathbb{R}$, and written as:

$$I(w_j) = \frac{1}{n}\left\{\left[\sum_{t=1}^{n} x_t \cos(2\pi w_j t)\right]^2 + \left[\sum_{t=1}^{n} x_t \sin(2\pi w_j t)\right]^2\right\}$$

can be computed quickly via the fast Fourier transform (FFT).

- $I(w_j)$ is a consistent estimator of $f(w)$, as next thm shows.

Thm: (Consistency of Periodogram).

Let $\{X_t\}$ be a linear process:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \qquad \{Z_t\} \overset{iid}{\sim} (0, \sigma^2), \qquad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

Then, for any collection of m distinct frequency $w_j \in (0, 0.5)$, $j = 1, \dots, m$, with $w_{j:n} \to w_j$, where $w_{j:n} = \frac{j_n}{n}$, and $\{j_n\}$ is a sequence such that $\frac{j_n}{n}$ is the closest Fourier frequency to w_j, we have, for $j = 1, \dots, m$:

(i) $E\, I(w_{j:n}) \to f(w_j)$, and

(ii) $\dfrac{2I(w_{j:n})}{f(w_j)} \xrightarrow{d}$ iid χ^2_2, for $j = 1, \dots, m$; provided $f(w_j) > 0$.

<u>Note</u> : (ii) can be used as an asymptotic pivot to construct a $100(1-\alpha)\%$ CI for $f(w)$:

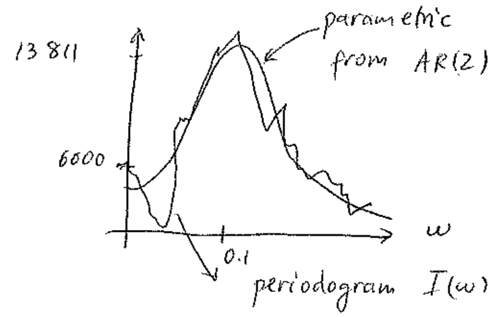
$$P\left(\frac{2I(w_{j:n})}{\chi^2_{2,1-\frac{\alpha}{2}}} \le f(w) \le \frac{2I(w_{j:n})}{\chi^2_{2,\frac{\alpha}{2}}} \right) = 1-\alpha$$

where $\chi^2_{\nu,\alpha}$ satisfies $P(\chi^2_\nu \le \chi^2_{\nu,\alpha}) = \alpha$.

<u>Ex</u> / <u>Sunspots</u> : $n = 100$, $w = \{0, 0.01, 0.02, \dots, 0.99\}$ = Fourier fundamental frequencies w_j.

$I(w_j)$: is max at $j = 10$, $w_j = 0.10$

$\Rightarrow T = \frac{1}{w} = 10$



95% CI for $f(0.10)$ is :

$$\left(\frac{2I(0.10)}{\underbrace{\chi^2_{2,0.975}}_{7.378}}, \frac{\overbrace{2I(0.10)}^{13,811}}{\underbrace{\chi^2_{2,0.025}}_{0.0506}} \right) = (3444,\ 545488)$$

$\Rightarrow$ peak not significantly different from the background (rest of spectral).

$\boxed{4/19}$ $I(w)$ = periodologram = nonparametric estimator of $f(w)$. Very rough, we want to smooth it.

$I(\omega)$ is susceptible to large uncertainty. Here we consider smoothing the periodogram $I(\omega)$ ($I(\omega) = \sum_{h=-\infty}^{\infty} \hat{\gamma}(h) e^{-2\pi i \omega h}$).

i.e. we consider estimators of form :

$$\hat{f}(\omega) = \sum_{|h|<r} w\left(\frac{h}{r}\right) \hat{\gamma}(h) e^{-2\pi i \omega h} \quad \leftarrow \text{ similar to kernel density estimation. (KPE)}$$

where $W(\cdot)$ is a weight / kernel function called the " lag window", satisfying :

i) $w(0) = 1$

ii) $|w(x)| \leq 1$ and $w(x) = 0 \;\; \forall \; |x| > 1$.

iii) $w(x) = W(-x)$.

And r is the "bandwidth".

<u>Note</u> : If $w(x) = 1$ for $|x| < 1$ and $r = n$, then $\hat{f}(\omega_j) = I(\omega_j)$.

This shows $I(\cdot)$ gives too much weight to $\hat{\gamma}(h)$ when h is large (unreliable).

The "smoothing window" / "kernel operator" is defined as :

associated

$$W(\omega) = \sum_{h=-r}^{r} w\left(\frac{h}{r}\right) e^{-2\pi i \omega h}$$

and determines which part of the spectrum will be used to get $\hat{f}(\omega) = W(\omega) \hat{\gamma}(h)$.

- <u>Rectangular lag window : (Dirichlet kernel)</u>.

$$w(x) = \mathbb{1}\, I(|x| \leq 1) = \begin{cases} 1 , & |x| \leq 1 \\ 0 , & o.w. \end{cases}$$

$$\Rightarrow \quad W(\omega) = \frac{\sin(2\pi r + \pi)\omega}{\sin(\pi \omega)}$$

- <u>Sinc lag window (Daniell kernel) :</u> $\quad w(x) = \dfrac{\sin(\pi x)}{\pi x} I(|x| \leq 1)$

$$\Rightarrow \quad W(\omega) = \begin{cases} r , & |\omega| \leq \frac{1}{2r} \\ 0 & o.w. \end{cases}$$

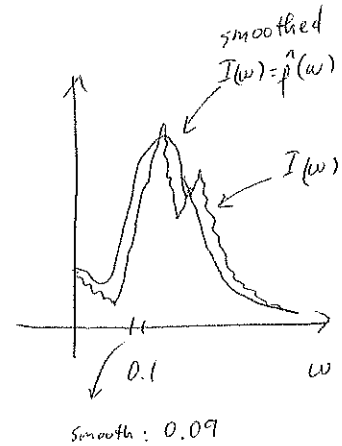
. Barlett / Fejer lag window / kernel :

$$w(x) = (1-|x|)\, I(|x| \le 1) \Rightarrow W(\omega) = \frac{\sin^2(\pi r \omega)}{r \sin^2(\pi\omega)} .$$

Ex/ (Smooth sunspots) AR(2) fit.

