PSTAT 174/274: Time Series

Prof. Dr. Gareth W. Peters

(CStat-RSS, CMath-IMA, FIOR, SIRM, FRSS, FIMA, YAS-RSE)

Professor of Statistics for Risk and Insurance &

Janet and Ian Duncan Endowed Chair in Actuarial Science

University of California Santa Barbara

January 7, 2023

UC SANTA BARBARA UCSB

PART V: Models for Linear Time Series and Properties.

Linear Process Time Series Models

White Noise (aslo called "shocks" or "innovation noise") revisited...

We know already:

- $\epsilon_t \sim WN(0, \sigma^2)$ is a TS s.t.

$$\mathbb{E}[\epsilon_t] = 0, \;\; \gamma(s,t) = \mathbb{E}[\epsilon_t \epsilon_s] = \sigma^2 \text{ if } s = t \;\&\; \gamma(s,t) = 0 \text{ if } t \neq s.$$

- $\epsilon_t \sim IID(0, \sigma^2)$ is a WN with independent observations
- If WN ϵ_t is Gaussian WN then $\epsilon_t \sim N(0, \sigma^2)$ for all t.

We can also consider the WN as the simplest linear TS Model:

- WN is also a special Moving Average model $\Rightarrow$ MA(0) model
- Moving Average of "order" 0

We will learn more about MA models shortly after we see a general specification of a linear process.

- Many forecasters are persuaded of the benefits of **parsimony**: *using as few parameters as possible!*
- Although complicated models can track data very well over the historical period for which parameters are estimated, they often perform poorly when used for out-of-sample forecasting!

The Box-Jenkins methodology for forecasting

1. **Model identification**
2. **Parameter estimation**
3. **Verification**

   Check model obtained from 1. & 2.

   - Good? Goto 4.
   - Bad? Goto 1. & decide on new model
4. **Forecasting**

Simple Temporal Models in TS settings:

Example

A non-stationary process (simplest) is given by

$$Y_t = \kappa_t + \epsilon_t = \underbrace{\beta t}_{deterministic\ \ component} + \underbrace{\epsilon_t}_{stochastic\ \ component}, \quad \epsilon_t \sim WN$$

- *set deterministic process* κ_t *as linear time trend s.t.* $\mathbb{E}[Y_t] = \beta t$ *depends on* t
- *However,* $X_t = Y_t - \beta t$ *is covariance stationary.*

Example

Another example could be a signal in noise given by a model:

$$Y_t = f(t; \boldsymbol{\theta}) + \epsilon_t$$

Note - it could be linear or non-linear in parameters of model/signal $\boldsymbol{\theta}$.

Example of a signal in noise model where we consider κ_t as the realisations of some function at times $t = t_1, t_2, \ldots$ (typically default $t = 1, 2, 3, \ldots$) for instance given by:

$$Y_t = A \cos\left(\frac{2\pi t}{p} + \theta\right) + \epsilon_t$$

- the first component is deterministic and considered a "signal" and
- the second component is stochastic, denoted $\epsilon_t \sim WN(0, \sigma^2)$ and considered observation noise or innovation error or stochastic driver or risk driver - many names given in various literature's depending on interpretation and application context.

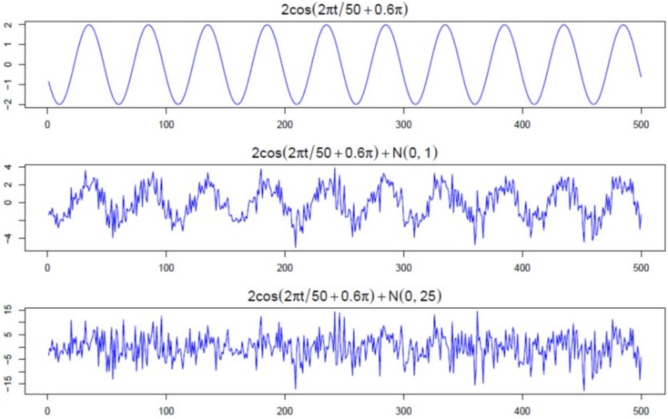
- NOTE: Many realistic models for generating time series assume an underlying signal with some consistent periodic variation, contaminated by adding a random noise.

TS models with time-trend

UC SANTA BARBARA
Prof. Gareth W. Peters

(source: Dewei Wang https://people.stat.sc.edu/wang52)

```
set.seed(100); cs = 2*cos(2*pi*1:500/50 + .6*pi); w = rnorm(500,0,1)
par(mfrow=c(3,1), mar=c(3,2,2,1), cex.main=1.5)
plot.ts(cs, main=expression(2*cos(2*pi*t/50+.6*pi)),col="blue")
plot.ts(cs+w, main=expression(2*cos(2*pi*t/50+.6*pi) + N(0,1)),col="blue")
plot.ts(cs+5*w, main=expression(2*cos(2*pi*t/50+.6*pi) + N(0,25)),col="blue")
```



Example of another type of linear TS model:

Example

Random Walk example

$$Y_t = Y_{t-1} + \epsilon_t, \;\; \epsilon_t \sim WN, \;\; Y_0 \;\; \text{constant}$$

1. *Solving recursively:*

$$Y_t = \sum_{j=1}^{t} \epsilon_j + Y_0$$

2. $\mathbb{E}\left[Y_t\right] = Y_0$ *time-invariant mean.*
3. $\mathbb{V}\mathrm{ar}\left[Y_t\right] = t\sigma^2$ *time dependent*
4. $X_t = Y_t - Y_{t-1}$ *is covariance stationary*

A very general representation of linear processes (linear time series models):

Proposition

Let $\{X_t\}$ *be a stationary time series with zero mean and covariance function* γ_X*. If* $\sum_{j=-\infty}^{\infty} |\psi_j| \leq \infty$*, then the linear process time series*

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

is stationary with zero mean and autocovariance function

$$\gamma_Y(k) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \psi_j \gamma_X(k - i + j)$$

From these general linear process forms, one can see that setting $\{X_t\} = \{\epsilon_t\}$ white noise, MA(1), AR(1), MA(q), and MA(∞) are all examples of special cases of statistical time series models:

- White noise: set $\mu = 0$ and $\psi_j = \mathbb{I}[j = 0]$ to get $Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j} = \epsilon_t \sim WN(0, \sigma^2)$,
- MA(1): set $\mu = 0$ and $\psi_j = \mathbb{I}[j = 0] + \theta \mathbb{I}[j = 1]$ to get $Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j} = \epsilon_t + \theta \epsilon_{t-1}$,
- AR(1): set $\mu = 0$ and $\psi_j = \phi^j \mathbb{I}[j \geq 0]$ to get $Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j} = \sum_{j=-\infty}^{\infty} \phi^j \epsilon_{t-j} = \epsilon_t + \phi \sum_{j=-\infty}^{\infty} \phi^j \epsilon_{t-1-j} = \epsilon_t + \phi Y_{t-1}$,
- MA(q): set $\mu = 0$ and $\psi_j = \mathbb{I}[j = 0] + \sum_{k=1}^{q} \theta_k \mathbb{I}[j = k]$ to get $Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j} = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$
- MA(∞): set $\mu = 0$ and $\psi_j = \sum_{k=0}^{\infty} \theta_k \mathbb{I}[j = k]$ to get $Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j} = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j}$

A generic linear model representation can be stated for weakly stationary processes.

Every second-order stationary process is either a linear process or can be transformed to a linear process by subtracting a deterministic component (a time-invariant linear filter representation)

Definition

A time series $\{Y_t\}$ is a linear process if it has the representation

$$Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \quad \text{for all } t$$

where $\{\epsilon_t\} \sim WN(0, \sigma^2)$, μ is a real constant, and coefficient sequence $\{\psi_j\}$ is bounded by an absolutely decreasing sequence $|\psi_j| \leq h_j$ with $h_j \downarrow 0$ as $j \to \pm\infty$ and that satisfies absolute summability $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

Qn: Why do we need the condition on the coefficients $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$?

Ans: condition on coefficients $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ allows us to guarantee that Y_t is a sensible representation in the sense that $|Y_t| < \infty$ almost surely (probability that $|Y_t|$ is finite is 1)

We can see this as follows, since $\mathbb{E}[|\epsilon_t|] < \sigma$ for all t and an application of Markov's inequality shows:

$$\begin{aligned}
\mathbb{P}r\left(|Y_t| \geq a\right) &\leq \frac{1}{a}\mathbb{E}\left[|Y_t|\right] = \frac{1}{a}\mathbb{E}\left[\left|\mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}\right|\right] \\
&\leq \frac{1}{a}\left\{|\mu| + \sum_{j=-\infty}^{\infty} |\psi_j| \mathbb{E}\left[|\epsilon_{t-j}|\right]\right\} \\
&\leq \frac{1}{a}\left\{|\mu| + \sigma \sum_{j=-\infty}^{\infty} |\psi_j|\right\} \rightarrow 0, \text{ as } a \rightarrow \infty.
\end{aligned}$$

Note: last step uses fact that ϵ_t has zero mean and fact that mean absolute deviation will always be less-than-equal-to standard deviation.

One can also show that the condition on coefficients $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ also allows us to show that Y_t will converge in mean square sense, i.e. if $Y_t \stackrel{ms}{\rightarrow} Y$ and $X_t \stackrel{ms}{\rightarrow} X$, then as $t \rightarrow \infty$ one has:

$$\mathbb{E}[Y_t] \rightarrow \mathbb{E}[Y]$$
$$\mathbb{E}\left[|Y_t|^2\right] \rightarrow \mathbb{E}\left[|Y|^2\right]$$
$$\mathbb{E}[X_t Y_t] \rightarrow \mathbb{E}[XY]$$

First we need to recall the L^2 completeness theorem known as the Riesz-Fisher Theorem (Cauchy criterion) of convergence in a mean squared sense.

- This theorem is often easier to establish whether or not a mean square limit exists for time series $\{Y_t\}$ without needing to know what it is.

Sequences that satisfy this theorem are called Cauchy sequences in L^2 space (Hilbert space of square-integrable functions)

Theorem (Riesz-Fisher Theorem)

Let time series $\{Y_t\}$ *be a sequence in* L^2. *Then, there exists a* Y *in* L^2 *such that* $Y_t \overset{ms}{\to} Y$ *if and only if the following asymptotic limit holds*

$$\lim_{m,n\to\infty} \mathbb{E}\left[(Y_m - Y_n)^2\right] = 0$$

Note: one may see it also expressed as:

$$\lim_{m\to\infty} \sup_{n\geq m} \mathbb{E}\left[(Y_m - Y_n)^2\right] = 0$$

Now consider a linear process $Y_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$ and if one restricts the coefficients such that $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ then using Reisz-Fisher Theorem we can show that this linear process representation is logical in that Cauchy limits exist, i.e. subsequences of this linear representation, given by

$$Y_{t,m} = \sum_{j=-m}^{m} \psi_j \epsilon_{t-j}$$

converge in a mean square sense as $m \rightarrow \infty$.

proof We can apply this theorem as follows by considering the subsequences of the time series representation as follows

$$Y_{t,m} = \sum_{j=-m}^{m} \psi_j \epsilon_{t-j}$$

for $m = 1, 2, 3, \ldots$ and we want to show that this sequence of partial sums will admit a mean square limit as $m \rightarrow \infty$.

By Riesz-Fisher Theorem it suffices to show that $\mathbb{E}\left[(Y_{t,m} - Y_{t,n})^2\right] = 0$ as $m, n \to \infty$. For $n > m > 0$ one has

$$\begin{aligned}
\mathbb{E}\left[(Y_{t,n} - Y_{t,m})^2\right] &= \mathbb{E}\left[\left(\sum_{m<|j|<n} \psi_j \epsilon_{t-j}\right)^2\right] \\
&= \sum_{m<|j|<n} \sum_{m<|k|<n} \psi_j \psi_k \mathbb{E}\left[\epsilon_{t-j}\epsilon_{t-k}\right] \\
&= \sum_{m<|j|<n} \psi_j^2 \mathbb{E}\left[\epsilon_{t-j}^2\right] \\
&= \sum_{m<|j|<n} \psi_j^2 \left[\gamma_\epsilon(0) + \mu_\epsilon^2\right] \\
&\leq \gamma_\epsilon(0) \left(\sum_{m<|k|<n} |\psi_j|\right)^2 \to 0
\end{aligned}$$

As $m, n \to \infty$ since $\gamma_\epsilon(0) < \infty$ is constant and $\{\psi_j\}$ are assumed absolutely summable, and bounded by decreasing absolute sequence, then the tail sums $\to 0$!

Although this allows us to show that sequence of partial sums $\{Y_{t,m}\}$ converges in mean square sense, the limit has not been explicitly established.

If we wanted to show that the limit as $m \to \infty$ is the linear filter representation given by

$$Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \quad \text{for all } t$$

Then we can proceed by application of Fatou's Lemma as follows, where we assume the mean square limit of $Y_{t,m}$ is denoted generically by S and we want to show that $S = Y_t$

$$\mathbb{E}\left[|S - Y_t|^2\right] = \mathbb{E}\left[\lim \inf_{n\to\infty} (S - Y_{t,n})^2\right] \leq \lim \inf_{n\to\infty} \mathbb{E}\left[(S - Y_{t,n})^2\right] = 0$$

where the inequality is via Fatou's lemma and then we have the result from the mean square convergence to apply to get 0, which establishes that Y_t is the mean square limit of sequence $Y_{t,n}$ when it has this linear process form.

Lastly, we can also use the mean square convergence results to show that this linear process is stationary with

$$
\begin{aligned}
\mu_Y &= \mathbb{E}\left[Y_t\right] = \mu + \mu_\epsilon \sum_{j=-\infty}^{\infty} \psi_j = \mu \\
\gamma_Y(k) &= \mathbb{E}\left[\sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_i \left(\epsilon_{t+k-i} - \mu_\epsilon\right) \psi_j \left(\epsilon_{t-j} - \mu_\epsilon\right)\right] \\
&= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \gamma_\epsilon(k-j+1) \psi_j
\end{aligned}
$$

where we recall that $\mu_\epsilon = 0$ by definition and in the case of ϵ_t being a WN sequence, furthermore, one can simplify the ACVF as follows, based on WN

$$
\begin{aligned}
\gamma_Y(k) &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \gamma_\epsilon(k-j+1) \psi_j = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \psi_j \mathbb{I}\left[j = i - k\right] \sigma^2 \\
&= \sigma^2 \sum_{i=-\infty}^{\infty} \psi_i \psi_{i-k} = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_{j+k} \psi_j
\end{aligned}
$$

In the case of a general linear model of this form, one can obtain the following confidence intervals for estimation of the mean μ_Y based on $\gamma_Y(\cdot)$.

Definition

For the linear process $Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$, *with* $\{\epsilon_t\} \sim WN(0, \sigma^2)$, *coefficient sequence* $\{\psi_j\}$ *satisifies* $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ *and* $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$, *then*

$$\sqrt{n}\left(\overline{Y}_n - \mu_Y\right) \sim N(0, \nu), \quad \text{as } n \to \infty$$

with

$$\nu = \sum_{k=-\infty}^{\infty} \gamma_Y(k) = \sigma^2 \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_{j+k}\psi_j.$$

This result for the linear process also hold for ARMA models we will see next. This gives us a 95% approximated confidence interval for mean μ_Y as follows

$$\left(\overline{Y}_n - 1.96\sqrt{\nu/n}, \overline{Y}_n + 1.96\sqrt{\nu/n}\right).$$

Typically one would have to estimate $\nu = \sum_{k=-\infty}^{\infty} \gamma_Y(k)$ from the observed time series sample where an estimator can be obtained:

- a first intuitive approach would be to consider to use $\widehat{\nu} = \sum_{k=-\infty}^{\infty} \widehat{\gamma}_Y(k)$ that is constructed based on a finite sample of observed time series $\{y_1, \ldots, y_n\}$. However, it is impossible to obtain a reasonable estimator for $\gamma_Y(k)$ for $k \geq n$.

So why not use a truncated estimator such as

$$\widehat{\nu} = \sum_{k=-(n-1)}^{n-1} \widehat{\gamma}_Y(k),$$

however interestingly you will find that this estimator is always 0!

$$\begin{aligned}
\widehat{\nu} &= \sum_{k=-(n-1)}^{n-1} \widehat{\gamma}_Y(k) = \sum_{k=-(n-1)}^{n-1} \frac{1}{n} \sum_{t=1}^{n-|k|} \left(Y_{t+|k|} - \overline{Y}_n\right)\left(Y_t - \overline{Y}_n\right) \\
&= \frac{1}{n} \sum_{t=1}^{n} \left(Y_t - \overline{Y}_n\right)^2 + \frac{2}{n} \sum_{k=1}^{n-1} \sum_{t=1}^{n-k} \left(Y_{t+k} - \overline{Y}_n\right)\left(Y_t - \overline{Y}_n\right) \\
&= \frac{1}{n} \sum_{t=1}^{n} Y_t^2 - n\overline{Y}_n^2 + \frac{2}{n} \sum_{k=1}^{n-1} \sum_{t=1}^{n-k} Y_{t+k} Y_t = 0
\end{aligned}$$

A compromise is to use the estimator given by

$$\nu = \sum_{k=-[\sqrt{n}]}^{[\sqrt{n}]} \left(1 - \frac{|k|}{[\sqrt{n}]}\right) \widehat{\gamma}_Y(k)$$

Alternatively, if we know what type of time series model has generated the samples (or will assume such a model), i.e. we have an explicit formula for $\gamma_Y(k)$, then we can construct more accurate estimates of $\widehat{\nu}$.

Proposition

If $\{Y_t\}$ *is a stationary process*

$$Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$$

where $\epsilon_t \sim IID(0, \sigma^2)$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ *and* $\mathbb{E}\left[\epsilon_t^4\right] < \infty$, *then for each* k, *one has*

$$\widehat{\rho}_Y(k) = \begin{pmatrix} \widehat{\rho}_Y(1) \\ \widehat{\rho}_Y(2) \\ \vdots \\ \widehat{\rho}_Y(k) \end{pmatrix} \sim N\left(\rho_Y(k) = \begin{pmatrix} \rho_Y(1) \\ \rho_Y(2) \\ \vdots \\ \rho_Y(k) \end{pmatrix}, T^{-1}\Omega\right), \quad as\ T\infty$$

where $\Omega = [\omega_{ij}]_{i,j=1}^{k}$ *is the covariance matrix whose* (i,j)*-element is given by Bartlett's formula*

$$\omega_{ij} = \sum_{k=1}^{\infty} \widetilde{\rho}_Y(k,i)\widetilde{\rho}_Y(k,j)$$

with $\widetilde{\rho}_Y(k,\cdot) := \rho_Y(k+\cdot) + \rho_Y(k-\cdot) - 2\rho_Y(\cdot)\rho_Y(k)$

Linear Time Series Models Built from WN: Special Case of Moving Averages

Definition (Moving average $MA(q)$ process)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. Then, $\{Y_t\}$ is a moving average process of order q, written $MA(q)$, if

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

An MA model defines a process which, at time t, involves a random event at time t plus weighted random events from near past. e.g. economic indicators.

Moving Average Models: MA(q)

$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$ where $Z_t \sim$ WN $(0, \sigma_Z^2)$

For example,

MA(2): $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$

MA(3): $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \theta_3 Z_{t-3}$

What order is $X_t = Z_t - 0.2Z_{t-1} - 0.7Z_{t-2}$?

What order is $X_t = Z_t - 0.6Z_{t-1} + 0.08Z_{t-2} + 0.3\, Z_{t-5}$?

$\theta_0 = 1$. All missing terms correspond to a zero coefficient.

Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

Example ($MA(1)$ Model ACVF)

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1}$$

It is instructive to derive the autcovariance function (ACVF) $\gamma(\cdot)$ of an $MA(1)$ process. This can be done in numerous 'different' ways:

- *via Wold*
- *Expansion based approach with Y_{t-k} and taking $\mathbb{E}(\cdot)$*

ACVF of $MA(1)$ via Wold Result

Via Wold Write $MA(1)$ in similar form as Wold representation:

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} = \sum_{j=0} \beta_j \epsilon_{t-j} \quad \text{with } \beta_j := \begin{cases} 1, & j = 0 \\ \theta_1, & j = 1 \\ 0, & j \geq 2 \end{cases}$$

From Wold Decomposition recall that the ACVF $\gamma(\cdot)$ is given by:

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k}.$$

Note, for $MA(1)$ that $\beta_{j+k} = 0$ for $j + k \geq 2$. i.e. we can restrict to a finite sum as follows:

$$\gamma(k) = \sigma^2 \sum_{j=0}^{1-k} \beta_j \beta_{j+k}$$

$$
\begin{aligned}
\gamma(k) &= \sigma^2 \sum_{j=0}^{1-k} \beta_j \beta_{j+k} \\
&= \sigma^2 \left(\beta_0 \beta_k + \sum_{j=1}^{1-k} \beta_j \beta_{j+k} \right) \\
&= \sigma^2 \left(\beta_k + \sum_{j=1}^{1-k} \theta_j \theta_{j+k} \right) \\
&= \begin{cases} \sigma^2(1+\theta_1^2), & k=0 \\ \sigma^2\theta_1, & k=1 \\ 0, & k \geq 2 \end{cases} \\
[\gamma \text{ is symmetric}] &= \begin{cases} \sigma^2(1+\theta_1^2), & k=0 \\ \sigma^2\theta_1, & |k|=1 \\ 0, & |k| \geq 2 \end{cases}
\end{aligned}
$$

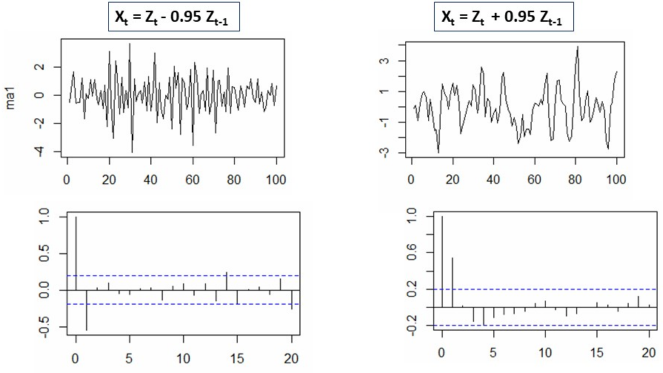
Alternative derivation approach when $\{Y_t\}$ is $MA(1)$, i.e.

$$\begin{aligned} Y_t &= \epsilon_t + \theta_1 \epsilon_{t-1} \\ Y_{t-k} &= \epsilon_{t-k} + \theta_1 \epsilon_{t-k-1} \end{aligned}$$

Hence

$$\begin{aligned} \gamma_Y(k) = \operatorname{cov}(Y_t, Y_{t-k}) &= \operatorname{cov}(\epsilon_t + \theta_1\epsilon_{t-1}, \epsilon_{t-k} + \theta_1 \epsilon_{t-k-1}) \\ &= \operatorname{cov}(\epsilon_t, \epsilon_{t-k}) + \theta_1 \operatorname{cov}(\epsilon_t, \epsilon_{t-k-1}) \\ &\quad + \theta_1 \operatorname{cov}(\epsilon_{t-1}, \epsilon_{t-k}) + \theta_1^2 \operatorname{cov}(\epsilon_{t-1}, \epsilon_{t-k-1}) \\ [\text{note } \operatorname{cov}(\epsilon_s, \epsilon_t) = \sigma^2 \delta_{s,t}] &= \sigma^2(\delta_{0,k} + \theta_1 \delta_{-1,k} + \theta_1 \delta_{1,k} + \theta_1^2 \delta_{0,k}) \\ &= \sigma^2(1+\theta_1^2)\delta_{0,k} + \sigma^2\theta_1(\delta_{-1,k} + \delta_{1,k}) \\ &= \sigma^2(1+\theta_1^2)\delta_{0,k} + \sigma^2\theta_1\delta_{1,|k|} \\ &= \begin{cases} \sigma^2(1+\theta_1^2), & k = 0 \\ \sigma^2\theta_1, & |k| = 1 \\ 0, & \text{oth. } (|k| \geq 2) \end{cases} \end{aligned}$$

100 simulated values of MA(1) and its ACF

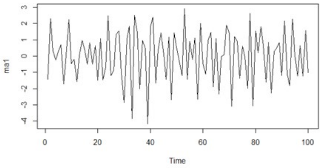


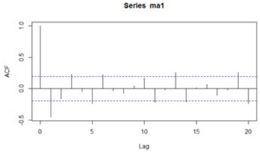
$X_t = Z_t - 0.95\, Z_{t-1}$
Correlation is 0 after lag 1.
Graph rough -- strong negative correlation

$X_t = Z_t + 0.95\, Z_{t-1}$
Correlation is 0 after lag 1.
Graph smooth -- strong positive correlation

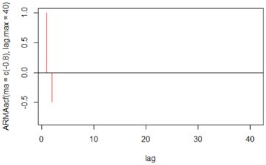
Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

100 simulated values of MA(1) $X_t = Z_t - 0.8\, Z_{t-1}$





ACF=0, k > 1. Negative correlation at lag 1. Series is choppy.



R Commands:

```
> ma1 <- arima.sim(model=list(ma=c(-0.8)), n=100)
> plot(ma1)
> acf(ma1)
> plot(ARMAacf(ma=c(-0.8), lag.max=40), col="red", type="h", xlab="lag", ylim=c(-.8,1));
abline(h=0)
```

Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

Review of $MA(1)$ Properties

ACVF and ACF of Moving Average Models: MA(1)

<u>REVIEW:</u> MA(1) Moving Average of Order One

$X_t = Z_t + \theta_1 Z_{t-1}$ **where** $Z_t \sim$ **WN** $(0, \sigma_Z^2)$ **and** $|\theta_1| < 1$.

X is stationary with $\mu_x = E(X_t) = 0$**, and ACF**

$\rho_X(0) = 1$; $\rho_X(k) = 0$ for $|k| > 1$, and $\rho_X(1) = \theta_1/(1+\theta_1^2)$

Note: For MA(1),

Autocorrelations for lags $k \neq 0, 1$ are 0: $\rho_X(k) = 0$ for $|k| > 1$.

Sign $(\rho_X(1))$ = Sign (θ_1)

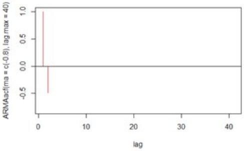


Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

Expanding to Moving Average Order 2.

Example ($MA(2)$)

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

Find ACVF γ via Wold

Via Wold Write $MA(2)$ in similar form as Wold representation:

$$Y_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j} \quad \text{with } \beta_j := \begin{cases} 1, & j = 0 \\ \theta_j, & j = 1, 2 \\ 0, & j \geq 3 \end{cases}$$

From Wold, $\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k}$. Note, for $MA(2)$ that $\beta_{j+k} = 0$ for $j + k \geq 3$. I.e.

$$\gamma(k) = \sigma^2 \sum_{j=0}^{2-k} \beta_j \beta_{j+k}$$

ACVF of $MA(2)$ via Wold Result

$$
\begin{aligned}
\gamma(k) &= \sigma^2 \sum_{j=0}^{2-k} \beta_j \beta_{j+k} \\
&= \sigma^2 \left(\beta_0 \beta_k + \sum_{j=1}^{2-k} \beta_j \beta_{j+k} \right) \\
&= \sigma^2 \left(\beta_k + \sum_{j=1}^{2-k} \theta_j \theta_{j+k} \right) \\
&= \begin{cases} \sigma^2 (1 + \theta_1^2 + \theta_2^2), & k = 0 \\ \sigma^2 (\theta_1 + \theta_1 \theta_2), & k = 1 \\ \sigma^2 \theta_2, & k = 2 \\ 0, & k \geq 3 \end{cases}
\end{aligned}
$$

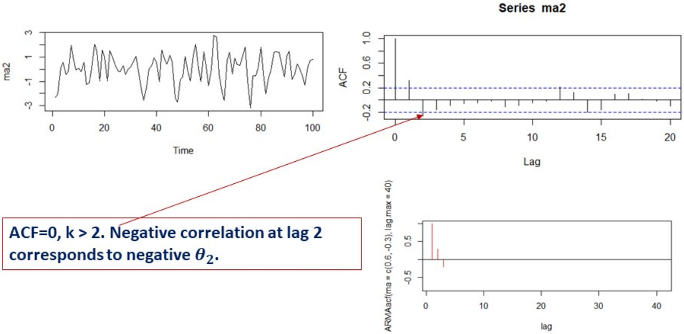
ACF of $MA(2)$

$$\gamma(k) = \begin{cases} \sigma^2(1+\theta_1^2+\theta_2^2), & k=0 \\ \sigma^2(\theta_1+\theta_1\theta_2), & |k|=1 \\ \sigma^2\theta_2, & |k|=2 \\ 0, & |k|\geq 3 \end{cases}$$

$$\therefore \rho(k) = \frac{\gamma(k)}{\gamma(0)} = \begin{cases} 1, & |k|=0 \\ \dfrac{\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}, & |k|=1 \\ \dfrac{\theta_2}{1+\theta_1^2+\theta_2^2}, & |k|=2 \\ 0, & |k|\geq 3 \end{cases}$$

Understanding ACF for MA(2)

100 simulated values of MA(2) $X_t = Z_t + 0.6\, Z_{t-1} - 0.3\, Z_{t-2}$



$$\rho_X(1) = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_X(2) = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_X(k) = 0, \; k > 2.$$

Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

Generalised to Moving Average order $q \in \mathbb{N}$

Example ($MA(q)$)

An $MA(q)$ process can be written $Y_t = \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$,

c.f. Wold: $$Y_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j}\,, \qquad \text{with } \beta_j = \begin{cases} 1, & j = 0 \\ \theta_j & j = 1, 2, \ldots, q \\ 0, & \text{oth.} \end{cases}$$

From Wold:

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k}$$

Now, $\beta_j = 0\,, \forall j \geq q+1$, i.e. $\beta_{j+k} = 0\,, \forall j \geq q-k+1$

$$\gamma(k) \quad = \quad \sigma^2 \sum_{j=0}^{q-k} \beta_j \beta_{j+k} = \sigma^2 \left(\beta_k + \sum_{j=1}^{q-k} \theta_j \theta_{j+k} \right)$$

ACF of $MA(q)$

$$\gamma(k) = \sigma^2 \left(\beta_k + \sum_{j=1}^{q-k} \theta_j \theta_{j+k} \right)$$

In particular: $\gamma(0) = \sigma^2 \left(1 + \sum_{j=1}^{q} \theta_j^2\right)$. i.e.

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \begin{cases} 1, & |k| = 0 \\ \dfrac{\theta_k + \sum_{j=1}^{q-k} \theta_j \theta_{j+k}}{1 + \sum_{j=1}^{q} \theta_j^2}, & |k| = 1, \ldots, q \\ 0, & |k| \geq q+1 \end{cases}$$

Remark (ACF 'cut-off')

Note that the ACF of an $MA(q)$ process 'cuts off' at $|k| = q + 1$ (is zero for $|k| > q$). Hence, ACF can be used as a model identification tool!

Moving Average Models: MA(q)

$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$ where $Z_t \sim$ WN $(0, \sigma_Z^2)$

Moment Calculation: $E(X_t) = 0$ (linearity; $E(Z_t) = 0$ by def'n of WN)

Denote $\theta_0 = 1$. Then,

$$X_t = 1 \cdot Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} = \sum_{i=0}^{q} \theta_i Z_{t-i}$$

$$\gamma_X(k) = E(X_t X_{t+k}) = E\{(\sum_{i=0}^{q} \theta_i Z_{t-i})(\sum_{j=0}^{q} \theta_j Z_{t+k-j})\} \qquad \text{(k-j = -m)}$$

$$= E\{(\sum_{i=0}^{q} \theta_i Z_{t-i})(\sum_{m=-k}^{q-k} \theta_{k+m} Z_{t-m})\} \qquad \text{(t-i=t-m)}$$

$$= \sum_{i=0}^{q-k} \theta_i \, \theta_{k+i} \, E(Z_{t-i} Z_{t-i})$$

$$= \sigma_Z^2 \sum_{i=0}^{q-k} \theta_i \, \theta_{k+i}$$

$$= \sigma_Z^2 (\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \ldots + \theta_{q-k} \theta_q)$$

Note: $\gamma_X(0) \equiv Var(X_t) = \sigma_Z^2 (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)$.

Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

Summary on MA(q) Processes

$$Y_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \text{ with } Z_t \sim WN\left(0, \sigma_Z^2\right)$$

Conclude: $MA(q)$ is always stationary with

- Mean given by $\mu_Y = 0$ and $Var(Y) = \sigma_Z^2 \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)$ which are constants.
- Autocovariance for lags $k = 1, 2, \ldots, q$ given by

$$\gamma_Y(k) = \sigma_Z^2 \left(\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q\right)$$

  Autocovariance for lags $k > q$ are 0: $\gamma_Y(k) = 0$ for $|k| > q$.
- Autocorrelation is given by

$$\rho_Y(k) = \frac{\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2}, \quad k = 1, \ldots, q$$

  and $\rho_Y(k) = 0$ for $k > q$.

Reminder: Operators on Time Series and Invertible Time Series Representations

Definition (Backshift operator)

Let $\{Y_t\}$ be some time series. Then the backshift operator (a.k.a lag operator) B is defined by

$$B^k Y_t = Y_{t-k}$$

Remark

Some texts your will see the Lag operator

$$L^k Y_t = Y_{t-k}$$

Note, the backshift or lag operator can be raised to arbitrary integer powers.

$$B^{-1} Y_t = Y_{t+1}$$

Proposition

Let B be the backshift operator. Let $a, c \in \mathbb{R}$, and $\{X_t\}$ be some sequence. Then, $\forall i, j \in \mathbb{N} = \{0, 1, 2, \ldots\}$:

1. $Bc = c$
2. $B^j(cX_t) = cB^jY_t = cY_{t-j}$
3. $(aB^i + cB^j)Y_t = aB^iY_t + cB^jY_t = aY_{t-i} + cY_{t-j}$
4. $B^iB^jY_t = B^iY_{t-j} = Y_{t-j-i} = B^{i+j}Y_t$
5. $(1 - aB)^{-1}Y_t = \sum_{j=0}^{\infty} a^jB^jY_t = \sum_{j=0}^{\infty} a^jY_{t-j}\,, \quad$ *if* $|a| < 1$

Proof 1 – 4: by definition(!). For 5 want: $(1-aB)^{-1}Y_t = \sum_{j=0}^{\infty} a^j B^j Y_t$.
'Proof' Consider:

$$\begin{aligned}
(1-aB)\left(\sum_{j=0}^{\infty} a^j B^j\right) Y_t &= \left(\sum_{j=0}^{\infty} a^j B^j - \sum_{j=0}^{\infty} a^{j+1} B^{j+1}\right) Y_t \\
&= \left(\sum_{j=0}^{\infty} a^j B^j - \sum_{j=1}^{\infty} a^j B^j\right) Y_t \\
&= \left(1 + \cancel{\sum_{j=1}^{\infty} a^j B^j} - \cancel{\sum_{j=1}^{\infty} a^j B^j}\right) Y_t \\
&= Y_t
\end{aligned}$$

$\therefore \quad \sum_{j=0}^{\infty} a^j B^j = (1-aB)^{-1}$, provided sum converges, i.e. $|a| < 1$ ■

Remark

It can be shown that other expressions involving the backshift operator can often be manipulated as if B were a number or variable...