Daniell kernel with orders c(1,2).

In R, > spec.pgram(Sunspots, taper = 0, log = "no", kernel = "modified-daniell", c(1,2))



$\Rightarrow$ period $= T = \dfrac{1}{0.09} = 11.1$.

§6.3 Time Invariant linear Filters (TILF).

Defn. The linear process $\{Y_t\}$ is the output of a TILF applied to an input process $\{X_t\}$ if :

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} , \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \qquad (*)$$

Note : A linear process is a TILF where $X_t \sim WN(0, \sigma^2)$.

Notes : . Coefficients $\{\psi_j\}$ (impulse response fn) are time-invariant & causal if $\psi_j = 0$ for $j < 0$.

. $\underset{\sim}{\psi}(e^{-2\pi i\omega}) = \sum_{j=-\infty}^{\infty} \psi_j e^{-2\pi i \omega j}$ = frequency response or transfer fn.

$|\underset{\sim}{\psi}(e^{-2\pi i \omega})|^2$ = power transfer fn = squared frequency response fn.

. TILF theory is important. We have already used these filters extensively in course, e.g. : - smoothing data : MA filter, exponential smoothing, smoothed by elimination of high frequency components, etc.

- removal of trend & seasonality by differencing.

Thm : The Spectral densities of X_t & Y_t in ⊛ are related by :

$$f_Y(\omega) = \left| \psi(e^{-2\pi i\omega}) \right|^2 f_X(\omega).$$

Note : This was used to derive the result for the $f(\omega)$ of ARMA (p,q).

Ex/ SOI = southern oscillation index (SS).

$X_t \sim$ SOI monthly index for years 1950-1987, measuring changes in air pressure related to sea surface temperature in Central Pacific. (measure El Niño effect).

Consider 2 following TIFLs :

- $Y_t = (1-B)X_t = X_t - X_{t-1} = \sum_j \psi_j X_{t-j}$ where $\psi_j = \begin{cases} 1 & , j=0 \\ -1 & , j=1 \\ 0 & , \text{o.w.} \end{cases}$

- $Y_t = \frac{1}{24}\left(X_{t-6} + X_{t+6}\right) + \frac{1}{12}\sum_{k=-5}^{5} X_{t-k}$ where $\psi_j = \begin{cases} \frac{1}{24} & , j=\pm 6 \\ \frac{1}{12} & , -5 \le j \le 5 \\ 0 & , \text{o.w.} \end{cases}$

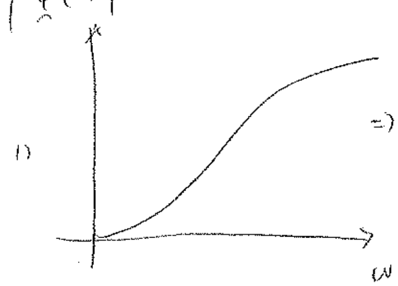
Compute corresponding transfer function :

1) $\psi(e^{-2\pi i\omega}) = \sum_{-\infty}^{\infty} \psi_j e^{-2\pi i\omega j} = 1 - e^{-2\pi i\omega}$

Fig. 4.17 SS.

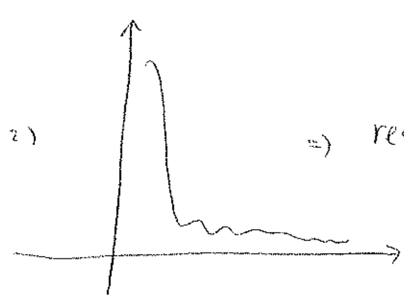
2) $\psi(e^{-2\pi i\omega}) = \sum_{-\infty}^{\infty} \psi_j e^{-2\pi i\omega j} = \frac{1}{12}\left[1 + \cos(12\pi\omega) + 2\sum_{k=1}^{5} \cos(2\pi\omega k)\right]$

$|\psi(\cdot)|^2$

1) 

$\Rightarrow$ result looks rough

eliminate low freq (high-pass)

2) 

$\Rightarrow$ result looks smooth

eliminate high freq. (low-pass)

04/24 CH.7 Extra Topics.

§7.1 Long Memory Models.

- The ACF of a causal ARMA converges rapidly to zero, in following sense:

$\rho(h) = o(r^{-h}) \Rightarrow r^h \rho(h) \to 0$ as $h \to \infty$, for some $r > 1$.

Ex/ Recall general solution for $\rho(h)$ from Ch. 3:

$$\rho(h) = \frac{1}{\gamma(0)} \sum_{i=1}^{p} \alpha_i \xi_i^{-h}, \quad \xi_i = i^{th} \text{ root of } \phi(z) \quad \left(\text{Causal} \Rightarrow |\xi_i| > 1\right).$$

$$\Rightarrow \rho(h) \approx \frac{c}{\xi^h} < \frac{c}{r^h} \quad \text{for some } 1 < r < |\xi|. \Rightarrow \left(-1 < \frac{r}{\xi} < 1\right).$$

$\left|\frac{r}{\xi}\right| < 1$

$$\Rightarrow r^h \rho(h) \approx c\left(\frac{r}{\xi}\right)^h \longrightarrow 0 \quad \text{because } -1 < \frac{r}{\xi} < 1 \text{ condition.}$$

Thus: ARMA $\rho(h) \to 0$ exponentially fast. (an ARMA has "short-memory").

- We can define a stationary process with slower decreasing ACF, one for which:

(*) $\rho(h) = O(h^{2d-1}) \Rightarrow h^{1-2d} \rho(h) \to c \neq 0$, as $h \to \infty$, where $0 < |d| < 0.5$.