Next we will introduce the operator that is analogous to differentiation in a discret time TS setting.

- Define the lag-1 difference operator (like a first derivative)

$$\nabla Y_t = \frac{Y_t - Y_{t-1}}{t-(t-1)} = Y_t - Y_{t-1} = (1-B)Y_t$$

Definition (Difference operator)

Let $\{Y_t\}$ be some time series. Then the difference operator (a.k.a finite difference operator) ∇^k is defined by

$$\nabla^k Y_t = (1-B)^k Y_t$$

Note: sometimes you see it written as $\nabla_k Y_t$

Example

$$
\begin{aligned}
\nabla Y_t &= (1-B)Y_t = Y_t - Y_{t-1} \\
\nabla^2 Y_t &= \nabla(\nabla Y_t) = (1-B)\nabla Y_t = \nabla Y_t - \nabla Y_{t-1} \\
&\vdots \\
\nabla^k Y_t &= (1-B)^k Y_t = \sum_{j=0}^{k} (-1)^j \binom{k}{j} B^j Y_t = \sum_{j=0}^{k} (-1)^j \binom{k}{j} Y_{t-j}
\end{aligned}
$$

Some examples:

- Note that if $Y_t = b_0 + b_1 t + \epsilon_t$ then:

$$\nabla Y_t = b_1 + \nabla \epsilon_t$$

- Note that if $Y_t = \sum_{i=0}^{k} b_i + \epsilon_t$ then:

$$\nabla^k Y_t = k! b_k + \nabla^k \epsilon_t$$

where $\nabla^k Y_t = \nabla(\nabla^{k-1} Y_t)$ and $\nabla^1 Y_t = \nabla Y_t$

- Note that if $Y_t = m_t + s_t + \epsilon_t$ where s_t has period k (i.e. $s_t = s_{t-k}$ for all t), then:

$$\nabla^k Y_t = m_t - m_{t-k} + \nabla^k \epsilon_t$$

Definition (Invertible Processes)

A time series $\{Y_t\}$ is invertible if the shocks (WN) $\{Z_t\}$ can be expressed via past values of Y_t as a convergent series given by

$$Z_t = Y_t + \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \pi_3 Y_{t-3} + \cdots = \sum_{j=0}^{\infty} \pi_j(B) Y_t = \Pi(B) Y_t$$

where $\Pi(B)Y_t$ is an $AR(\infty)$ representation.

This is important as can then deduce from an invertible TS that

$$Y_t = Z_t - \pi_1 Y_{t-1} - \pi_2 Y_{t-2} - \pi_3 Y_{t-3} - \cdots$$

Since the series is convergent for $|\pi_n| \to 0$ as $n \to \infty$, so that the contribution of the most recent terms are most important (we forget past fast enough).

Remark

Invertibility roughly says that remote past has less influence on current values.

Using the shift operator B, one can express $MA(1)$ TS as follows for WN process $\{Z_t\}$

$$Y_t = Z_t + \theta_1 Z_{t-1} = Z_t + \theta_1 B Z_t = (1 + \theta_1 B) Z_t$$

Now we can invert the process as follows

$$\begin{aligned} Z_t &= \frac{1}{(1 + \theta_1 B)} Y_t \\ &= (1 + (-\theta_1)B + (-\theta_1)^2 B^2 + \cdots + (-\theta_1)^k B^k + \cdots) Y_t \\ &= Y_1 - \theta_1 Y_{t-1} - \cdots + (-\theta_1)^k Y_{t-k} - \cdots \end{aligned}$$

where this geometric series converges only for $|\theta_1| < 1$.

We can work with the backshift operator to factorise for the MA(q) model as follows

$$\begin{aligned} Y_t &= Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + +\theta_q Z_{t-q} \\ &= Z_t + \theta_1 B Z_t + \theta_2 B^2 Z_t + \cdots + +\theta_q B^q Z_t \end{aligned}$$

Lets introduce then the Characteristic polynomial $\theta(\cdot)$ of order q given by

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

then we can write

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

Hence we can write $Y_t = \theta(B)Z_t$, and we can say that Y_t is invertible MA(q) process if polynomial $\theta(z) \neq 0$ for $|z| \leq 1$, that is, the roots of the polynomial $\theta(z)$ lie outside of the unit circle.

In this case we can write $\theta^{-1}(z) = 1 + \pi_1 z + \pi_2 z^2 + \ldots$ which has a representation as a convergent series so that for our MA(q) process one gets a representation for the current shock given in terms of past values of MA(q) process as follows

$$Z_t = \theta^{-1}(B)X_t = X_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots$$

In other words:

Definition

An MA model is said to be invertible if it can be written as an AR model (AR models to be explored next sections).

- Are MA models unique?

We will consider how invertibility restriction will resolve this question.

Example

Consider the two $MA(1)$ models

$$\begin{aligned} \textit{model I:} \quad Y_t &= \epsilon_t + \theta_1 \epsilon_{t-1} \\ \textit{model II:} \quad Y_t &= \epsilon_t + \theta_1^{-1} \epsilon_{t-1} \end{aligned}$$

Recall Model I has ACF with $\rho_{\mathrm{I}}(1) = \theta_1(1+\theta_1^2)^{-1}$ and $\rho_{\mathrm{I}}(k) = 0\,, \forall |k| \geq 2$. Model II has

$$\rho_{\mathrm{II}}(1) = \frac{\theta_1^{-1}}{1+\theta_1^{-2}} = \frac{\theta_1}{1+\theta_1^2} = \rho_{\mathrm{I}}(1)\,.$$

and $\rho_{\mathrm{II}}(k) = 0\,, \forall\,|k| \geq 2$. Hence $\rho_{\mathrm{I}} \equiv \rho_{\mathrm{II}}$. This is **not** a nice situation!

This uniqueness problem can be solved by asking the question:

Under what conditions can we represent an MA model as an AR model?

Theorem (roots of $MA(q)$ char. eqn. outside unit circle $\Leftrightarrow$ invertible)

Consider $MA(q)$ process: $Y_t = \epsilon_t + \sum_{j=1}^{p} \theta_j \epsilon_{t-j}$. Then $\{Y_t\}$ is invertible iff all roots of characteristic equation:

$$\theta(x) := 1 + \theta_1 x + \theta_2 x^2 + \ldots + \theta_q x^q \,,$$

are outside unit circle.

Example

$MA(1)$: $Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} =: \theta(B)\epsilon_t$

$Y_t = (1 + \theta_1 B)\epsilon_t$. I.e if the $MA(1)$ root $|\theta_1^{-1}| > 1$ then

$$\epsilon_t = (1 + \theta_1 B)^{-1} Y_t = \sum_{j=0}^{\infty} (-\theta_1 B)^j Y_t = \sum_{j=0}^{\infty} (-\theta_1)^j Y_{t-j}$$

Recall the two $MA(1)$ models:

$$\begin{array}{llll} \text{model I:} & Y_t = \epsilon_t + \theta_1 \epsilon_{t-1}\,, & \text{root} & = \theta_1^{-1} \\ \text{model II:} & Y_t = \epsilon_t + \theta_1^{-1} \epsilon_{t-1}\,, & \text{root} & = \theta_1 \end{array}$$

Then, for $|\theta_1^{-1}| > 1$, model I is invertible and model II is not. The invertibility condition ensures ACF of MA models is unique.

Invertibility check for MA(1) & MA(2)

Invertibility of MA(q)

MA(q) $X_t = \theta(B)\, Z_t$ **with** $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$ **is Invertible if** $\theta(z) \neq 0$ **for** $|z| \leq 1$, **that is, the roots of the polynomial** $\theta(z)$ **lie outside of the unit circle.**

Check for MA(1):

$$X_t = Z_t + \theta_1 Z_{t-1} \equiv (1 + \theta_1 B) Z_t \equiv \theta(B) Z_t \text{ with } \theta(z) = 1 + \theta_1 z.$$

$$\theta^{-1}(z) = \frac{1}{1+\theta_1 z} = \frac{1}{1-(-\theta_1 z)} = 1 + (-\theta_1 z) + (-\theta_1 z)^2 + \cdots$$

if $1 + \theta_1 z \neq 0$. Series convergent if $|\theta_1 z| < 1$.

To substitute z = B, we need $|\theta_1 B| < 1$ or $|\theta_1| < 1$. Then,

$$Z_t = \theta^{-1}(B) X_t = \frac{1}{1+\theta_1 B} X_t = (1 - \theta_1 B + \theta_1^2 B^2 - \cdots) X_t$$

$$= X_t - \theta_1 X_{t-1} + \theta_1^2 X_{t-2} - \cdots \text{ (convergent series)}$$

Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

- What are the coefficient restrictions on MA(2) model for invertibility?

Example

Consider the two $MA(2)$ *models*

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

with $\epsilon_t \sim WN(0, \sigma^2)$

Write

$$\Theta(z) := 1 + \theta_1 z + \theta_2 z^2 = (1 - \alpha_1 z)(1 - \alpha_2 z)$$

and check that $|\alpha_i| < 1$ since then roots of $\Theta(z)$ are given by $\frac{1}{\alpha_i}$ and satisfy that $|\frac{1}{\alpha_i}| > 1$ outside unit disc.

$$\theta_1 = -(\alpha_1 + \alpha_2) \;\; \& \;\; \theta_2 = \alpha_1 \alpha_2$$

which after some algebra gives constraint regions

$$\theta_1 + \theta_2 > -1, \;\; \theta_2 - \theta_1 > -1, \;\; \theta_2 \in [-1, 1]$$

or

$$|\theta_1| < 1 \;\; \& \;\; |\theta_1| + \theta_2 < 1$$

Invertibility check for MA(1) & MA(2)

Results of MA(2) Example

$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$ **with $Z_t \sim$ WN $(0, \sigma_Z^2)$**

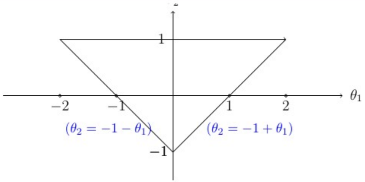
Corresponding polynomial $\theta(z) = 1 + \theta_1 z + \theta_2 z^2$;

Stationary with mean zero and ACF:

$\rho_X(1) = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$, $\rho_X(2) = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}$, $\rho_X(k) = 0$, $k > 2$.

Invertible if

$\theta_2 + \theta_1 > -1$ $\theta_2 - \theta_1 > -1$, $-1 < \theta_2 < 1$



$\theta_2 + \theta_1 > -1$
$\theta_2 - \theta_1 > -1$
$-1 < \theta_2 < 1.$

Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

General Solution for Invertibility

The general method to validate if coefficients are suitable for invertible solution is to solve the system of questions as follows, based on invertibility condition

$$Y_t = \Theta(B)\epsilon_t \;\; \& \;\; \epsilon_t = \Pi(B)Y_t$$

which gives

$$\Rightarrow \quad 1 = \Pi(B)\Theta(B)$$
$$\Rightarrow \quad 1 = (\pi_0 + \pi_1 B + \pi_2 B^2 + \cdots)(1 + \theta_1 B + \cdots + \theta_q B^q)$$

Matching coefficients gives system to solve

$$1 = \pi_0,$$
$$0 = \pi_1 + \theta_1 \pi_0,$$
$$0 = \pi_2 + \theta_1 \pi_1 + \theta_2 \pi_0,$$
$$\vdots$$

Linear Time Series Models Built from WN: Special Case of Autoregressive

Definition (Autoregressive process)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. *Then,* $\{Y_t\}$ *is an autoregressive process of order p, written* $AR(p)$*, if*

$$
\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t \\
&= \phi_1 B Y_t + \phi_2 B^2 Y_t + \ldots + \phi_p B^p Y_t + \epsilon_t \\
&= \left(\phi_1 B + \phi_2 B^2 + \ldots + \phi_p B^p\right) Y_t + \epsilon_t
\end{aligned}
$$

which can be rewritten as follows:

$$Y_t \left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right) = Y_t \phi(B) = \epsilon_t$$

with Characteristic AR Polynomial

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p.$$

An $AR(p)$ model defines a process which, at time t, depends linearly on past values: $Y_{t-1}, Y_{t-2}, \ldots$ (together with a random term / White noise term).

$AR(1)$ Model: Invertible & Stationary

Lets consider the case of an AR(1) model given by

$$Y_t = \phi_1 Y_{t-1} + Z_t, \quad \text{where } Z_t \sim WN(0, \sigma_Z^2) \;\; \& \;\; |\phi_1| < 1.$$

In AR(1) model, current observation Y_t depends on todays shock Z_t and yesterdays observation Y_{t-1}

AR(1) processes are created from past observations and WN.

Note: Unlike in MA models, AR(1) is always invertible:

$$Z_t = Y_1 - \phi_1 Y_{t-1}$$

However - AR models are not always stationary!
Condition for stationarity is $|\phi_1| \leq 1$
(proof to come - we will see numerous ways to show this....)

$AR(1)$ Model ACVF

Lets calculate some basic summary of the AR(1) model moments:

- First moment: $\mathbb{E}[Y_t] = 0$ why? because MA has zero mean and can express as infinite sum of WN terms in MA form.
- ACVF can be calculated numerous ways, the approach we explore here is via a recursion:

$$\gamma_Y(k) = \mathbb{E}[Y_t Y_{t-k}] = \mathbb{E}[(\phi_1 Y_{t-1} + \epsilon_t) Y_{t-k}]$$

Firstly note: for $k \geq 1$ if you consider terms $\mathbb{E}[\epsilon_t Y_{t-k}]$ for each $k \geq 1$ it will be such that $\mathbb{E}[\epsilon_t Y_{t-k}] = 0$, since Y_{t-k} depends on $\epsilon_{t-k}, \epsilon_{t-k-1}$ etc. which are uncorrelated with ϵ_t according to WN assumption.

Therefore one has:

$$\begin{aligned}
\mathbb{E}\left[Y_t Y_{t-k}\right] &= \mathbb{E}\left[\left(\phi_1 Y_{t-1} + \epsilon_t\right) Y_{t-k}\right] \\
&= \phi_1 \mathbb{E}\left[Y_{t-1} Y_{t-k}\right] \\
&= \phi_1 \gamma_Y(k-1) \\
\Rightarrow \gamma_Y(k) &= \phi_1 \gamma_Y(k-1)
\end{aligned}$$

Hence, one would get the following expression (in terms of $\gamma_Y(0)$):

$$\begin{aligned}
\therefore \gamma_Y(1) &= \phi_1 \gamma_Y(0) \\
\gamma_Y(2) &= \phi_1 \gamma_Y(1) = \phi_1^2 \gamma_Y(0), \\
&\vdots \\
\gamma_Y(k) &= \phi_1^k \gamma_Y(0)
\end{aligned}$$

Which would produce an ACF given by $\rho_Y(k) = \frac{\gamma_Y(k)}{\gamma_Y(0)}$ giving

$$\rho(1) = \phi_1, \;\; \rho_Y(2) = \phi_1^2, \;\; \cdots, \;\; \rho_Y(k) = \phi_1^k$$

Example $AR(1)$ ACVF at $k = 0$

What about the term $\gamma_Y(0)$ to complete the picture for the ACVF?

We will use property of geometric series $\sum_{i=0}^{\infty} ar^k = \frac{a}{1-r}$ for $|r| < 1$.

We will explore two different ways to now find $\gamma(0)$, the first is via iterative substitution and an L2 / mean squared error convergence argument.

Example ($\gamma(0)$ of $AR(1)$ using successive substitution)

$$
\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \epsilon_t && (0.1)\\
Y_{t-1} &= \phi_1 Y_{t-2} + \epsilon_{t-1} && (0.2)\\
Y_{t-2} &= \phi_1 Y_{t-3} + \epsilon_{t-2} && (0.3)
\end{aligned}
$$

Now substitute (0.2) into (0.1):

$$
\begin{aligned}
Y_t &= \phi_1(\phi_1 Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= \phi_1^2 Y_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t
\end{aligned}
$$

$$Y_t = \phi_1^2 Y_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \tag{0.4}$$

Now substitute (0.3) into (0.4):

$$\begin{aligned} Y_t &= \phi_1^2(\phi_1 Y_{t-3} + \epsilon_{t-2}) + \phi_1 \epsilon_{t-1} + \epsilon_t \\ &= \phi_1^3 Y_{t-3} + \phi_1^2 \epsilon_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \end{aligned}$$

By similar successive substitutions:

$$\begin{aligned} Y_t &= \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots + \phi_n \epsilon_{t-n} + \phi_1^{n+1} Y_{t-n-1} \\ &= \phi_1^{n+1} Y_{t-n-1} + \sum_{j=1}^{n} \phi_1^j \epsilon_{t-j} \, , \end{aligned}$$

Assuming $|\phi| < 1$, it is tempting to let $n \to \infty$ to get $Y_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}$. Further assuming that $\mathbb{E} Y_t^2 < \infty$ we can indeed say that this holds (albeit in the mean square sense).

Example $AR(1)$ ACVF at $k=0$

$$Y_t = \phi_1^{n+1} Y_{t-n-1} + \sum_{j=1}^{n} \phi_1^j \epsilon_{t-j} = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \cdots + \cancelto{0}{\lim_{j\to\infty} \phi_1^j Y_{t-j}}$$

if sum exists...

We can now show that Y_t converges in mean square to $\sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}$, sometimes written as $Y_t \xrightarrow{L_2} \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}$. I.e. want:

$$\lim_{n\to\infty} \mathbb{E}\left(\left| Y_t - \sum_{j=1}^{n} \phi_1^j \epsilon_{t-j} \right|^2 \right) = 0$$

We have:

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}\left(\left| Y_t - \sum_{j=1}^{n} \phi_1^j \epsilon_{t-j} \right|^2 \right) &= \lim_{n\to\infty} \mathbb{E}\left(\left| \phi_1^{n+1} Y_{t-n-1} \right|^2 \right) \\
&= \lim_{n\to\infty} \phi_1^{2n+2} \, \mathbb{E}\left(Y_{t-n-1}^2 \right) \\
&= 0
\end{aligned}$$

if $|\phi_1| < 1$ and $\mathbb{E}(Y_t^2) < \infty$.

Example $AR(1)$ ACVF at $k = 0$

Hence, we would have as $n \to \infty$:

$$
\begin{aligned}
\mathbb{E}(Y_t^2) &= \mathbb{E}\left(\sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}\right)^2 \\
&= \mathbb{E}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \phi_1^j \phi_1^k \epsilon_{t-j}\epsilon_{t-k} \\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \phi_1^j \phi_1^k \mathbb{E}\left(\epsilon_{t-j}\epsilon_{t-k}\right) \\
&= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} \phi_1^j \phi_1^k \sigma^2 \delta_{j,k} = \sigma^2 \sum_{j=0}^{\infty} \phi_1^{2j} = \frac{\sigma^2}{1-\phi_1^2}
\end{aligned}
$$

i.e. sum converges (in the mean square, w.p. 1), if $|\phi_1| < 1$ (c.f. geometric progression). So, provided that $|\phi_1| < 1$, then $\{Y_t\}$ can be written in form $\sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}$.

Definition (Causal Process)

A linear process Y_t *is* **causal** *(strictly, a causal function of* $\{\epsilon_t\}$*), if there is a characteristic polynomial*

$$\Psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$$

with $\sum_{j=1}^{\infty} |\psi_j| < \infty$ *s.t.*

$$Y_t = \Psi(B)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

Lets now use this result to obtain a second derivation of ACF for AR(1) process.

Then we will discuss causality in further detail follow this.

Alternatively one can also get to this same result for $\gamma(0)$ as follows (useful to see both approaches):

Consider for $k = 0$,

$$\gamma_Y(0) = \mathbb{E}\left[(\phi_1 Y_{t-1} + \epsilon_t) Y_t\right] = \phi_1 \mathbb{E}\left[Y_{t-1} Y_t\right] + \mathbb{E}\left[\epsilon_t Y_t\right]$$

assume Y_t admits a causal representation with respect to a WN process ($Y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$), with $|\phi_1| < 1$, then first consider mixed moments given by expression:

$$\mathbb{E}\left[\epsilon_t Y_t\right] = \mathbb{E}\left[\epsilon_t \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}\right]$$

where we will need to consider swapping order of integration and infinite summation - this is a direct application of **Fubini's Theorem**.

Recall: Fubini's theorem is a powerful tool that provides conditions for interchanging the order of integration in a double integral. Given that sums are essentially special cases of integrals, it also gives conditions for interchanging the order of summations, or the order of a summation and an integration.

- One may switch the order of integration if the double integral yields a finite answer when the integrand is replaced by its absolute value.

Example $AR(1)$ ACVF at $k = 0$

Fubini's theorem states that for a general function f_n,

$$\text{if } \int \sum_n |f_n(x)| dx < \infty \text{ or } \sum_n \int |f_n(x)| dx < \infty$$

(by Tonelli's Theorem these two conditions are equivalent) then one has

$$\int \sum_n f_n(x) dx = \sum_n \int f_n(x) dx.$$

So we are interested in the following expression:

$$\begin{aligned}
\mathbb{E}\left[\epsilon_t Y_t\right] &= \mathbb{E}\left[\epsilon_t \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}\right] \\
&= \int_{-\infty}^{\infty} \sum_{j=0}^{\infty} \phi_1^j \epsilon_t \epsilon_{t-j} f_{\epsilon_t, \epsilon_{t-j}}(\epsilon_t, \epsilon_{t-j}) d\epsilon_t d\epsilon_{t-j} \\
&= \int_{-\infty}^{\infty} \sum_{j=0}^{\infty} \phi_1^j \epsilon_t \epsilon_{t-j} f_{\epsilon_t}(\epsilon_t) f_{\epsilon_{t-j}}(\epsilon_{t-j}) d\epsilon_t d\epsilon_{t-j}
\end{aligned}$$

Example $AR(1)$ ACVF at $k = 0$

We wish to swap order of summation and integration

$$\int_{-\infty}^{\infty} \sum_{j=0}^{\infty} \phi_1^j \epsilon_t \epsilon_{t-j} f_{\epsilon_t}(\epsilon_t) f_{\epsilon_{t-j}}(\epsilon_{t-j}) d\epsilon_t d\epsilon_{t-j}$$

$$= \sum_{j=0}^{\infty} \phi_1^j \int_{-\infty}^{\infty} \epsilon_t \epsilon_{t-j} f_{\epsilon_t}(\epsilon_t) f_{\epsilon_{t-j}}(\epsilon_{t-j}) d\epsilon_t d\epsilon_{t-j} = \sum_{j=0}^{\infty} \phi_1^j \mathbb{E}\left[\epsilon_t \epsilon_{t-j}\right],$$

via Fubini's theorem.
To show this is possible, we can consider showing the following holds,

$$\sum_{j=0}^{\infty} \int_{-\infty}^{\infty} \left| \phi_1^j \epsilon_t \epsilon_{t-j} f_{\epsilon_t}(\epsilon_t) f_{\epsilon_{t-j}}(\epsilon_{t-j}) \right| d\epsilon_t d\epsilon_{t-j} < \infty,$$

then apply Tonelli and Fubini Theorems as follows:

$$\sum_{j=0}^{\infty} \int_{-\infty}^{\infty} \left| \phi_1^j \epsilon_t \epsilon_{t-j} f_{\epsilon_t}(\epsilon_t) f_{\epsilon_{t-j}}(\epsilon_{t-j}) \right| d\epsilon_t d\epsilon_{t-j}$$

$$= \sum_{j=0}^{\infty} |\phi_1^j| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\epsilon_t||\epsilon_{t-j}| f_{\epsilon_t}(\epsilon_t) f_{\epsilon_{t-j}}(\epsilon_{t-j}) d\epsilon_t d\epsilon_{t-j}$$

$$= \sum_{j=0}^{\infty} |\phi_1|^j \int_{-\infty}^{\infty} |\epsilon_t| f_{\epsilon_t}(\epsilon_t) d\epsilon_t \int_{-\infty}^{\infty} |\epsilon_{t-j}| f_{\epsilon_{t-j}}(\epsilon_{t-j}) d\epsilon_{t-j}$$

For WN one has $\mathbb{E}[|\epsilon_t|] \leq \sigma$ for all t and therefore:

$$= \sum_{j=0}^{\infty} |\phi_1|^j \int_{-\infty}^{\infty} |\epsilon_t| f_{\epsilon_t}(\epsilon_t) d\epsilon_t \int_{-\infty}^{\infty} |\epsilon_{t-j}| f_{\epsilon_{t-j}}(\epsilon_{t-j}) d\epsilon_{t-j}$$

$$\leq \sum_{j=0}^{\infty} |\phi_1|^j \sigma^2 = \frac{\sigma^2}{1 - |\phi_1|} < \infty$$

when $|\phi_1| < 1$ and $\sigma^2 < \infty$.

Example $AR(1)$ ACVF at $k = 0$

e.g. lets take WN from a Gaussian distribution, giving:

$$
\begin{aligned}
&= \sum_{j=0}^{\infty} |\phi_1|^j \int_{-\infty}^{\infty} |\epsilon_t| f_{\epsilon_t}(\epsilon_t) d\epsilon_t \int_{-\infty}^{\infty} |\epsilon_{t-j}| f_{\epsilon_{t-j}}(\epsilon_{t-j}) d\epsilon_{t-j} \\
&= 4 \sum_{j=0}^{\infty} |\phi_1|^j \int_{0}^{\infty} \epsilon_t f_{\epsilon_t}(\epsilon_t) d\epsilon_t \int_{0}^{\infty} \epsilon_{t-j} f_{\epsilon_{t-j}}(\epsilon_{t-j}) d\epsilon_{t-j} \\
&= \sum_{j=0}^{\infty} |\phi_1|^j = \frac{1}{1 - |\phi_1|} < \infty \text{ if } |\phi_1| < 1.
\end{aligned}
$$

We will see that this condition of the coefficient $|\phi_1| < 1$ is also consistent with a causal representation of an AR(1) model as we will see later.

Example $AR(1)$ ACVF at $k = 0$

Now that we have established that we can swap order of integration and summation we can work with

$$\mathbb{E}\left[\epsilon_t Y_t\right] = \sum_{j=0}^{\infty} \phi_1^j \mathbb{E}\left[\epsilon_t \epsilon_{t-j}\right]$$

Now, recall that WN is an uncorrelated (no autocorrelation) process s.t.

$$\mathbb{E}\left[\epsilon_t Y_t\right] = \sum_{j=0}^{\infty} \phi_1^j \mathbb{E}\left[\epsilon_t \epsilon_{t-j}\right] = \phi_1^0 \mathbb{E}\left[\epsilon_t^2\right] = \sigma^2.$$

Hence, one gets

$$\begin{aligned} \gamma_Y(0) &= \mathbb{E}\left[(\phi_1 Y_{t-1} + \epsilon_t) Y_t\right] = \phi_1 \gamma_Y(1) + \sigma^2 = \phi_1^2 \gamma_Y(0) + \sigma^2 \\ \Rightarrow & \quad \gamma_Y(0) = \frac{\sigma^2}{1 - \phi_1^2} \end{aligned}$$

which completes the derivation of $\gamma(0)$ and we can conclude that AR(1) is stationary when $|\phi_1| < 1$

Third way to arrive at result is directly invoke Wold Representation and the corresponding known form of ACVF for a Wold representation of linear stationary process:

Remark (Linear Stationary Process as $MA(\infty)$!)

A linear stationary process can always be represented as an infinite combination of lagged WN terms (via Wold Theorem):

$$Y_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j} \ \ \text{with } \beta_0 = 1 \ \ \& \ \ \sum_{j=0}^{\infty} \beta_j^2 < \infty$$

Then, the resulting ACVF is given by

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k}$$

To associate our AR(1) process to the Wold representation, consider rewriting the AR(1) model in terms of a Backshift operator and end up with a Wold representation (linear combination of lagged WN terms)

Example

Consider $AR(1)$ process: $Y_t = \phi_1 Y_{t-1} + \epsilon_t$. In terms of backshift operator:

$$
\begin{aligned}
Y_t &= \phi_1 B Y_t + \epsilon_t \\
Y_t - \phi_1 B Y_t &= \epsilon_t \\
(1 - \phi_1 B) Y_t &= \epsilon_t \\
Y_t &= (1 - \phi_1 B)^{-1} \epsilon_t = \sum_{j=0}^{\infty} (\phi_1 B)^j \epsilon_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}
\end{aligned}
$$

provided $|\phi_1| < 1$. N.b. $\phi(B) := (1 - \phi_1 B)$ is called the $AR(1)$ charateristic polynomial/equation. Hence $AR(1)$ is stationary iff the root $(1/\phi_1)$ of $\phi(B)$ lies outside unit circle (i.e. if $|1/\phi_1| > 1$).

$AR(1)$ is $MA(\infty)$!

Hence, in this case for the AR(1) process we can identify that this agrees with Wold decomposition under the condition:

$$\beta_j = \phi_1^j \, .$$

Hence, we have that an $AR(1)$ process can be written as an infinite order moving average process $MA(\infty)$.
Then apply the known result for ACF of Wold Process for AR(1) setting as follows:

Corollary (mean and γ for $AR(1)$)

Let $\{Y_t\}$ be $AR(1)$, with $|\phi_1| < 1$. Then,

1. $\mathbb{E}(Y_t) = 0$
2. $\gamma(k) = \dfrac{\sigma^2 \phi_1^k}{1 - \phi_1^2}$

$AR(1)$ Model ACVF & ACF

From Wold:

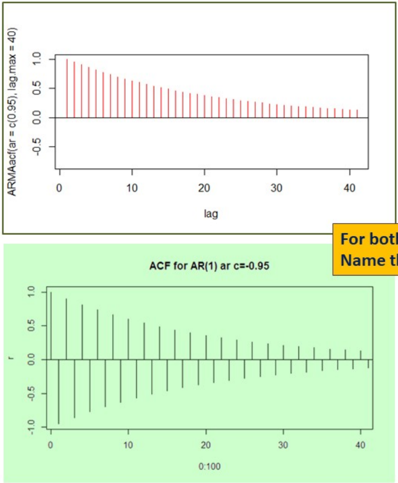
$$
\begin{aligned}
\gamma(k) &= \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k} \\
&= \sigma^2 \sum_{j=0}^{\infty} \phi_1^j \phi_1^{j+k}, \qquad [k \in \mathbb{N}, \text{ i.e. } k = 0, 1, 2, \ldots] \\
&= \sigma^2 \phi_1^k \sum_{j=0}^{\infty} (\phi_1^2)^j \\
&= \frac{\sigma^2 \phi_1^k}{1 - \phi_1^2} \qquad [\text{geo. prog., } |\phi_1| < 1] \quad \Rightarrow \text{Eqn } 2
\end{aligned}
$$

In particular $\mathrm{var}(Y_t) = \gamma(0) = \sigma^2/(1 - \phi_1^2)$. Hence:

$$
\rho(k) = \gamma(k)/\gamma(0) = \phi_1^k, \quad k \in \mathbb{N}
$$

and ACF decays exponentially with increasing lag k.

$AR(1)$ ACF

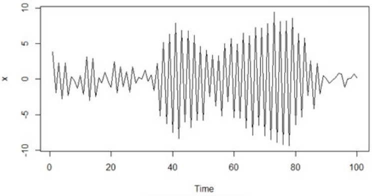


For both graphs $|\phi_1|$ = 0.95.
Name the graph with ϕ_1 <0.

Notice: ACFs of AR(1) decay exponentially, but are never zero!

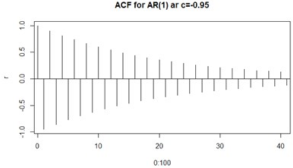
Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

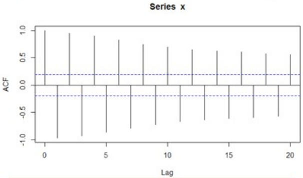
100 simulated values of AR(1) $X_t = -0.95 X_{t-1} + Z_t$ and its ACF



Correlations die out slowly because $\varphi = -0.95$ is near -1.
Signs alternate, exponential decay.

Graph choppy because of strong negative correlation.





(the same acf with confidence intervals)

Commands used in R:

```
> ar1 <- arima.sim(model=list(ar=c(-0.95)), n=100, sd=1)
> plot(ar1)
> plot(ARMAacf(ar=c(-0.95), lag.max=40), ylab="r",type="h", main="ACF for AR(1)  c=-0.95"); abline(h=0)
```

25

Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

$AR(1)$ ACF

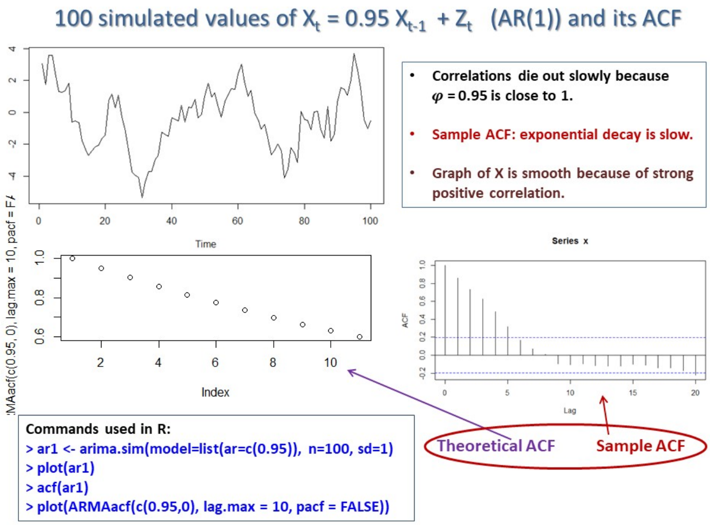
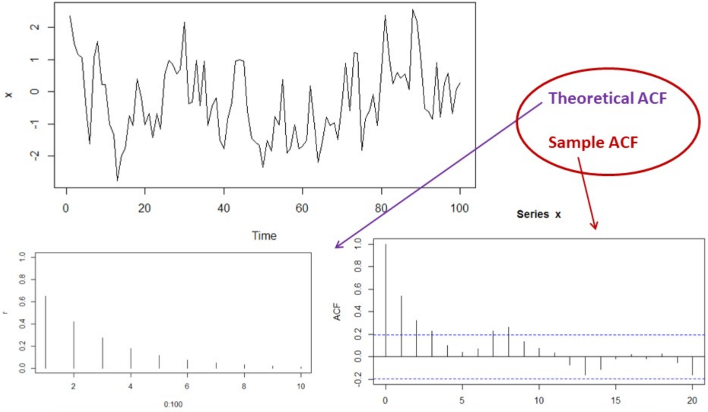


Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

100 simulated values of $X_t = 0.65\, X_{t-1} + Z_t$ (AR(1)) and its ACF



- ACF dies out faster than on the previous slide because φ=0.65 is not close to 1;
- Graph of X less smooth (less correlation, less linear dependence) .

Figure: (Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021)

Revisit Causality in the AR(1) Case.

Definition (causality)

A time series $\{Y_t\}$ is called <u>*causal*</u> *if it, and any noise terms $\{\epsilon_t\}$, satisfy:*

$$Y_{t-k} \perp \epsilon_t \,, \quad \forall k > 0$$

Remark

Causality means that 'future noise' is independent of present or past values of the time series. This should, intuitively, make sense. We will restrict our attention to causal models in this course.