Ex/ $\rho(h) \approx \frac{c}{h^p}$, power decay law (with $0 < p < 3$).

$$\Rightarrow h^{1-2d} \rho(h) \approx c h^{1-2d-p} \to c \iff 1 - 2d - p = 0 \iff d = \frac{1-p}{2}, \text{ which implies}$$

$0 < |d| < 0.5$.

Such processes will then have "long-memory." They are a compromise between ARMA & ARIMA:

	ARMA	ARFIMA	ARIMA
Memory	short	long	∞.

Defn: (ARFIMA Model) An ARFIMA (p, d, q) with $0 < |d| < 0.5$ satisfies:
fractionally

$$(1-B)^d \phi(B) X_t = \theta(B) Z_t \quad , \quad Z_t \sim WN(0, \sigma^2).$$

where $\begin{cases} \phi(z) \\ \theta(z) \end{cases}$ are the usual AR / MA poly. of degrees p/q with all the roots outside the unit circle.

Notes: (on ARFIMA)

- The $\rho(h)$ of an ARFIMA satisfies (*), so that it's a long-memory process.
- Note that we have:

	ARMA	ARFIMA	ARIMA
Memory	short	long	∞
d	$d=0$	$-0.5 < d < 0.5$ ($d \neq 0$)	$d = 1, 2, 3, \ldots$

- Can rewrite model in 2 equations:

fractionally integrated WN

$$\begin{cases} \phi(B) X_t = \theta(B) W_t \\ (1-B)^d W_t = Z_t \end{cases} \Rightarrow W_t \sim FI\,WN \text{ with ACVF } \gamma_W \text{ and ACF } \rho_W.$$

where $\gamma_W(0) = \sigma^2 \dfrac{\Gamma(1-2d)}{\Gamma(1-d)}$; $\rho_W(h) = \dfrac{\Gamma(h+d)\,\Gamma(1-d)}{\Gamma(h-d+1)\,\Gamma(d)} = O(h^{2d-1})$. ; $\Gamma(\cdot)$: gamma fn $\Gamma(x) = \int_0^\infty x^{x-1} e^{-x}\,dx$

- $\{X_t\}$ is a linear process, so involving that theory:

$$\gamma_X(h) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \psi_k \gamma_W(h+j-k),$$

where $\psi(z) = \dfrac{\theta(z)}{\phi(z)}$.

- Spectral density of $\{X_t\}$: $f(\omega) = \sigma^2 \dfrac{|\theta(e^{-2\pi i\omega})|^2}{|\phi(e^{-2\pi i\omega})|^2} \, |1 - e^{-2\pi i\omega}|^{-2d}$

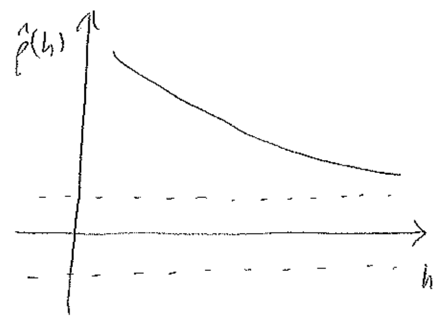
  - Parameter estimates for $\{\underset{\sim}{\phi}, \underset{\sim}{\theta}, \sigma^2, d\}$ can be obtained via maximum Gaussian likelihood, similar to ARIMA case ... (computationally intensive).

- It's common to use the simpler Whittle likelihood (spectrum-based approximation to full likelihood).

- Typical applications: hydrology (floods), climatology (glacial varves).

<u>Ex/</u> nile.itsm (B & D).

Famous hydrology series of minimum annual water level of Nile River measured at Cairo for years 622-871 ($n = 250$).



- Sample ACF suggests power decay rate ...

  $X_t = Y_t - 1119$ (mean-corrected).

- Fit the following 2 models: by best model search:

  ↳ subset ARMA: $\phi(B) X_t = \theta(B) Z_t$, $Z_t \sim WN(0, 5670)$.

$$\left.\begin{aligned} \phi(z) &= 1 + 0.2412z - 0.6563z^3 - 0.2263z^5 \\ \theta(z) &= 1 + 0.6264z + 0.2573z^2 - 0.5033z^3 \end{aligned}\right\} \text{ subset ARMA } (5,3)$$

$3\phi, 3\theta, \sigma^2$.

$-2\ell = 2873.5$ (Whittle) $\Rightarrow$ $AIC_c = 2888.0$ (k = # parameters = 7)

- full set $\overset{FI}{\text{ARMA}}$ ($p = 2 = q$):

$$(1-B)^{0.3855}(1 - 0.1694B + 0.9704B^2) X_t = (1 - 0.1800B - 0.9278B^2) Z_t$$

$$Z_t \sim WN(0, 5827)$$

$2\phi, 2\theta, \sigma^2, d$

$-2\ell = 2872.9$ (Whittle) $\Rightarrow$ $AIC_c = 2884.9$ ($k = 6$)

Recall: $AIC_c = -2\ell + \dfrac{2nk}{n-k-1}$, k = # parameters.

4/26 CH.8 State-Space Models (SSMs)

- SSMs are a rich class that include ARIMA & classical decomposition as special cases.
- Model (S&S notation): (lower case: vector, upper case: matrix).

$$\begin{cases} y_t = A_t x_t + \Gamma u_t + v_t & \{v_t\} \sim WN(0, R) \rightarrow \text{observation eqn} \\ x_t = \Phi x_{t-1} + \Upsilon u_t + w_t & \{w_t\} \sim WN(0, Q) \rightarrow \text{state eqn} \end{cases}$$

- Initial conditions: $(t = 1, \dots, n)$
  $x_0 \sim N(\mu_0, \Sigma_0)$

$\{y_t\}$ is q-dim t.s. of observations (observation variable)

$\{x_t\}$ is p-dim t.s (state variable)

$\{u_t\}$ is r-dim exogenous series (external covariate)

$\{A_t\}$ is sequence of $q \times q$ matrices (known)

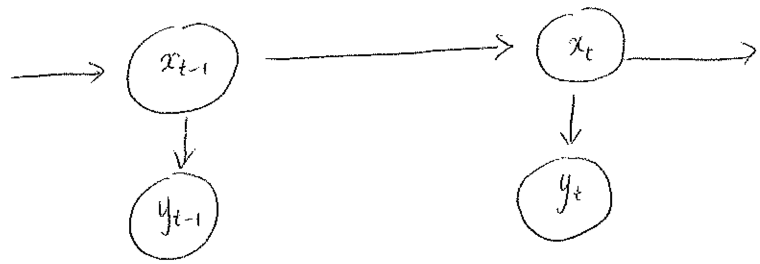
Φ is a $p \times p$ matrix of parameters

Γ & Υ: constant matrices (also parameters)

- Together with Kalman Filter, SSMs provide a flexible & computationally efficient framework, for handling: estimation, smoothing & prediction, even in the presence of missing data.
- Original idea: control engineering in context of ---> tracking a moving object:
  - State eqn: describes physical laws governing object's motion. (also called transition or system eqn).
  - observation eqn: describes the relationship between state variables & what is actually observed. (also called measurement equation).
- Other names for SSMs:

- unobserved component models
- dynamic linear models
- structural / classical decomposition models.

- Diagrammatic representation of an SSM:



- Obs. $\{y_t\}$ are independent conditional on $\{x_t\}$.
- $\{x_t\}$ is a Markov process:

$$\underbrace{\{\ldots, x_{t-1}\}}_{\text{past}} \text{ and } \underbrace{\{x_{t+1}, \ldots\}}_{\text{future}} \text{ are independent conditional on } \underbrace{x_t}_{\text{present}}$$

§ 8.1 Examples:

1. AR(1): $Y_t = \phi Y_{t-1} + Z_t$, $Z_t \sim WN(0, \sigma^2)$ ← $R = 0 \Rightarrow$ constant $\neq 0$.

let $A_t = 1$ ⟶ $\begin{cases} y_t = x_t + \cancel{v_t}, & v_t \sim WN(0, R), \; R = 0 \\ x_t = \phi x_{t-1} + w_t, & w_t \sim WN(0, \underset{Q}{\sigma^2}) \end{cases}$

$\Rightarrow$ State-space structure. $\quad y_t = x_t \sim N\left(0, \dfrac{\sigma^2}{1-\phi^2}\right)$ = stationary dist^{n}.

with initial (stationary) condition:

$$x_0 = y_0 \sim N\left(\underbrace{0}_{\mu_0}, \underbrace{\frac{\sigma^2}{1-\phi^2}}_{\Sigma_0}\right) \Rightarrow x_0 = \sum_{j=0}^{\infty} \phi^j Z_{0-j} \quad \text{and} \quad w_t = Z_t$$

2. ARMA(1,1): $Y_t = \phi Y_{t-1} + Z_t + \theta Z_{t-1}$ (causal & invertible)

Write as: $\underbrace{y_t = (\theta, 1)\begin{pmatrix} x_{t-1} \\ x_t \end{pmatrix}}_{\text{obs. eqn}}$ and $\underbrace{\begin{pmatrix} x_{t-1} \\ x_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & \phi \end{pmatrix}\begin{pmatrix} x_{t-2} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ Z_t \end{pmatrix}}_{\text{state eqn}}$

with: $x_0 = \begin{pmatrix} \sum_{j=0}^{\infty} \phi^j Z_{-1-j} \\ \sum_{j=0}^{\infty} \phi^j Z_{0-j} \end{pmatrix}$, $\Sigma_0 = ?$

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

To see this, note:

$$\left.\begin{array}{ll} ① & y_t = \theta x_{t-1} + x_t \\ ② & x_{t-1} = x_{t-1} \\ ③ & x_t = \phi x_{t-1} + Z_t \end{array}\right\}$$

From ① : $y_t - \phi y_{t-1} = \underbrace{(\theta x_{t-1} + x_t)}_{y_t} - \phi \underbrace{(\theta x_{t-2} + x_{t-1})}_{y_{t-1}}$

$$\xrightarrow{\text{from ③}} \quad = \underbrace{(x_t - \phi x_{t-1})}_{Z_t} + \theta \underbrace{(x_{t-1} - \phi x_{t-2})}_{Z_{t-1}} = Z_t + \theta Z_{t-1}.$$

3. Random walk plus noise.

Recall classical decomp. model: $X_t = \underbrace{m_t}_{\text{trend}} + \underbrace{S_t}_{\text{seasonality}} + Y_t$

In many applications primary interest is on m_t & s_t but modeling as deterministic is too restrictive. We can allow for randomness leading to notion of a "structural model." (econometrics). One such model is:

$$\begin{cases} y_t = m_t + v_t, & v_t \sim WN(0, \sigma_v^2) \\ m_t = m_{t-1} + w_t, & w_t \sim WN(0, \sigma_w^2). \end{cases}$$

$\{m_t\}$ is now a random trend / level at time t. Widely applicable, e.g. quality control, where the unobserved $\{m_t\}$ is intended to be within specified limits. Can extend to include a locally linear trend with slope β_t at time t:

$$m_{t+1} = m_t + \beta_t + w_t$$

and introduce randomness by the additional eqn:

$$\beta_t = \beta_{t-1} + u_t, \quad \{u_t\} \sim WN(0, \sigma_u^2).$$

In SSM form we have:

$$x_t = \begin{pmatrix} m_t \\ \beta_t \end{pmatrix}, \quad w_t = \begin{pmatrix} w_t \\ u_t \end{pmatrix} \Rightarrow x_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} x_{t-1} + w_t \qquad t = 1, 2, \ldots$$

↖ state eqn.

$$y_t = (1, 0)\, x_t + v_t$$

↖ obs. eqn.