Lemma

Let $\{Y_t\}$ be a causal time series and $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$, then $\mathbb{E}(Y_{t-k}\epsilon_t) = 0 \,, \forall k > 0$.

<u>Proof</u> $\forall k > 0$:

$$\begin{aligned} 0 &= \mathrm{cov}(Y_{t-k}, \epsilon_t)\,, \quad \text{since } \{Y_t\} \text{ causal } \Rightarrow Y_{t-k} \perp \epsilon_t \\ &= \mathbb{E}(Y_{t-k}\epsilon_t) - \mathbb{E}(Y_{t-k})\cancelto{0}{\mathbb{E}(\epsilon_t)} \quad \blacksquare \end{aligned}$$

Remark

- **Causality:** *is an important concept as it ensures that the time series representation of a process is FUTURE INDEPENDENT - hence has meaningful interpretation.*

- **Invertibility:** *observations in remote past have decreasing influence on current values of the process.*

We study closer the investigation on the AR(1) process in terms of causality representation. Consider $Y_t = \phi Y_{t-1} + \epsilon_t$ where $\epsilon \sim WN(0, \sigma^2)$

- When $\phi = 0$ it is trivial case that $Y_t = \epsilon_t$.
- When $|\phi| < 1$ we can show that a causal representation exists with

$$Y_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$$

- When $|\phi| > 1$ then such a causal representation does not converge and so does not exists (instead we will get the exact opposite of a causal relation representation).

Causal Linear Process AR(1) if $|\phi| < 1$

To show the causal representation holds for the parameter restriction in the unit disc $|\phi| < 1$ then we can go back to the characteristic equation $\phi(B) = 1 - \phi B$ for an AR(1) and consider $\Psi(B) = \sum_{j=0}^{\infty} \phi^j B^j$

$$\begin{aligned} Y_t - \phi Y_{t-1} &= \epsilon_t \\ \Rightarrow \Psi(B)(Y_t - \phi Y_{t-1}) &= \Psi(B)\epsilon_t \\ \Rightarrow \Psi(B)\phi(B) Y_t &= \Psi(B)\epsilon_t \\ \Rightarrow Y_t &= \Psi(B)\epsilon_t \end{aligned}$$

where the last step arises since

$$\Psi(B)\phi(B) = (1 - \phi B)\sum_{j=0}^{\infty} \phi^j B^j = \sum_{j=0}^{\infty} \phi^j B^j - \sum_{j=1}^{\infty} \phi^j B^j = 1$$

Hence, we see that this model has a causal relationship when we have $|\phi| < 1$.

We now show that if $|\phi| > 1$ then we wont have a causal representation for the AR(1) model $Y_t = \phi Y_{t-1} + \epsilon_t$ in this case. Note, in this case we cannot simply invert the AR(1) Characteristic polynomial as it will not be a convergent sequence.
However, we can rewrite the AR(1) model in this case as follows:

$$Y_t = -\phi^{-1}\epsilon_{t+1} + \phi^{-1}Y_{t+1}$$

Then we can iterate as follows via substitution:

$$Y_t = -\phi^{-1}\epsilon_{t+1} - \cdots - \phi^{-k-1}\epsilon_{t+k+1} + \phi^{-k-1}Y_{t+k+1}$$

to get the unique representation (in terms of future WN values – opposite of causality):

$$Y_t = -\sum_{j=1}^{\infty} \phi^{-j}\epsilon_{t+j}$$

This is hard to interpret since Y_t now is defined to be correlated with future values of WN terms.

Note, we could however define a new sequence as follows

$$
\begin{aligned}
\epsilon_t^* &= Y_t - \phi^{-1} Y_{t-1} \\
&= \phi Y_{t-1} + \epsilon_t - \phi^{-1} Y_{t-1} \\
&= (\phi - \phi^{-1}) Y_{t-1} + \epsilon_t \\
& \quad \text{now substitute } Y_{t-1} = -\sum_{j=1}^{\infty} \phi^{-j} \epsilon_{t-1+j} \\
&= -(\phi - \phi^{-1}) \sum_{j=1}^{\infty} \phi^{-j} \epsilon_{t-1+j} + \epsilon_t \\
&= -(\phi - \phi^{-1})\phi^{-1} \epsilon_t + \epsilon_t - (\phi - \phi^{-1}) \sum_{j=2}^{\infty} \phi^{-j} \epsilon_{t-1+j} \\
&= \phi^{-2} \epsilon_t - \frac{(1 - \phi^{-2})}{\phi} \sum_{j=2}^{\infty} \phi^{-j+1} \epsilon_{t-1+j} \\
& \quad \text{change index variable } -\tilde{j} = -j + 1 \\
&= \phi^{-2} \epsilon_t - (1 - \phi^{-2}) \sum_{\tilde{j}=1}^{\infty} \phi^{-\tilde{j}} \epsilon_{t+\tilde{j}}
\end{aligned}
$$

One can then show that WN sequence $\{\epsilon^*\}$ has

$$\mathbb{E}\left[\epsilon_t^*\right] = 0, \quad \gamma_{\epsilon^*}(k) = \frac{\sigma^2}{\phi^2}\mathbb{I}[k=0]$$

which means that with this new WN sequence we could rewrite the AR(1) model for $|\phi| > 1$ according to new parameters $|\phi^*| = 1/|\phi| < 1$ to get causal representation

$$Y_t = \phi^* Y_{t-1} + \epsilon_t^* = \sum_{j=0}^{\infty} \phi^{*j} \epsilon_{t-j}^*.$$

which under this transform does not depend on future values.

- Thus, for an AR(1) model, one typically assumes that $|\phi| < 1$.

Example: $AR(1)$ as RW with drift

Here we will consider an example of an AR(1) model with the boundary case $|\phi| = 1$ and a constant DC level, and we will show that in this case the resulting Causal representation will result in a RW model (causal representation) with addition of a time drift (temporal trend) that arises in this instance $\Rightarrow$ non-stationary.

Example

Let Y_t be a time series given by model:

$$Y_t = \beta + Y_{t-1} + \epsilon_t$$
$$\Rightarrow Y_t(1 - B) = \beta + \epsilon_t$$

for $t = 1, 2, 3, \ldots$ with $X_0 = 0$ and $\epsilon \sim WN(0, \sigma^2)$.

We can re-express the process as follows:

$$X_t = \beta t + \sum_{j=1}^{t} \epsilon_j$$

```
set.seed(150); w = rnorm(200,0,1); x = cumsum(w);
wd = w +.2; xd = cumsum(wd); par(mar=c(4,4,2,.5))
plot.ts(xd, ylim=c(-5,45), main="random walk",col="blue")
lines(x); lines(.2*(1:200), lty="dashed",col="blue")
```

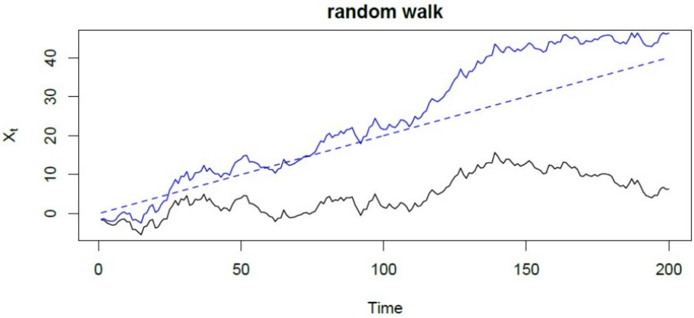


Figure Random walk, $\sigma = 1$, with drift $\delta = 0.2$ (upper jagged line), without drift, $\delta = 0$ (lower jagged line), and a straight line with slope .2 (dashed line).

Figure: (source: Dewei Wang https://people.stat.sc.edu/wang52)

We have learnt that we refer to a causal autoregressive process $\{Y_t\}$ given by $Y_t = \phi Y_{t-1} + \epsilon_t$ if it admits a unique linear representation in terms of a WN process $\{\epsilon_s, s \leq t\}$, such that the current time t behaviour of the process is only expressed in terms of the past event "shocks" or innovations and not the future.

- If $|\phi| < 1$ then such an AR(1) process $\{Y_t\}$ is a causal function of $\{\epsilon_t\}$.
- If $|\phi| > 1$, AR(1) process is not causal.

Proposition

AR(1) process $\phi(B)Y_t = \epsilon_t$ with $\phi(B) = 1 - \phi B$ is causal iff $|\phi| < 1$ or the root z_1 of the characteristic polynomial $\phi(z) = 1 - \phi z$ satisfies $|z_1| > 1$

Higher Order AR(p) Models ($p > 1$): Invertibility, Stationarity and Causality.

$AR(p)$ Models

Recall that the AR(p) model is generically given by

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t, \;\; \epsilon_t \sim WN(0, \sigma_2)$$

For example

- AR(2): $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$
- AR(3): $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \epsilon_t$

What order AR process is

$$Y_t = -0.2 Y_{t-1} - 0.7 Y_{t-2} + \epsilon_t \;\; ?$$

or

$$Y_t = 0.6 Y_{t-1} + 0.08 Y_{t-2} + 0.3 Y_{t-5} + \epsilon_t \;\; ?$$

Note: $\phi_0 = 1$

Example

An $AR(2)$ model can be represented as an $MA(\infty)$ model.

$AR(2)$: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$ can be written, in terms of the backshift operator, as:

$$(1 - \phi_1 B - \phi_2 B^2) Y_t = \epsilon_t \, .$$

Write the (2) characteristic polynomial as

$$\phi(B) = (1 - \lambda_1 B)(1 - \lambda_2 B) \, ,$$

—here λ_1^{-1} and λ_2^{-1} are roots of $\phi(B)$. Then

$$\begin{aligned}
(1 - \lambda_1 B)(1 - \lambda_2 B) Y_t &= \epsilon_t \\
Y_t &= \left((1 - \lambda_1 B)(1 - \lambda_2 B)\right)^{-1} \epsilon_t \\
Y_t &= \frac{1}{(1 - \lambda_1 B)(1 - \lambda_2 B)} \epsilon_t \, ,
\end{aligned}$$

Now use partial fractions to deal with the RHS.

Partial fractions: find α_1, α_2, s.t.

$$\frac{1}{(1-\lambda_1 x)(1-\lambda_2 x)} = \frac{\alpha_1}{1-\lambda_1 x} + \frac{\alpha_2}{1-\lambda_2 x}$$

Multiply both sides by denominator of LHS; then put $x = \lambda_1^{-1}$ to get $\alpha_1 = \lambda_1(\lambda_1 - \lambda_2)^{-1}$ and put $x = \lambda_2^{-1}$ to get $\alpha_2 = -\lambda_2(\lambda_1 - \lambda_2)^{-1}$ i.e.

$$\frac{1}{(1-\lambda_1 x)(1-\lambda_2 x)} = \frac{1}{\lambda_1 - \lambda_2}\left(\frac{\lambda_1}{1-\lambda_1 x} - \frac{\lambda_2}{1-\lambda_2 x}\right)$$

I.e. $$Y_t = \frac{1}{\lambda_1 - \lambda_2}\left(\lambda_1(1-\lambda_1 B)^{-1} - \lambda_2(1-\lambda_2 B)^{-1}\right)\epsilon_t$$

Now, provided $|\lambda_1|, |\lambda_2| < 1$:

$$\begin{aligned}
Y_t &= \frac{1}{\lambda_1 - \lambda_2}\left(\lambda_1 \sum_{j=0}^{\infty} \lambda_1^j B^j - \lambda_2 \sum_{j=0}^{\infty} \lambda_2^j B^j\right)\epsilon_t \\
&= \frac{1}{\lambda_1 - \lambda_2} \sum_{j=0}^{\infty} (\lambda_1^{j+1} - \lambda_2^{j+1}) B^j \epsilon_t \\
&= \frac{1}{\lambda_1 - \lambda_2} \sum_{j=0}^{\infty} (\lambda_1^{j+1} - \lambda_2^{j+1}) \epsilon_{t-j} . \qquad \blacksquare
\end{aligned}$$

Roots of $AR(p)$ char. eqn. outside circle $\Leftrightarrow$ stationarity

Theorem (AR Char. Eqn and Stationarity)

Consider $AR(p)$ process: $Y_t = \epsilon_t + \sum_{j=1}^{p} \phi_j Y_{t-j}$. Then $\{Y_t\}$ is (weakly) stationary iff all roots of characteristic equation

$$\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$$

are outside unit circle.

Sketch Proof In terms of backshift operator, $Y_t - \sum_{j=1}^{p} \phi_j Y_{t-j} = \epsilon_t$ can be written:

$$\begin{aligned} (1 - \sum_{j=1}^{p} \phi_j B^j) Y_t &= \epsilon_t \\ \phi(B) Y_t &= \epsilon_t \, , \qquad \phi(B) \text{ is a } p\text{-order polynomial} \end{aligned}$$

AR(p) Characteristic Polynomials

Factorising (the AR characteristic polynomial) $\phi(B)$:

$$\begin{aligned}
\lambda_0 \prod_{j=1}^{p} (\lambda_j - B) Y_t &= \epsilon_t \\
\lambda_0 \prod_{j=1}^{p} \lambda_j (1 - B/\lambda_j) Y_t &= \epsilon_t
\end{aligned}$$

Hence, we can write Y_t as:

$$Y_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j}, \quad \beta_0 = 1, \quad \sum_{j=0}^{\infty} \beta_j^2 < \infty, \quad \text{c.f. Wold}$$

(and therefore claim $\{Y_t\}$ is weakly stationary) iff all roots of AR characteristic polynomial have magnitude greater than one. E.g. if $|\lambda_k| > 1$, then $|1/\lambda_k| < 1$, and we have (via geometric progression):

$$\begin{aligned}
\lambda_0 \prod_{j \neq k} \lambda_j (1 - B/\lambda_j) Y_t &= (1 - B/\lambda_k)^{-1} \epsilon_t \\
&= \sum_{j=0}^{\infty} \left(\frac{1}{\lambda_k} B \right)^j \epsilon_t = \sum_{j=0}^{\infty} \left(\frac{1}{\lambda_k} \right)^j \epsilon_{t-j} \qquad \blacksquare
\end{aligned}$$

summary: the following are basic properties of the AR(p) model:

- AR(p) is always invertible by its construction: $\epsilon_t = \phi(B)Y_t$
- AR(p) model admits an MA(∞) representation and is stationary and causal when the Characteristic polynomial

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \neq 0, \ |z| \leq 1,$$

that is, the roots of $\phi(z)$ lie outside the unit circle.

Lets now prove the causal aspect of this condition on the roots of the AR(p) char. polynomial.

Theorem

A (unique) stationary solution to $\phi(B)Y_t = \epsilon_t$ *exists for* Y_t *expressed in terms of* $\{\epsilon_t\}$ *if and only if*

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \quad \Rightarrow \quad |z| \neq 1.$$

Furthermore, this AR(p) process is causal if and only if

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \quad \Rightarrow \quad |z| > 1$$

We have stated that when $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \quad \Rightarrow \quad |z| > 1$ is satisfied, based on causality, we can write $Y_t = \Psi(B)\epsilon_t$ where $\Psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$ for some ψ_j's satisfying $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Proof.

This means that $\phi(B) Y_t = \epsilon_t$ & $Y_t = \Psi(B)\epsilon_t$, hence we can show:

$$\begin{aligned}
& \forall z \in \mathbb{C},\ |z| \leq 1 \rightarrow \phi(z) \neq 0 \\
\Leftrightarrow \quad & \exists \left\{\psi_j\right\},\ \ \delta > 0,\ \ \forall |z| \leq 1 + \delta,\ \ \frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j \\
& \Rightarrow \forall |z| \leq 1 + \delta,\ \ |\psi_j z^j| \rightarrow 0,\ \ \left(|\psi_j|^{1/j} |z|\right)^j \rightarrow 0 \\
& \Rightarrow \exists j_0,\ \ \forall j \geq j_0,\ \ |\psi_j|^{1/j} \leq \frac{1}{1 + \delta/2} \ \ \Rightarrow \sum_{j=0}^{\infty} |\psi_j| < \infty.
\end{aligned}$$

So if $|z| \leq 1 \ \Rightarrow \phi(z) \neq 0$, then $S_m = \sum_{j=0}^{m} \psi_j B^j \epsilon_t$ converges in mean square, so we have a stationary, causal time series given by

$$Y_t = \phi^{-1}(B)\epsilon_t$$

$AR(p)$: finding the causal coefficients?

When $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \quad \Rightarrow \quad |z| > 1$ is satisfied, based on causality, we can write

$$Y_t = \Psi(B)\epsilon_t$$

where $\Psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$ for some ψ_j's satisfying $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

The practical question then becomes how to calculate the coefficients ψ_j's?

We could do this for instance via coefficient matching between equations:

$$\begin{aligned}
&\phi(B)Y_t = \epsilon_t \;\; \& \;\; Y_t = \Psi(B)\epsilon_t \\
&\Rightarrow 1 = \Psi(B)\phi(B) \\
&\Rightarrow 1 = (\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - \phi_1 B - \cdots - \phi_p B^P) \\
&\Rightarrow 1 = \psi_0 \\
&\quad 0 = \psi_1 - \phi_1\psi_0 \\
&\quad 0 = \psi_2 - \phi_1\psi_1 - \phi_2\psi_0 \\
&\quad \vdots
\end{aligned}$$

$AR(p)$: finding the causal coefficients?

We can solve these *linear difference equations*

$$\Rightarrow 1 = \psi_0,\ 0 = \psi_j\ (j < 0),\ 0 = \phi(B)\psi_j\ (j > 0),$$

in several ways:

- numerically, or
- guessing the form of the solution and using an inductive proof, or
- using the theory of linear difference equations.