$\boxed{4/29}$ SSM eqn.

$$\begin{cases} y_t = A_t x_t + \underset{\sim}{\Gamma} u_t + v_t, & v_t \sim (0, R) \quad \rightarrow \text{obs.} \\ x_t = \Phi x_{t-1} + \underset{\sim}{\Upsilon} u_t + w_t, & w_t \sim (0, Q) \quad \rightarrow \text{state.} \end{cases}$$

(con't) $\boxed{4.}$ Seasonal series plus noises. Example:

Recall $\{s_t\}$ in classical decomposition model has period d, which means:

$$s_{t+d} = s_t \quad \text{and} \quad \sum_{t=1}^{d} s_t = 0$$

Given starting values of $\{s_1, s_0, \ldots, s_{-d+3}\}$, we can generate such a sequence recursively and make it random by adding WN:

$$s_{t+1} = -s_t - \cdots - s_{t-d+2} + w_t, \quad w_t \sim WN(0, \sigma_w^2) \Rightarrow s_d = -s_{d-1} - \cdots - s_1.$$

$t = 1: \quad s_2 = -s_1 - \cdots - s_{3-d}$

$t = 2: \quad s_3 = -s_2 - \cdots - s_{2-d}$

A SSM representation is as follows:

switch notation : $s_t \to y_t$

(def. of state variable)

$$x_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-d+2} \end{pmatrix}$$

Obs. eqn. is : $y_t = (1, 0, \ldots, 0)\, x_t \iff y_t = y_t$ (vacuous)

State eqn. is :

State eqn:

$$x_t = \begin{pmatrix} -1 & -1 & \cdots & -1 & -1 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} x_{t-1} + w_t$$

→ vacuous portion

Could also mix a random trend & random seasonality.

$$y_t = m_t + v_t \quad \to \text{obs.}$$

$$\left.\begin{aligned} m_{t+1} &= m_t + w_{t+1} \\ s_{t+1} &= -s_t - \cdots - s_{t-d+2} + u_{t+1} \end{aligned}\right\} \to \text{state}$$

Thm: Every regression with ARMA errors can be expressed as SSM (the SSM form is generally not unique!)

Proof. $y_t = \underbrace{\Gamma u_t}_{\equiv \beta_t} + \varepsilon_t$; $\varepsilon_t \sim ARMA(p,q)$.

ARMA (p,q): $\underbrace{\phi(B)}_{p\text{-dim}} e_t = \underbrace{\theta(B)}_{q\text{-dim}} v_t$, $v_t \sim WN(0,\sigma^2)$.

SSM form: assume $p \geq q$, otherwise, set $\phi_{p+1} = \cdots = \phi_q = 0$ and use formulas below with $p \equiv q$.

$$\begin{cases} y_t = \beta_t + A x_t + v_t \\ x_t = F x_{t-1} + G v_{t-1} \end{cases}$$

or dimensions: $q \times q$, $q \times 1$, $q \times 1$ resp. for MA(q).

$$\underset{(p\times p)}{F} = \begin{pmatrix} \phi_1 & 1 & \cdots & 0 \\ \vdots & 0 & 1 \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_p & 0 & \cdots & 1 \end{pmatrix}, \quad \underset{(p\times 1)}{G} = \begin{pmatrix} \theta_1 + \phi_1 \\ \vdots \\ \theta_q + \phi_q \\ \phi_{q+1} \\ \vdots \\ \phi_p \end{pmatrix}, \quad \underset{(p\times 1)}{A} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Identity

5. Ex: Regression with AR(1) / MA(1) errors: $y_t = \beta_t + \varepsilon_t$.

AR(1)

$A = 1$, $F = \phi = G$

$$\Rightarrow \begin{cases} y_t = \beta_t + \underbrace{x_t + v_t}_{\varepsilon_t} \\ x_t = \phi(x_{t-1} + v_{t-1}) \end{cases}$$

set $\varepsilon_t = x_t + v_t$

MA(1)

$A = 1$, $F = 0$, $G = \theta$

$$\Rightarrow \begin{cases} y_t = \beta_t + \underbrace{x_t + v_t}_{\varepsilon_t} \\ x_t = \theta v_{t-1} \end{cases}$$

Set $\varepsilon_t = x_t + v_t$

$$\Rightarrow \begin{cases} y_t = \beta_t + \varepsilon_t \\ x_t + v_t = \varepsilon_t = \phi \varepsilon_{t-1} + v_t \end{cases}$$

$$\Rightarrow \begin{cases} \varepsilon_t = \phi \varepsilon_{t-1} + v_t \\ y_t = \beta_t + \varepsilon_t \end{cases} \quad ; \text{ reg. with AR(1) errors}$$

$$\begin{cases} y_t = \beta_t + \varepsilon_t \\ \varepsilon_t - v_t = \theta v_{t-1} \end{cases}$$

$$\Rightarrow \begin{cases} \varepsilon_t = \theta v_{t-1} + v_t \\ y_t = \beta_t + \varepsilon_t \end{cases} \quad ; \text{ reg. with MA(1) errors}$$

§ 8.2 The Kalman Recursions and Estimation.

Objectives of SSM: to estimate / predict values of the unobserved state vector x_t in terms of obs. $\{y_1, y_2, \dots\}$ and y_0 (which in many cases is just a constant). These estimates / predictions are min MSE as usual. These cases are defined.

"Estimation" of x_t (don't obs.) in terms of:

(i) $y_0, \dots, y_{t-1}$: is called the predicting problem.

(ii) $y_0, \dots, y_t$: is called the filtering problem.

(iii) $y_0, \dots, y_n$ $(n > t)$: is called the smoothing problem.

These can be solved recursively using Kalman Recursions (computationally efficient, a.k.a. Kalman filtering). These recursions give form of "estimates" in terms of the parameters of the SSM.

$$\Theta = \{ \Phi, \Upsilon, Q, \Gamma, A_t, R, \mu_0, \Sigma_0 \} .$$

$\boxed{05/01}$

Estimation of x_t in terms of:

i) $y_0, \cdots, y_{t-1}$ $\rightarrow$ prediction

ii) $y_0, \cdots, y_t$ $\rightarrow$ filtering

iii) $y_0, \cdots, y_n$ $\rightarrow$ smoothing.

Use Kalman Recursions, everything is a function of parameters:

$$\Theta = \{ \Phi, \Upsilon, Q, \Gamma, A_t, R, \mu_0, \Sigma_0 \}$$

$\leftarrow$ can be estimated via Gaussian ML by assuming $\{w_t\}$ & $\{v_t\}$ jointly Gaussian.

Notation: (S & S)

they use $P^{s}_{t_1,t_2}$

- $x_t^s : E(x_t \mid y_1, \dots, y_s)$

if $t_1 = t_2$. $\equiv$ BLP. $\equiv$ MSE of BLP.

- $\Omega^s_{t_1,t_2} = E\left[(x_{t_1} - x^s_{t_1})(x_{t_2} - x^s_{t_2})'\right] = \Omega^s_{t_1}$

if process is Gaussian

$= E\left[(x_{t_1} - x^s_{t_1})(x_{t_2} - x^s_{t_2})' \mid y_1, \dots, y_s\right]$ $\iff$ $\{w_t\}$ & $\{v_t\}$ are jointly Gaussian.

Thm: (Kalman - Filter)

This main result allows for both filtering & prediction. With initial conditions:

$x_0^0 = \mu_0$ and $\Omega_0^0 = \Sigma_0$, then for $t = 1, \dots, n$:

$$\left.\begin{aligned} x_t^{t-1} &= \Phi x_{t-1}^{t-1} + \Upsilon u_t \\ \Omega_t^{t-1} &= \Phi \Omega_{t-1}^{t-1} \Phi' + Q \end{aligned}\right\} \text{ use for prediction with } t > n.$$

$$\left.\begin{aligned} x_t^t &= x_t^{t-1} + K_t (y_t - A_t x_t^{t-1} - \Gamma u_t) \\ \Omega_t^t &= (I - K_t A_t)\, \Omega_t^{t-1} \end{aligned}\right\} \text{ gives filtered values at } t.$$

$$K_t = \Omega_t^{t-1} A_t' (A_t \Omega_t^{t-1} A_t' + R)^{-1} \rightarrow \text{Kalman gain.}$$

A byproduct of this are the innovations (prediction errors):

$\varepsilon_t = y_t - E(y_t \mid y_1, \dots, y_{t-1})$ with $E(\varepsilon_t) = 0$

(*) $= y_t - \underbrace{A_t x_t^{t-1} - \Gamma u_t}_{E(y_t \mid y_1, \dots, y_{t-1})}$

$Var(\varepsilon_t) = \Sigma_t = A_t \Omega_t^{t-1} A_t' + R$

$E(y_t) - A_t E(x_t^{t-1}) - \Gamma u_t = 0$

$A_t E(x_t^{t-1}) + \Gamma u_t$?

$\underbrace{Var(y_t)}_{R} + \underbrace{Var(A_t x_t^{t-1})}_{A_t \underbrace{Var(x_t^{t-1})}_{\Omega_t^{t-1}} A_t'} - 2\underbrace{Cov(\quad)}_{0}$

Notes:

- Can show $\begin{pmatrix} x_t \\ \varepsilon_t \end{pmatrix} \Bigg| y_1, \dots, y_{t-1} \sim$ Normal

$$\sim N\left[\begin{pmatrix} x_t^{t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_t^{t-1} & \Omega_t^{t-1} A_t' \\ (*)' & \Sigma_t \end{pmatrix}\right]$$

- Extends straightforwardly to time-varying parameter case:

$$\Phi \to \Phi_t, \quad \underset{\sim}{\Upsilon} \to \underset{\sim}{\Upsilon}_t, \quad Q \to Q_t, \quad R \to R_t, \quad \Gamma \to \Gamma_t.$$

- Similar thm for Kalman Smoother (for smoothing):

<u>Thm</u> (Kalman Smoother)

With initial conditions x_n^n and Ω_n^n obtained from Kalman Filter (KF), we have:

for $t = n, n-1, \ldots, 1$:

$$x_{t-1}^{n} = x_{t-1}^{t-1} + J_{t-1}\left(x_t^n - x_t^{t-1}\right)$$

$$\Omega_t^{t-1} = \Omega_{t-1}^{t-1} + J_{t-1}\left(\Omega_t^n - \Omega_t^{t-1}\right) J_{t-1}' \, .$$

$$J_{t-1} = \Omega_{t-1}^{t-1}\,\Phi'\left(\Omega_t^{t-1}\right)^{-1}.$$

- KF still hold even if the process is non-Gaussian, under the following interpretations:

$$x_t^s = E\left(x_t \mid y_1, \ldots, y_s\right) = P_{\{1, y_1, \ldots, y_s\}}\, x_t \equiv BLP \text{ of } x_t.$$

(I.e. KF gives BLPs, but under normality, BLP = BP).

- <u>Estimation of Parameters</u>: (MLE)

Assume: $x_0 \sim N(\mu_0, \Sigma_0)$, $\begin{pmatrix} v_t \\ w_t \end{pmatrix} \overset{\text{ind. \& uncorr}}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} R & 0 \\ 0 & Q \end{pmatrix}\right]$

$\Rightarrow \quad \varepsilon_t \overset{\text{ind.}}{\sim} N(0, \Sigma_t) \longleftarrow$ innovations from (A).

$\Rightarrow \quad \log \varepsilon_t = -\frac{1}{2}\log|\Sigma_t| - \frac{1}{2}\varepsilon_t'\Sigma_t^{-1}\varepsilon_t$

$\Rightarrow \quad -2\log L\left(\varepsilon_1, \ldots, \varepsilon_t \mid \Theta\right) = -\sum_{t=1}^{n}\log f(\varepsilon_t) = \sum_t \log|\Sigma_t(\Theta)| + \sum_t \varepsilon_t(\Theta)'\Sigma_t(\Theta)^{-1}\varepsilon_t(\Theta)$

$\Rightarrow$ MLE $\hat{\Theta}_n$.

<u>Optimation Algorithms</u>:
- Newton-Ralphson
- EM.