Example ($AR(2)$ process)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t \tag{0.5}$$

Writing in terms of backshift operator:

$$\begin{aligned} Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} &= \epsilon_t \quad \text{(invertible)} \\ (1 - \phi_1 B - \phi_2 B^2) Y_t &= \epsilon_t \\ \phi(B) Y_t = \epsilon_t \end{aligned}$$

Now, $\{Y_t\}$ is stationary iff roots of $\phi(B) = (1 - \phi_1 B - \phi_2 B^2)$ lie outside unit circle. It can be shown that this is true iff:

$$\phi_1 + \phi_2 < 1\,, \quad \phi_2 - \phi_1 < 1\,, \quad |\phi_2| < 1 \tag{0.6}$$

Note also, if $\mu := \mathbb{E}(Y_t)$, then take $\mathbb{E}$ of (0.5):

$$\mu = (\phi_1 + \phi_2)\mu$$

But (0.6) $\Rightarrow (\phi_1 + \phi_2) \neq 1$. Hence $\mu = 0$.

Example ($AR(2)$ process example)

Consider the case $\phi_1 = 1/6, \phi_2 = -1/6$*:*

$$Y_t = -\frac{1}{6} Y_{t-1} + \frac{1}{6} Y_{t-2} + \epsilon_t \, .$$

Is this process stationary?

Writing in terms of backshift operator:

$$\begin{aligned} Y_t + \frac{1}{6} Y_{t-1} - \frac{1}{6} Y_{t-2} &= \epsilon_t \\ (1 + \frac{1}{6} B - \frac{1}{6} B^2) Y_t &= \epsilon_t \end{aligned}$$

Factorise AR characteristic polynomial:

$$1 + \frac{1}{6} B - \frac{1}{6} B^2 = -\frac{1}{6}(B-2)(B+3) \, .$$

$\Rightarrow$ roots at 2 and -3. Hence, Y_t is stationary. Also note that

$$\phi_1 + \phi_2 < 1 \, , \quad \phi_2 - \phi_1 < 1 \, , \quad |\phi_2| < 1$$

are satisfied.

Lets consider a second example AR(2) time series model

$$Y_t = 1.3Y_{t-1} - 0.7Y_{t-2} + \epsilon_t$$

This AR(2) process see in this case the roots are outside the unit circle using for instance quadratic equation or the R command polyroot.

Roots of $\phi(z) = 1 - 1.3z + 0.7z^2$ *corresponding to*
AR(2): $X_t = 1.3X_{t-1} - 0.7X_{t-2} + Z_t$.
Roots are outside unit circle:

```
> polyroot(c(1, -1.3, 0.7))
0.9285714+0.7525467i
0.9285714-0.7525467i
```

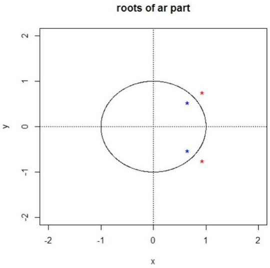


Figure: Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021

Properties of AR(p):

- $AR(p)$ is always invertible by its construction: $\epsilon_t = \phi(B)Y_t$.
- Has $MA(\infty)$ representation when

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \neq 0 \text{ for } |z| \leq 1,$$

  that is, the roots of the Char. Fn. polynomial $\phi(z)$ lie outside of the unit cirlce.
- Stationary when it has $MA(\infty)$ representation, that is, if $\phi(z) \neq 0$ for $|z| \leq 1$

Summary of Understanding of Model Characterisation by Charts from Correlograms (ACF).

Summary of ACF structures

Time Series Model	Sample ACF Behaviour
White Noise WN	Zero for lags $\|k\| > 0$
Trend	Slow decay
Periodic	Periodic decay
MA(q)	Zero for lags $\|k\| > q$
AR(1)	Decays to zero exponentially

Shortly, we will see yet another way to find the ACF of AR(p) models via the Yule-Walker equations, which is also useful for estimation of model parameters (coefficients of model).

Some basic examples in R to illustrate ACF examples

```
set.seed(100);
par(mfrow=c(5,2))
par(mar=c(4,4,2,.5))
     #White Noise
WN=rnorm(100,0,1);
plot(1:n,WN,type="o",col="blue",main="White Noise",ylab=expression(X[t]),xlab="t");
acf(WN)
```

```
     #Trend
t=seq(1,100,1);
Tt=1+.1*t;
Xt=Tt+rnorm(length(t),0,4)
plot(t,Xt,xlab="t",ylab=expression(X[t]),main="Trend")
lines(t,Tt,col="blue")
acf(Xt)
```

```
     #MA(1)
w = rnorm(550,0,1)
v = filter(w, sides=1, c(1,.6))[-(1:50)]
plot.ts(v, main="MA(1)",col="blue",ylab=expression(X[t]),xlab="t")
acf(v)
```

```
     #Periodic
t=seq(1,150,1)
St=2*cos(pi*t/5)+3*sin(pi*t/3)
Xt=St+rnorm(length(t),0,2)
plot(t,Xt,xlab="t",ylab=expression(X[t]), main="Periodic")
lines(t,St,col="blue")
acf(Xt)
```

```
     #AR(1)
w = rnorm(550,0,1)
x = filter(w, filter=c(.6), method="recursive")[-(1:50)]
plot.ts(x, main="AR(1)",col="blue",ylab=expression(X[t]),xlab="t")
acf(x)
```

Figure: (source: Dewei Wang https://people.stat.sc.edu/wang52)

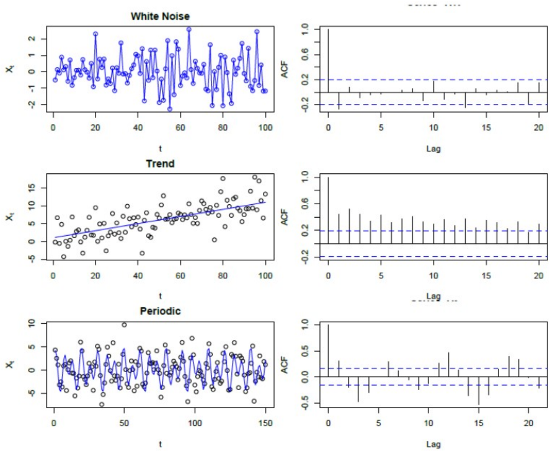


Figure: (source: Dewei Wang https://people.stat.sc.edu/wang52)

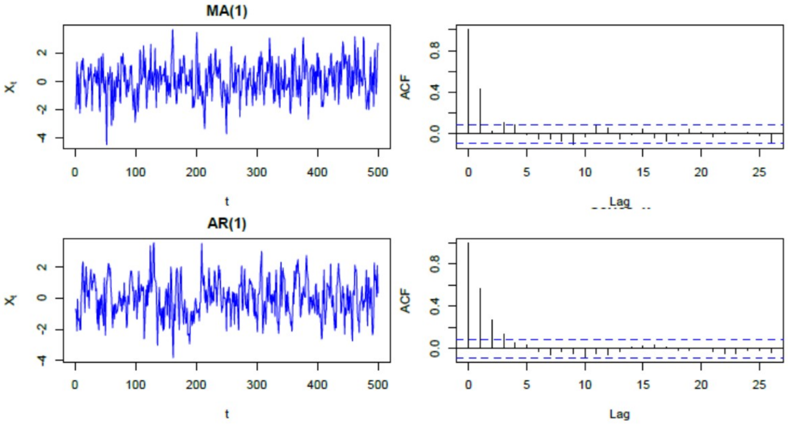


Figure: (source: Dewei Wang https://people.stat.sc.edu/wang52)

Summary of ACF structures for AR(p)

For an $AR(p)$ process, it can be shown that if the Characteristic polynomial has:

- real roots $\Rightarrow$ exponentially decaying ACF; and
- non-real roots $\Rightarrow$ an ACF with exponential decay multiplied by sum of sinusoids.

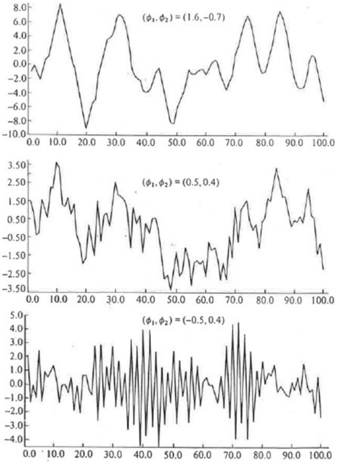


Figure 10.5 Realizations of the process: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$

Compare graphs of AR(2) models for different combinations of signs of coefficients.

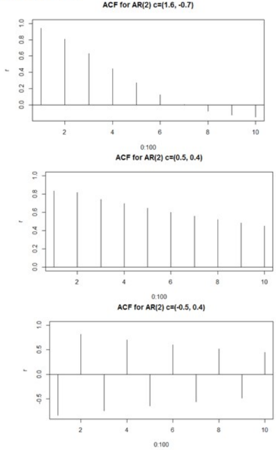
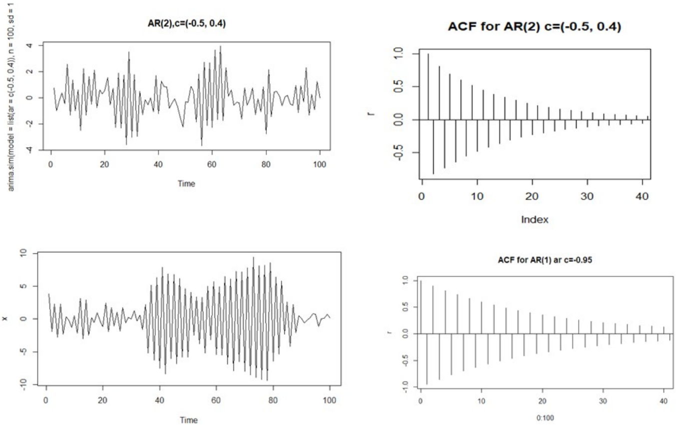


Figure: Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021

Example $AR(2)$ process



Compare plots and graphs ACFs for AR(1) and AR(2) models.
Can we guess the order from these plots?

Figure: Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021

Some simple R commands used to create previous slides

- To simulate 100 values of AR(1): $X_t = -0.95X_{t-1} + Z_t$

```
ar1 <- arima.sim(model=list(ar=c(-0.95)), n=100, sd=1)
plot(ar1)
```

- To plot ACF for simulated values stored in file ar1:

`acf (ar1, type="correlation", plot=T)` or simply `acf(ar1)`

- To plot theoretical ACF for this AR(1)

```
plot(ARMAacf(ar=c(-0.95), lag.max=40), col="red", type="h", xlab="lag", ylim=c(-.8,1));
abline(h=0)
```

Figure: Source: Dr. Raya Feldman, UCSB Time Series Lectures 2021

Yule-Walker Equations,
& Partial Autocorrelation Function (PACF)

Remark

Assuming that $\{Y_t\}$ is stationary, the ACF of an $AR(p)$ process can be found via the system of equation based on the Autocorrelation known as the Yule-Walker equations.

Example ($AR(2)$ process)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$$

Multiply both sides by Y_{t-k} and take expectations (n.b. $\{Y_t\}$ causal):

$$\mathbb{E}(Y_{t-k}Y_t) = \phi_1 \mathbb{E}(Y_{t-k}Y_{t-1}) + \phi_2 \mathbb{E}(Y_{t-k}Y_{t-2}) + \cancel{\mathbb{E}(Y_{t-k}\epsilon_t)} \rightarrow 0,\ k>0$$

Now, for an $AR(2)$, we know that $\mathbb{E}(Y_t) = 0$. Hence,

$$\text{cov}(Y_t, Y_{t-k}) = \mathbb{E}(Y_t Y_{t-k}) - \cancel{\mathbb{E}(Y_t)\mathbb{E}(Y_{t-k})} \rightarrow 0,$$

I.e. for $k > 0$:

$$\begin{aligned} \gamma(k) &= \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) \\ \rho(k) &= \phi_1 \rho(k-1) + \phi_2 \rho(k-2), \qquad \text{divide by } \gamma(0) \end{aligned}$$

Yule Walker Equations: AR(2) Example

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2)$$

E.g., put $k = 1$:

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

But, $\rho(0) = 1$, and $\rho(-k) = \rho(k) \Rightarrow \rho(-1) = \rho(1)$:

$$\rho(1) = \phi_1 + \phi_2 \rho(1) \Rightarrow \rho(1) = \frac{\phi_1}{1-\phi_2}$$

We can continue by putting $k = 2$ to get:

$$\rho(2) = \phi_1 \rho(1) + \phi_2 \rho(0)$$

I.e.

$$\rho(2) = \frac{\phi_1^2}{1-\phi_2} + \phi_2 \,,$$

and so on. (Exercise: compute $\rho(3)$).

Yule Walker Equations: AR(p)

In general, the Yule-Walker equations of $AR(p) : Y_t = \epsilon_t + \sum_{j=1}^{p} \phi_j Y_{t-j}$ can also be found in the same manner.

Assuming stationarity, multiply by Y_{t-k}, take $\mathbb{E}$, and divide by $\gamma(0)$:

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \cdots + \phi_p \rho(k-p)$$

We can then solve these difference equations with standard methods. These can be written as:

$$\begin{pmatrix} \rho(1) \\ \vdots \\ \rho(p) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \rho(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix}$$

I.e.

$$\boldsymbol{\rho} = \boldsymbol{\Gamma} \boldsymbol{\phi} \,,$$

with

$$\boldsymbol{\rho} := (\rho(k))_{k=1}^{p} \,, \quad \boldsymbol{\phi} := (\phi_j)_{j=1}^{p} \,, \quad \boldsymbol{\Gamma} := (\rho(k-j))_{k,j=1}^{p}$$

Yule Walker Equations: AR(p)

Note that this is a well-posed system of equations (with a square coefficients matrix $\mathbf{\Gamma}$), i.e., with the same number of constraints (equations, $\mathbf{\Gamma}$'s rows) as unknowns (the elements ϕ_j of the unknown vector $\boldsymbol{\phi}$).

Further, $\mathbf{\Gamma}$ is full-rank and symmetric, so that invertability is guaranteed,

$$\widehat{\boldsymbol{\phi}} = \mathbf{\Gamma}^{-1}\boldsymbol{\rho}$$

Remark

A matrix is full row rank when each of the rows of the matrix are linearly independent and full column rank when each of the columns of the matrix are linearly independent.

- Since in this case $\mathbf{\Gamma}$ is a square matrix, we can test that it is full rank by just computing the determinant. If $det(\mathbf{\Gamma}) \neq 0$, then it is a full rank matrix. (prove for yourself this is the case for $\mathbf{\Gamma}$.

The partial autocorrelation function (PACF) gives the partial correlation of a stationary time series with its own lagged values, regressed against the values of the time series at all shorter lags. It contrasts with the autocorrelation function, which does not control for other lags.

- This function plays an important role in data analysis aimed at identifying the extent of the lag in an autoregressive model i.e. the order p of an AR(p) model.

Given a time series Y_t, the PACF of lag k, denoted $\alpha(k)$, is

- the autocorrelation between Y_t and Y_{t+k} after having removed the linear dependence of Y_t on Y_{t+1} through Y_{t+k-1} removed. (OR equivalently)
- the autocorrelation between Y_t and Y_{t+k} that is not accounted for by lags 1 through to $k-1$, inclusive.

$$\begin{aligned}
\alpha(1) &= corr\left(Y_{t+1}, Y_t\right), \quad k = 1 \\
\alpha(k) &= corr\left(Y_{t+k} - P_{t,k}(Y_{t+k}), Y_t - P_{t,k}(Y_t)\right), \quad k \geq 2,
\end{aligned}$$

where we denote $P_{t,k}(x)$ as the surjective (onto function) operator of orthogonal projection of x onto the linear subspace of Hilbert space spanned by $Y_{t+1}, \ldots, Y_{t+k-1}$.

So we see that like the autocorrelation function, the PACF is another tool that conveys vital information regarding the dependence structure of a stationary process and depends only on the second order properties of the process.

To be more explicit we can state:

The partial autocorrelation function PACF $\alpha_Y(\cdot)$ of a stationary time series is defined by

$$\alpha_Y(1) = \phi_{11} = Corr(Y_2, Y_1) = \rho_Y(1)$$

and

$$\alpha_Y(k) = \phi_{kk} = Corr\left(Y_{k+1} - \mathbb{E}\left[Y_{k+1}|Y_k, \dots, Y_2\right], Y_1 - \mathbb{E}\left[Y_1|Y_k, \dots, Y_2\right]\right).$$

value $\alpha_Y(k)$ is the partial autocorrelation of $\{Y_t\}$ at lag k.

Note PACF $\alpha_Y(k)$ can be regarded as the correlation between Y_1 and Y_{k+1}, adjusted for the intervening observations $Y_2, \dots, Y_k$

It is also possible to rewrite this definition of $\alpha_Y(k)$ above, which was based on $\{Y_1, Y_2, \dots, Y_k, Y_{k+1}\}$. However, for a stationary process it is equivalent to the one based on $\{Y_{t+1}, Y_{t+2}, \dots, Y_{t+k}, Y_{t+k+1}\}$ for any $t > 0$ i.e re-stated as

$$\alpha_Y(k) = \phi_{kk} = Corr\left(Y_{t+k+1} - \mathbb{E}\left[Y_{t+k+1}|Y_{t+k}, \dots, Y_{t+2}\right], Y_{t+1} - \mathbb{E}\left[Y_{t+1}|Y_{t+k}, \dots, Y_{t+2}\right]\right).$$

The Yule-Walker system of equations $\boldsymbol{\rho} = \boldsymbol{\Gamma}\boldsymbol{\phi}$, provides a convenient recursion for computing what is known as the Partial Autocorrelation Function (PACF).

The first step is to compute the acf up to a reasonable cutoff, say $p \simeq N/4$.