Asymptotics of MLE :

Under very general regularity conditions, we have :

$$\sqrt{n}\,(\hat{\Theta}_n - \Theta) \xrightarrow{d} N\left(\underset{\sim}{0},\ J(\Theta)^{-1}\right)$$

$$J(\Theta) = \lim_{n\to\infty} \frac{1}{n} E\left[\frac{-\partial^2 \log L}{\partial \Theta\, \partial \Theta'}\right] = \text{asymptotic Information matrix.}$$

Regularity Conditions :

- Causality : $\Leftrightarrow$ stable $\Leftrightarrow$ all eigenvalues of Φ being less than 1 in abs. value.
- w_t & $v_t \sim$ Gaussian : but can be relaxed.

$\boxed{05/03}$ §8.3 Hidden Markov Models (HMM).

- SSM : measurements are on continuous scale.
- HMM : " discrete " .

In a simple 1-dim HMM we assume states are Markov Chain, taking values on a finite state-space. $\{x_t = i\ ,\ i = 1, \dots, m\}$, with :

stationary distⁿ

$\pi_j = P(x_t = j)$

$\{x_t\}$ is latent variable (don't obs.) "state eqns"

stationary transition probabilities

$\pi_{ij} = P(x_{t+1} = j \mid x_t = i)$

for $t = 0, 1, \dots$ and $i, j = 1, \dots, n$.

The obs. are conditionally independent given the states, with corresponding distributions :

$$p_j(y_t) = p(y_t \mid x_t = j)\ ,\quad j = 1, \dots, m$$

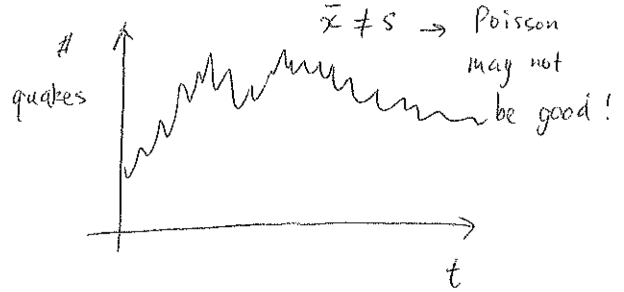
$\{y_t\}$ is the observed process "obs. eqns"

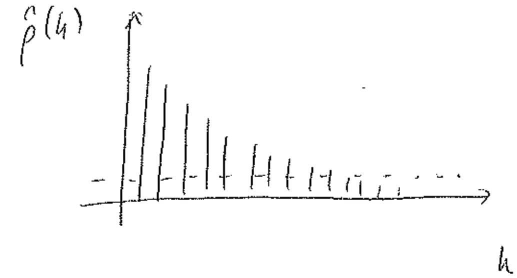
Marginally, y_t is "mixed" accordingly : law of total probability.

$$p(y_t) = \sum_{j=1}^{m} \pi_j\, p_j(y_t)$$

Ex/ Poisson HMM for # major quakes "EQcount", SS Ex/ 6.15.

EQcount = {annual # of quakes magnitude ≥ 7, 1900-2006}.





- Series is autocorrelated (AR(1)?)
- Could model marginally using a discrete dist^n. (Poisson, Negative Binomial, etc.)

but how to account for ACF?

Sol^n: let $y_t \mid x_t \sim$ Poisson and capture the autocorrelation via the latent 2-state (m-state) stationary Markov chain (MC) x_t.

Some MC Theory: The MC is characterized by its "transitioned matrix":

$$P = (\pi_{ij})_{i,j=1}^{m}, \quad \text{where} \quad \sum_{j=1}^{m} \pi_{ij} = 1.$$

The stationary (steady state) dist^n $\{\pi_j\}_{i,j=1}^{m}$, when it exists, is unique sol^n of:

$$\pi_j = \sum_{i=1}^{m} \pi_{ij}\pi_i \quad \& \quad \sum_{j=1}^{m} \pi_j = 1.$$

For m = 2 case:

$$P = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}, \quad \begin{array}{l} \pi_{11} + \pi_{12} = 1 \Rightarrow \pi_{12} = 1 - \pi_{11} \\ \pi_{21} + \pi_{22} = 1 \Rightarrow \pi_{21} = 1 - \pi_{22} \end{array}$$

Can show that the stationary dist^n always exist:

$$\pi_1 = \frac{\pi_{21}}{\pi_{12} + \pi_{21}}, \quad \pi_2 = \frac{\pi_{12}}{\pi_{12} + \pi_{21}}.$$

Now specify :

$$p_1(y) \downarrow$$

$$\left.\begin{array}{l} y_t \mid x_t = 1 \sim \text{Poisson}(\lambda_1) \\ y_t \mid x_t = 2 \sim \text{Poisson}(\lambda_2) \end{array}\right\} \quad p_i(y) = \frac{e^{-\lambda_i}\lambda_i^{\,y}}{y!}$$

$$\uparrow p_2(y)$$

k = counts

$y = 0, 1, 2, \ldots$

$E(y_t) = \lambda_i = Var(y_t)$

$$\Rightarrow p(y_t) = \begin{cases} p_1(y), & \text{if } x_t = 1 \quad (\text{occurs w.p. } \pi_1) \\ p_2(y), & \text{if } x_t = 2 \quad (\text{occurs w.p. } \pi_2) \end{cases}$$

$$y_t = Y_1 I(x_t = 1) + Y_2 I(x_t = 2).$$

$$\Rightarrow p(y_t) = \pi_1 p_1(y_t) + \pi_2 p_2(y_t) \quad \rightarrow \text{mixture dist}^{\underline{n}}.$$

via mixture

$$E(y_t) = \pi_1 EY_1 + \pi_2 EY_2 = \pi_1\lambda_1 + \pi_2\lambda_2$$

via conditioning

$$\begin{aligned} E(y_t) &= E\left[E(y_t \mid x_t)\right] \\ &= E\left[I(x_t=1)\,EY_1 + I(x_t=2)\,EY_2\right] \\ &= E\left[I(x_t=1)\,\lambda_1 + I(x_t=2)\,\lambda_2\right] \\ &= \pi_1\lambda_1 + \pi_2\lambda_2. \end{aligned}$$

In general,

$$E(y_t^k) = \pi_1 E Y_1^k + \pi_2 E Y_2^k.$$

Thus : $E(y_t) = \pi_1\lambda_1 + \pi_2\lambda_2$.