Next, let $\boldsymbol{\rho}^{(i)}$ denote the value from the Yule-Walker system for the $k = i$ case and $\boldsymbol{\Gamma}^{(i)}$ denote the coefficient matrix components for the same case.
Then to compute the PACF iterate as follows

- Loop on i, $1 \leq i \leq p$
  - Computer $\boldsymbol{\Gamma}^{(i)}$ and $\boldsymbol{\rho}^{(i)}$
  - Invert for $\widehat{\boldsymbol{\phi}}^{(i)}$ given by

$$\widehat{\boldsymbol{\phi}}^{(i)} = \left(\boldsymbol{\Gamma}^{(i)}\right)^{-1} \boldsymbol{\rho}^{(i)} = \left[\widehat{\phi}_1, \widehat{\phi}_2, \ldots, \widehat{\phi}_i\right]^T$$

  - Discard all $\widehat{\phi}_j$ for $1 \leq j \leq i-1$, retaining $\widehat{\phi}_i$
  - Set PACF $\alpha(i) = \widehat{\phi}_i$
- End loop on i
- Plot pacf $\alpha(i)$ as a function of i.

AR(p) process PACF Example

Example: Consider the AR(p) process

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

then

$$\mathbb{E}\left[Y_t | Y_{t-1}, \ldots, Y_{t-p}\right] = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p}$$

Then

$$\phi_{kk} = \begin{cases} \phi_k, & \text{if } 1 \leq k \leq p \\ 0, & \text{otherwise .} \end{cases}$$

Estimation: $\widehat{\phi}_k = \widehat{\boldsymbol{\Gamma}}^{-1} \widehat{\gamma}_Y(k)$ which gives a recursive system to be solved (Durbin-Levinson can be used).

In more detail, we can explore the PACF as follows....

(How) can we design something like the ACF which cuts-off at pth lag for an $AR(p)$ process (in the same way the ACF cuts-off at qth lag for an $MA(q)$ process)?

Remark

$AR(p)$ processes do not have an ACF that cuts-off at lag p because future values of series depend (indirectly) on **all** *previous values.*

Example

Recall $AR(1)$: $Y_t = \phi_1 Y_{t-1} + \epsilon_t$, with ACF $\rho(k) = \phi_1^{|k|}$.

Using successive substitution:

$$\begin{aligned} Y_2 &= \phi_1 Y_1 + \epsilon_2 \\ Y_3 &= \phi_1 Y_2 + \epsilon_3 = \phi_1^2 Y_1 + \phi_1 \epsilon_2 + \epsilon_3 \,, \end{aligned}$$

i.e. Y_3 depends on Y_1.

In general, Y_t depends on $Y_{t-1}, Y_{t-2}, \ldots$ I.e. Y_t depends on **all** previous values of the process $\{Y_t\}$.

- Consider representing an $AR(1)$ process as a linear combination of previous two values:

$$Y_t = \phi_{2,1} Y_{t-1} + \phi_{2,2} Y_{t-2} + \epsilon_t \,, \quad \text{for some } \phi_{2,1}, \phi_{2,2} \in \mathbb{R} \,.$$

  Then, $\phi_{2,2}$ represents any linear dependence that Y_t has on Y_{t-2} which is not accounted for by Y_{t-1}.

- For an $AR(1)$ process $Y_t = \phi_1 Y_{t-1} + \epsilon_t$, we have, by definition, that $\phi_{2,1} = \phi_1$ and $\phi_{2,2} = 0$.
- If we now consider representing this $AR(1)$ by:

$$Y_t = \phi_{1,1} Y_{t-1} + \epsilon_t \quad \phi_{1,1} \in \mathbb{R} \,,$$

  then, again, $\phi_{1,1} = \phi_1 \neq 0$.

Remark

The numbers $\phi_{1,1}$ and $\phi_{2,2}$ are the first two partial autocorrelation coefficients (PACF). Note that, for the $AR(1)$ process, the PACF cuts off at lag 2.

Definition

Consider (the problem of estimating) the coefficients $\phi_{k,j} \in \mathbb{R}$, *where*

$$\begin{aligned}
Y_t &= \phi_{1,1} Y_{t-1} + \epsilon_t \\
Y_t &= \phi_{2,1} Y_{t-1} + \phi_{2,2} Y_{t-2} + \epsilon_t \\
Y_t &= \phi_{3,1} Y_{t-1} + \phi_{3,2} Y_{t-2} + \phi_{3,3} Y_{t-3} + \epsilon_t \\
&\vdots \\
Y_t &= \phi_{k,1} Y_{t-1} + \phi_{k,2} Y_{t-2} + \ldots + \phi_{k,k} Y_{t-k} + \epsilon_t
\end{aligned}$$

The sequence $\{\phi_{k,k}\} = \{\phi_{1,1}, \phi_{2,2}, \phi_{3,3}, \ldots\}$ *is called the partial autocorrelation function (PACF) coefficients.*

Example

For $AR(1)$, *by definition,* $\phi_{1,1} \neq 0$, *and* $\phi_{2,2}, \phi_{3,3}, \ldots = 0$.

Example

For $AR(2)$, *by definition,* $\phi_{1,1}, \phi_{2,2} \neq 0$, *and* $\phi_{3,3}, \phi_{4,4}, \ldots = 0$.

Remark

For (p), by definition:

$$\begin{aligned}\phi_{k,k} &\neq 0, \quad \textit{for } k = 1, \ldots p \\ \phi_{k,k} &= 0, \quad \textit{for } k \geq p+1\end{aligned}$$

Recall

$$Y_t = \phi_{k,1} Y_{t-1} + \phi_{k,2} Y_{t-2} + \ldots + \phi_{k,k} Y_{t-k} + \epsilon_t \tag{0.7}$$

Multiply both sides by Y_{t-j}, for $j \geq 0$:

$$Y_{t-j} Y_t = \phi_{k,1} Y_{t-j} Y_{t-1} + \phi_{k,2} Y_{t-j} Y_{t-2} + \ldots + \phi_{k,k} Y_{t-j} Y_{t-k} + Y_{t-j} \epsilon_t$$

Take expectations of both sides:

$$\gamma(j) = \phi_{k,1} \gamma(j-1) + \phi_{k,2} \gamma(j-2) + \ldots + \phi_{k,k} \gamma(j-k)$$

Divide by $\gamma(0)$:

$$\rho(j) = \phi_{k,1} \rho(j-1) + \phi_{k,2} \rho(j-2) + \ldots + \phi_{k,k} \rho(j-k)$$

$$\rho(j) = \phi_{k,1}\rho(j-1) + \phi_{k,2}\rho(j-2) + \ldots + \phi_{k,k}\rho(j-k)$$

For $j = 1$:

$$\rho(1) = \phi_{k,1}\rho(0) + \phi_{k,2}\rho(1) + \phi_{k,3}\rho(2) + \phi_{k,4}\rho(3) + \ldots + \phi_{k,k}\rho(k-1)$$

For $j = 2$:

$$\rho(2) = \phi_{k,1}\rho(1) + \phi_{k,2}\rho(0) + \phi_{k,3}\rho(1) + \phi_{k,4}\rho(2) + \ldots + \phi_{k,k}\rho(k-2)$$

For $j = 3$:

$$\rho(3) = \phi_{k,1}\rho(2) + \phi_{k,2}\rho(1) + \phi_{k,3}\rho(0) + \phi_{k,4}\rho(1) + \ldots + \phi_{k,k}\rho(k-3)$$

$$\vdots$$

For $j = k$:

$$\rho(k) = \phi_{k,1}\rho(k-1) + \phi_{k,2}\rho(k-2) + \phi_{k,3}\rho(k-3) + \phi_{k,4}\rho(k-4) + \ldots + \phi_{k,k}\rho(0)$$

Can be written as the Yule-Walker equations:

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \rho(4) \\ \vdots \\ \rho(k) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \rho(3) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & \rho(2) & \dots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \rho(1) & \dots & \rho(k-3) \\ \rho(3) & \rho(2) & \rho(1) & 1 & \dots & \rho(k-4) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \rho(k-4) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \phi_{k,3} \\ \phi_{k,4} \\ \vdots \\ \phi_{k,k} \end{bmatrix}$$

I.e.

$$\boldsymbol{\rho}_k = \boldsymbol{R}_k \boldsymbol{\phi}_k \,,$$

where $\boldsymbol{\rho}_k, \boldsymbol{\phi}_k \in \mathbb{R}^k$, and $\boldsymbol{R}_k \in \mathbb{R}^{k \times k}$. It can be shown ($\boldsymbol{R}_k$ is semi-positive definite) that $\boldsymbol{R}$ is invertible. Hence:

$$\boldsymbol{\phi}_k = \boldsymbol{R}_k^{-1} \boldsymbol{\rho}_k \,.$$

Note that $\boldsymbol{R}_k$ and $\boldsymbol{\rho}_k$ only contain ACF coefficients. Therefore we can compute the PACF coefficients from the ACF.

$$\boldsymbol{\phi}_k = \boldsymbol{R}_k^{-1} \boldsymbol{\rho}_k \, .$$

Example

$\boldsymbol{\phi}_1 = \phi_{1,1} = \boldsymbol{R}_1^{-1} \boldsymbol{\phi}_1 = 1\rho(1)$. *Hence*

$$\boxed{\phi_{1,1} = \rho(1)}$$

Example

$$\begin{aligned}
\boldsymbol{\phi}_2 = \begin{bmatrix} \phi_{2,1} \\ \phi_{2,2} \end{bmatrix} &= \begin{bmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix} \\
&= \frac{1}{1-\rho(1)^2} \begin{bmatrix} 1 & -\rho(1) \\ -\rho(1) & 1 \end{bmatrix} \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix} \\
&= \frac{1}{1-\rho(1)^2} \begin{bmatrix} \rho(1)(1-\rho(2)) \\ \rho(2)-\rho(1)^2 \end{bmatrix}
\end{aligned}$$

Hence

$$\boxed{\phi_{2,2} = \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2}}$$

$$
\begin{aligned}
\phi_{1,1} &= \rho(1)\,, \\
\phi_{2,2} &= \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}\,.
\end{aligned}
$$

Example

For $AR(1)$, recall $\rho(k) = \phi_1^{|k|}$. Then

$$\phi_{1,1} = \rho(1) = \phi_1\,,$$

and

$$\phi_{2,2} = \frac{\phi_1^2 - \phi_1^2}{1 - \phi_1^2} = 0\,.$$

I.e., 'cut-off' at lag 2.

We can find $\phi_{3,3}$, $\phi_{4,4}$, etc. in a similar way by solving higher order sets of Yule-Walker equations, e.g. $\boldsymbol{\phi}_3 = \boldsymbol{R}_3^{-1}\boldsymbol{\rho}_3$, $\boldsymbol{\phi}_4 = \boldsymbol{R}_4^{-1}\boldsymbol{\rho}_4$, etc. However, (perhaps unsurprisingly?) there is a more efficient way...

Theorem (Durbin-Levinson)

The PACF coefficients can be computed via:

$$\phi_{k,k} = \frac{\rho(k) - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho(k-j)}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho(j)}$$

where $\phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j}$, *for* $j = 1, 2, \ldots, k-1$.

Example

A time series model has $\rho(1) = 2/5$, $\rho(2) = -1/20$, $\rho(3) = -1/8$. *Find PACF at lags* $1, 2, 3$.

For $k = 1$: $\phi_{1,1} = \rho(1) = 2/5$. For $k = 2$:

$$\phi_{2,2} = \frac{\rho(2) - \sum_{j=1}^{1} \phi_{1,j}\rho(2-j)}{1 - \sum_{j=1}^{1} \phi_{1,j}\rho(j)} = \frac{\rho(2) - \phi_{1,1}\rho(1)}{1 - \phi_{1,1}\rho(1)} = \frac{-1/20 - (2/5)^2}{1 - (2.5)^2}$$

I.e. $\phi_{2,2} = -1/4$.

For $k = 3$:

$$\phi_{3,3} = \frac{\rho(3) - \sum_{j=1}^{2} \phi_{2,j}\rho(3-j)}{1 - \sum_{j=1}^{2} \phi_{2,j}\rho(j)} = \frac{\rho(3) - (\phi_{2,1}\rho(2) + \phi_{2,2}\rho(1))}{1 - (\phi_{2,1}\rho(1) + \phi_{2,2}\rho(2))}$$

where $\phi_{2,1} = \phi_{1,1} - \phi_{2,2}\phi_{1,1} = 4/10 - (1/4)4/10 = 1/2\,.$ I.e.

$$\phi_{3,3} = 0\,.$$

In practice, the **sample** PACF coefficients $\hat{\phi}_{k,k}$ can be computed from the **sample** ACF coefficients $\hat{\rho}(k)$, using Yule-Walker equations $\hat{\boldsymbol{\phi}}_k = \hat{}$ $\mathrm{R}_k^{-1}\hat{\boldsymbol{\rho}}_k$, or Durbin Levinson.

Model Identification with the ACF and PACF

Model	ACF	PACF
(1)	exponential decay	spike lag 1, then 0
$AR(p)$	exponential decay or damped sinusoid	spikes lags 1 to p, then 0
$MA(1)$	spike lag 1, then 0	exponential decay
$MA(q)$	spikes lags 1 to q, then 0	exponential decay or damped sinusoid
$ARMA(p,q)$	exponential decay or damped sinusoid (for lags $> q$)	exponential decay or damped sinusoid (for lags $> p$)

Autoregressive Moving Average Models ARMA(p,q)

Definition (Moving average $MA(q)$ process)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0,\sigma^2)$. Then, $\{Y_t\}$ is a moving average process of order q, written $MA(q)$, if

$$Y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q}$$

Definition (Autoregressive $AR(p)$ process)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0,\sigma^2)$. Then, $\{Y_t\}$ is an autoregressive process of order p, written $AR(p)$, if

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t\ ,$$

Definition (Autoregressive, moving average $ARMA(p,q)$ process)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0,\sigma^2)$. Then, $\{Y_t\}$ is an autoregressive, moving average process of order (p,q), written $ARMA(p,q)$, if it contains p-many AR terms and q-many MA terms:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q}$$

An ARMA(p,q) process $\{Y_t\}$ is a stationary process that satisfies

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \epsilon_1 + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

which can be also written in characteristic polynomial form with backshift operator

$$\phi(B) Y_t = \theta(B) \epsilon_t$$

where

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \\ \theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q. \end{aligned}$$

Remark

For an ARMA(p,q) process $\{Y_t\}$ *we always insist that* $\phi, \theta \neq 0$ *and that the polynomials*

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \text{ and } \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

have no common factors. This implies it is not a lower order ARMA model.

ARMA processes can accurately approximate many stationary processes.

Example (MA representation of $ARMA(1,1)$)

A stationary $ARMA(1,1)$ model can written as

$$Y_t = \left(1 + (\phi_1 - \theta_1)\sum_{j=1}^{\infty} \phi_1^{j-1} B^j\right) \epsilon_t$$

To see this, write $ARMA(1,1)$ as $(1-\phi_1 B)Y_t = (1+\theta_1 B)\epsilon_t$, i.e.

$$Y_t = (1-\phi_1 B)^{-1}(1+\theta_1 B)\epsilon_t \,.$$

Assume $|\phi_1| < 1$ and the rest is left as an exercise.

Remark

Note also, that writing the model in terms of $\psi(B)\epsilon_t$ offers an alternative way to compute the ACF.

In e.g. above, write $Y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_t$ with

$$\psi_j = \begin{cases} 1\,, & j = 0 \\ (\phi_1 - \theta_1)\phi_1^{j-1}\,, & \text{oth.} \end{cases}$$

and use (from Wold) $\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$.

Remark

Again, as in purely $AR(p)$ case, $\{Y_t\}$ is stationary iff all roots of AR characteristic equation are strictly greater than one.

Lemma

The $ARMA(p,q)$ process:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

has zero-mean.

Proof

$$Y_t = \sum_{j=0}^{\infty} \phi_j Y_{t-j} + \epsilon_t + \sum_{i=0}^{\infty} \theta_k \epsilon_{t-k} \tag{0.8}$$

Let $\mathbb{E}(Y_t) = \mu$. Take $\mathbb{E}$ of both sides of (0.8):

$$\begin{aligned} \mathbb{E}(Y_t) &= \sum_{j=0}^{\infty} \phi_j \mathbb{E} Y_{t-j} \\ \mu &= \mu \sum_{j=0}^{\infty} \phi_j \end{aligned}$$

$ARMA(p, q)$ and zero-mean

$$\mu = \mu \sum_{j=0}^{\infty} \phi_j$$

Hence, $\sum_{j=1}^{p} \phi_j \neq 1 \Rightarrow \mu = 0$. Now, recall $\{Y_t\}$ stationary $\Rightarrow AR$ characteristic equation has roots outside unit circle, i.e. roots of:

$$1 - \sum_{j=1}^{p} \phi_j x^j = 0 \tag{0.9}$$

must have magnitude greater than one. But, if $\sum_{j=1}^{p} \phi_j = 1$, then (0.9) has a solution at $x = 1$. Therefore, we must have that $\mu = 0$. $\blacksquare$

Example ($ARMA(1,1)$)

Consider the $ARMA(1,1)$ model.

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} \tag{0.10}$$

We can, again, use Yule-Walker-like arguments to derive the ACF.