Similar calculations give ACVF for $Y \equiv y_t$ at lag h:

$$\gamma_Y(h) = \pi_1\pi_2(\lambda_1 - \lambda_2)^2(1 - \pi_{12} - \pi_{21})^h = a \cdot b^h.$$

↳ exp. decaying ACVF, consistent with data.

Smoothing & Filtering :

$\pi_j(t|s) = P(x_t = j \mid y_1, \ldots, y_s)$. Filter gives recursions for $\pi_j(t \mid t-1)$ and $\pi_j(t|t)$ starting from : $\pi_j(1|0) = \pi_j$.

Smoother " $\pi_j(t|n)$.

MLE : Θ = parameters = $\{\pi_{11}, \underbrace{\pi_{12}, \pi_{21}}_{\text{from others}}, \pi_{22}, \lambda_1, \lambda_2\}$. Based on data $\{y_1, \dots, y_n\}$:

$$L(\Theta) = \prod_{t=1}^{n} p_\Theta(y_t \mid y_1, \dots, y_{t-1})$$

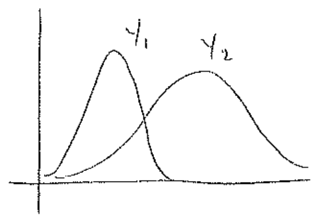
But $$p_\Theta(y_t \mid y_1, \dots, y_{t-1}) = \sum_{j=1}^{m} \underbrace{P(x_t = j \mid y_1, \dots, y_{t-1})}_{\pi_j(t \mid t-1)} \underbrace{p_\Theta(y_t \mid x = j, y_1, \dots, y_{t-1})}_{p(y_t \mid x_t = j) \equiv p_j(y_t).}$$

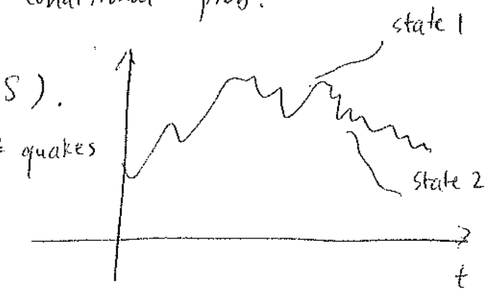
↳ total law of prob.

↳ by conditional prob.

$$\hat{\Theta} = \begin{pmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \\ \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} = \begin{pmatrix} 0.93 \\ 0.07 \\ 15.4 \\ 26.0 \end{pmatrix}$$

look at figure 6.13 (S& S).





$\boxed{05/06}$ Regression with ARMA errors & ARMAX Models.

A commonly used model in econometrics is ARMAX (ARMA with exogenous variables) — covariates

$$(*) \quad \underset{\sim}{Y_t} = \underset{\sim}{x_t}' \underset{\sim}{\beta} + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j Z_{t-j} + Z_t \quad ; \quad \{Z_t\} \sim WN(0, \sigma^2)$$

Can be shown (Thm) that : ARMAX <=> SSM

↳ proof : Trindate et. al (2016).

Recall : regression with ARMA errors :

$$(**) \quad \underset{\sim}{Y_t} = \underset{\sim}{x_t}' \underset{\sim}{\beta} + W_t \quad , \quad W_t \sim ARMA(p,q) \iff \phi(B) W_t = \theta(B) Z_t .$$

Fact : Regression with ARMA errors (**) is a special case of ARMAX.

Proof : (**) implies :

$$W_t = Y_t - x_t' \beta = Y_t - \mu_t \quad , \quad \mu_t = \underset{\sim}{x_t}' \beta \quad \text{call} : \tilde{\mu}_t .$$

$$\phi(B) W_t = W_t - \phi_1 W_{t-1} - \dots - \phi_p W_{t-p} = (Y_t - \mu_t) - \sum_{i=1}^{p} \phi_i (Y_{t-i} - \mu_{t-i})$$

$$= (\phi_1 \mu_{t-1} - \dots - \phi_p \mu_{t-p} - \mu_t) + \phi(B) Y_t .$$

$= \theta(B) Z_t$. Thus, $\tilde{\mu}_t + \phi(B) Y_t = \theta(B) Z_t$, where:

$$\tilde{\mu}_t = \sum_{i=1}^{p} \phi_i \mu_{t-i} - \mu_t$$

$$\Rightarrow \quad Y_t = \underbrace{-\tilde{\mu}_t}_{x_t' \beta} + \sum_{i=1}^{p} \phi_i \mu_{t-i} + \sum_{j=1}^{q} \theta_j Z_{t-j} + Z_t .$$

which is of the form (*) with $x_t' \beta = -\tilde{\mu}_t$.

<u>Connections</u>: $\rightarrow$ denote extension to accommodate features.

WN $\longrightarrow$ ARCH / GARCH (accommodate dependence).

$\downarrow$

(correlation) ARMA $\longrightarrow$ Regression with ARMA errors (non-stationarity via covariates)

ARMA $\searrow$ Classical decomposition (non-stationarity explicit via m_t & s_t)

$\downarrow$

(non-stat.) (via unit roots) ARIMA

$\downarrow$

(seasonal unit roots) SARIMA $\longrightarrow$ $\boxed{\text{SSM} \Leftrightarrow \text{ARMAX}}$ accommodates everything? (linearly...).

Classical decomposition $\downarrow$ $\boxed{\text{SSM} \Leftrightarrow \text{ARMAX}}$

<u>Final</u>: Mon May 13, 1:30–4:00. Open notes & open books (2).

S & S (Spectral, SSM)