Multiply both sides of (0.10) by Y_{t-k} and take expectations:

$$\mathbb{E}(Y_t Y_{t-k}) = \phi_1 \mathbb{E}(Y_{t-1} Y_{t-k}) + \mathbb{E}(\epsilon_t Y_{t-k}) + \theta_1 \mathbb{E}(\epsilon_{t-1} Y_{t-k}) \tag{0.11}$$

Now, look at 2nd and 3rd terms on RHS. Shift time index in Equation (0.10) by $-k$ and multiply through by ϵ_t; then $\forall k \geq 0$:

$$\begin{aligned}
\mathbb{E}(\epsilon_t Y_{t-k}) &= \phi_1 \cancelto{0}{\mathbb{E}(\epsilon_t Y_{t-k-1})} + \mathbb{E}(\epsilon_t \epsilon_{t-k}) + \theta_1 \cancelto{0}{\mathbb{E}(\epsilon_t \epsilon_{t-k-1})} = \sigma^2 \delta_{0,k} \, , \\
\mathbb{E}(\epsilon_{t-1} Y_{t-k}) &= \phi_1 \mathbb{E}(\epsilon_{t-1} Y_{t-k-1}) + \mathbb{E}(\epsilon_{t-1} \epsilon_{t-k}) + \theta_1 \mathbb{E}(\epsilon_{t-1} \epsilon_{t-k-1}) \\
&= \phi_1 \sigma^2 \delta_{0,k} + \sigma^2 \delta_{1,k} + \theta_1 \sigma^2 \delta_{0,k} \\
&= \sigma^2 (\phi_1 + \theta_1) \delta_{0,k} + \sigma^2 \delta_{1,k}
\end{aligned}$$

Hence, substituting these into (0.11):

$$\gamma(k) = \phi_1 \gamma(k-1) + \sigma^2 \delta_{0,k} + \theta_1 \sigma^2 \left((\phi_1 + \theta_1) \delta_{0,k} + \delta_{1,k} \right)$$

$$\gamma(k) = \phi_1\gamma(k-1) + \sigma^2\delta_{0,k} + \theta_1\sigma^2\big((\phi_1+\theta_1)\delta_{0,k} + \delta_{1,k}\big)$$

In particular, at $k = 0, 1, \ldots$:

$$\begin{aligned}
\gamma(0) &= \phi_1\gamma(1) + \sigma^2(1+\theta_1\phi_1+\theta_1^2) && (0.12)\\
\gamma(1) &= \phi_1\gamma(0) + \sigma^2\theta_1 && (0.13)\\
\gamma(k) &= \phi_1\gamma(k-1)\,, \quad \forall k \geq 2 && (0.14)
\end{aligned}$$

Substituting (0.13) into (0.12) gives:

$$\begin{aligned}
\gamma(0) &= \phi_1^2\gamma(0) + \sigma^2\theta_1\phi_1 + \sigma^2(1+\theta_1\phi_1+\theta_1^2)\\
\Rightarrow \gamma(0) &= \sigma^2\frac{1+2\theta_1\phi_1+\phi_1^2}{1-\phi_1^2}\\
\Rightarrow \gamma(1) &= \phi_1\sigma^2\frac{1+2\theta_1\phi_1+\phi_1^2}{1-\phi_1^2} + \sigma^2\theta_1\,, \quad \text{[sub. } \gamma(0) \text{ into (0.13)]}
\end{aligned}$$

which, together with (0.14), implies

$$\gamma(k) = \phi_1^k\sigma^2\frac{1+2\theta_1\phi_1+\phi_1^2}{1-\phi_1^2} + \phi_1^{k-1}\sigma^2\theta_1\,, \quad k \geq 1$$

Now, divide

$$\gamma(k) = \phi_1^k \sigma^2 \frac{1 + 2\theta_1\phi_1 + \phi_1^2}{1 - \phi_1^2} + \phi_1^{k-1}\sigma^2\theta_1 \,, \quad k \geq 1$$

by $\gamma(0)$ to get:

$$\rho(k) = \phi_1^k + \phi_1^{k-1} \frac{\theta_1(1 - \phi_1^2)}{1 + 2\theta_1\phi_1 + \phi_1^2} \,, \quad |\phi_1| < 1$$

i.e., for the ARMA(1,1) model, $\rho(k) = A\phi^{k-1}$ behaves similar to $AR(1)$ (decays exponentially).

Furthermore, for any stationary process with ACVF $\gamma(\cdot)$, any any $m > 0$, there exists an ARMA process $\{Y_t\}$ for which

$$\gamma_Y(k) = \gamma(k), \quad k = 0, 1, 2, \ldots, m$$

Recall the definition of causal and invertible. Let $\{Y_t\}$ be an ARMA(p,q) process defined by equation $\phi_p(B)Y_t = \theta_q(B)\epsilon_t$, then

- $\{Y_t\}$ is said to be causal if there exists constants $\{\psi_j\}$ s.t. $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$Y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots$$

- $\{Y_t\}$ is said to be invertible if there exists constants $\{\pi_j\}$ s.t. $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and

$$\epsilon_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots$$

- Neith causality nor invertibility is a property of $\{Y_t\}$ alone, but of the relationship between $\{Y_t\}$ and $\{\epsilon_t\}$.

$ARMA(p,q)$ Model ACF

Since the causal ARMA(p,q) process $\phi(B)Y_t = \theta(B)\epsilon_t$ has representation

$$Y_t = \Psi(B)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j},$$

where

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z), \quad |z| \leq 1.$$

The ACVF of $\{Y_t\}$ is then

$$\gamma_Y(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}.$$

To determine the coefficients ψ_j one can use the method of matching coefficients

$$\left(1+\psi_1 z+\psi_2 z^2+\psi_3 z^3+\psi_4 z^4+\cdots\right)\left(1-\phi_1 z-\phi_2 z^2-\cdots-\phi_p z^p\right)=\left(1+\theta_1 z+\theta_2 z^2+\cdots+\theta_q z^q\right)$$

which gives difference equations for ψ_k:

$$\begin{aligned}
\psi_1 - \phi_1 &= \theta_1 \\
\psi_2 - \phi_2 - \psi_1\phi_1 &= \theta_2 \\
\psi_3 - \phi_3 - \psi_2\phi_1 - \psi_1\phi_2 &= \theta_3 \\
&\cdots
\end{aligned}$$

$ARMA(p,q)$ Model ACF

By definining $\theta_0 = 1$, $\theta_j = 0$ for $j > q$ and $\phi_j = 0$ for $j > p$, one obtains a summary

$$\psi_j - \sum_{0<k\leq j} \phi_k \psi_{j-k} = \theta_j, \;\; 0 \leq j < \max\{p, q+1\} \tag{0.15}$$

$$\psi_j - \sum_{0<k\leq p} \phi_k \psi_{j-k} = 0, \;\; j \geq \max\{p, q+1\} \tag{0.16}$$

from which a general solution to the second component of the system (Eqn 0.16) is given by

$$\psi_n = \sum_{i=1}^{k} \sum_{j=0}^{r_i - 1} \alpha_{ij} n^j \zeta_i^{-n}, \;\; n \geq \max(p, q+1) - p$$

where ζ_i, $i = 1, \ldots, k$ are the distinct zeros of $\phi(z)$ and r_i is the multiplicity of ζ_i. The p constants α_{ij}s and the coefficients ψ_j, $0 \leq j < \max(p, q+1) - p$, are then determined uniquely by the $\max(p, q+1)$ boundary conditions in Eq 0.15

Theorem

Let $\{Y_t\}$ *be an ARMA(p,q) process. Then* $\{Y_t\}$ *is causal if and only if*

$$\phi(z) \neq 0 \quad \forall |z| \leq 1$$

The coefficients $\{\psi_j\}$ *in any causal expansion are determined by the expression*

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1.$$

Proof: First, we assume that $\phi(z) \neq 0$ if $|z| \leq 1$. Since we have

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = \phi_p(z - z_1)\cdots(z - z_p),$$

then $|z_i| > 1$ for $i = 1, \ldots, p$. For each i,

$$\frac{1}{z - z_i} = -\frac{1}{z_i(1 - z/z_i)} = -\frac{1}{z_i}\sum_{k=0}^{\infty}(z/z_i)^k, \quad \mathit{for} \quad |z| < |z_i|.$$

This implies that there exists $\delta > 0$ s.t. $1/\phi(z)$ has a power series expansion

$$\frac{1}{\phi(z)} = \sum_{j=0}^{\infty}\zeta_j z^j =: \zeta(z), \quad |z| < 1 + \delta \leq \min_i |z_i|.$$

Consequently, $\zeta_j(1+\delta/2)^j \to 0$ as $j \to \infty$ so that there exists $K > 0$ s.t.

$$|\zeta_j| < K(1+\delta/2)^{-j}, \quad \forall j = 0, 1, 2, \ldots$$

In particular we have $\sum_{j=0}^{\infty}|\zeta_j| < \infty$ and $\zeta(z)\phi(z) = 1$ for $|z| \leq 1$.

$ARMA(p,q)$ Causality

Therefore, we can apply $\zeta(B)$ to both sides of the equation $\phi(B)Y_t = \theta(B)\epsilon_t$ to obtain

$$Y_t = \zeta(B)\theta(B)\epsilon_t.$$

Thus we have the desired representation

$$Y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

where the sequence $\{\psi_j\}$ is determined by $\theta(z)/\phi(z)$.
Now assume that $\{Y_t\}$ is causal, i.e. $Y_t = \Psi(B)\epsilon_t$ with $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Then

$$\zeta(B)\phi(B)Y_t = \zeta(B)\theta(B)\epsilon_t.$$

If we let $\eta(z) = \phi(z)\Psi(z) = \sum_{j=0}^{\infty} \eta_j z^j, |z| \leq 1$, we can rewrite this equation as

$$\sum_{j=0}^{q} \theta_j \epsilon_{t-j} = \sum_{j=0}^{\infty} \eta_j \epsilon_{t-j},$$

and taking inner products of each side with W_{t-k} one obtains $\eta_k = \theta_k$, $k = 0, \ldots, q$ and $\eta_k = 0$, $k > q$.

Hence,

$$\theta(z) = \eta(z) = \phi(z)\Psi(z), \quad |z| \leq 1.$$

Since, $\theta(z)$ and $\phi(z)$ have no common zeros and since $|\Psi(z)| \leq \infty$ for $|z| \leq 1$, we conclude that $\phi(z)$ cannot be zero for $|z| \leq 1$. $\square$

Remark

If $\phi(z) = 0$ for some $|z| = 1$, there there is no stationary solution of the ARMA equations

$$\phi(B)Y_t = \theta(B)\epsilon_t.$$

$ARMA(p,q)$ Invertibility

Theorem

Let $\{Y_t\}$ be an ARMA(p,q) process. Then $\{Y_t\}$ is invertible if and only if

$$\theta(z) \neq 0 \;\; \forall |z| \leq 1$$

The coefficients $\{\pi_j\}$ in any inverted expansion are determined by the relation

$$\Pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1.$$

$ARMA(p, q)$ Invertibility

Proof follows analogously the logic of the causal case, except for different polynomial.

Proof: First assume that $\theta(z) \neq 0$ if $|z| \leq 1$. Hence, one can say that $1/\theta(z)$ has a power series expansion (see argument in causality proof previously)

$$1/\theta(z) = \sum_{j=0}^{\infty} \eta_j z^j =: \eta(z), \quad |z| < 1 + \delta,$$

for some $\delta > 0$ and $\sum_{j=0}^{\infty} |\eta_j| < \infty$. They apply $\eta(B)$ to both sides of ARMA equations, to get

$$\eta(B)\phi(B)Y_t = \eta(B)\theta(B)\epsilon_t = \epsilon_t.$$

Thus, we have the desired representation

$$\epsilon_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}$$

where the sequence $\{\pi_j\}$ is determined by $\phi(z)/\theta(z)$.

$ARMA(p,q)$ Invertibility

Conversely, if $\{Y_t\}$ is invertible then $\epsilon_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j} = \pi(B) Y_t$ for some $\sum_{j=0}^{\infty} |\pi_j| < \infty$, then

$$\phi(B)\epsilon_t = \phi(B)\pi(B)Y_t = \pi(B)\phi(B)Y_t = \pi(B)\theta(B)\epsilon_t$$

which leads to

$$\sum_{j=0}^{q} \phi_j \epsilon_{t-j} = \sum_{j=0}^{\infty} \zeta_j \epsilon_{t-j}$$

where $\zeta(z) = \pi(z)\theta(z) = \sum_{j=0}^{\infty} \zeta_j z^j$, $|z| \leq 1$. Taking inner products of each side with ϵ_{t-k} one obtains $\zeta_k = \phi_k$, $k = 0, 1, \ldots, q$ and $\zeta_k = 0$, $k \geq q$. Hence

$$\phi(z) = \zeta(z) = \pi(z)\theta(z), \quad |z| \leq 1.$$

Since $\theta(z)$ and $\phi(z)$ have no common zeros and since $|\pi(z)| < \infty$ for $|z| \leq 1$, we conclude that $\theta(z)$ cannot be zero for $|z| \leq 1$. $\square$

Remark

Uniqueness: *If* $\phi(z) \neq 0$ *for all* $|z| = 1$*, then the ARMA equations* $\phi(B)Y_t = \theta(B)\epsilon_t$ *have the* **unique stationary solution**

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$$

where ψ_j *comes from* $\theta(z)/\phi(z)$.

Example

Consider an ARMA(2,1) process given by

$$Y_t - Y_{t-1} + 0.25 Y_{t-2} = \epsilon_t + \epsilon_{t-1}.$$

This gives polynomials

$$\begin{aligned} \phi(z) &= 1 - z - (-0.25) z^2 \\ \theta(z) &= 1 + z \end{aligned}$$

The root of $\phi(z)$ *is 2* $(|2| > 1)$ *with multiplicity 2 and the root of* $\theta(z)$ *is* -1 $(|-1| = 1)$ *with multiplicity 1* $\Rightarrow$ $\{Y_t\}$ *is causal but not invertible.*

Theorem

Let Y_t satisfy an $ARMA(p,q)$ model, i.e. $\phi(B)Y_t = \theta(B)\epsilon_t$.

- *If the roots of the $AR(p)$ characteristic polynomial $\phi(B)$ are outside unit circle then Y_t is stationary; and*

$$\exists\, \psi(B) = \frac{\theta(B)}{\phi(B)}, \quad s.t.\ \ Y_t = \psi(B)\epsilon_t$$

- *If the roots of the $MA(q)$ characteristic polynomial $\theta(B)$ are outside unit circle then Y_t is invertible; and*

$$\exists\, \pi(B) = \frac{\phi(B)}{\theta(B)}, \quad s.t.\ \ \epsilon_t = \pi(B)Y_t$$

Integration and ARIMA(p,d,q) Family

Example

Consider the process

$$Y_t = \alpha_0 + \alpha_1 t\,, \quad \alpha_0, \alpha_1 \in \mathbb{R}\,.$$

In this case Y_t *is not stationary, since (the 'mean function')*

$$\mathbb{E} Y_t = \alpha_0 + \alpha_1 t\,,$$

which depends on t.

But, we can 'transform' the original non-stationary process to a stationary one by differencing. Consider:

$$\begin{aligned} Z_t &:= Y_t - Y_{t-1} \\ &= \cancel{\alpha_0} + \cancel{\alpha_1 t} - (\cancel{\alpha_0} + \alpha_1(\cancel{t} - 1)) \\ &= \alpha_1\,, \end{aligned}$$

which is now stationary.

Example

Consider an ARMA model where one (and only one) of the AR roots is equal to 1 and all the other roots are outside the unit circle. I.e. AR polynomial takes the form $\phi^*(B) := \phi(B)(1-B)$, where all the roots of $\phi(B)$ are outside the unit circle. We have:

$$\phi^*(B)Y_t = \phi(B)(1-B)Y_t = \theta(B)\epsilon_t$$

Again, define $Z_t := Y_t - Y_{t-1}$. I.e. $Z_t = (1-B)Y_t$, and we have

$$\phi(B)Z_t = \theta(B)\epsilon_t \,.$$

and Z_t is stationary.

Lemma

Y_t *stationary* $\Rightarrow$ $Y_t - Y_{t-1}$ *stationary.*

Proof $Y_t = \dfrac{\theta(B)}{\phi(B)}\epsilon_t$. Put $Z_t := Y_t - Y_{t-1}$:

$$\begin{aligned} Z_t &= (1-B)\frac{\theta(B)}{\phi(B)}\epsilon_t \\ \phi(B)Z_t &= (1-B)\theta(B)\epsilon_t \end{aligned}$$

Definition (Intrinsically Stationary Processes)

Stochastic process $\{Y_t\}$ is said to be intrinsically stationary of integer order $d > 0$ if $\{Y_t\}$,$\{\nabla Y_t\}$,...,$\{\nabla^{d-1} Y_t\}$ are non-stationary, but $\{\nabla^d Y_t\}$ is a stationary process.

Note: any stationary process is intrinsically stationary of order $d = 0$.

Definition (ARIMA(p,d,q) Process)

If d is a non-negative integer, then $\{Y_t\}$ is said to be an ARIMA(p,d,q) process if

- *$\{Y_t\}$ is intrinsically stationary of order d and*
- *$\{\nabla^d Y_t\}$ is a causal ARMA(p,q) process.*

With $\{\epsilon_t\} \sim WN\left(0, \sigma^2\right)$ we can express the model as

$$\phi(B)(1-B)^d Y_t = \theta(B)\epsilon_t$$

The simplest example of an ARIMA process is the ARIMA(0,1,0):

$$(1-B)Y_t = Y_t - Y_{t-1} = \epsilon_t$$

for which assuming existance of X_0 and $t \geq 1$

$$\begin{aligned}
Y_1 &= Y_0 + \epsilon_1 \\
Y_2 &= Y_1 + \epsilon_2 = Y_0 + \epsilon_1 + \epsilon_2 \\
Y_3 &= Y_2 + \epsilon_3 = Y_0 + \epsilon_1 + \epsilon_2 + \epsilon_3 \\
&\vdots \\
Y_t &= Y_0 + \sum_{k=1}^{t} \epsilon_t
\end{aligned}$$

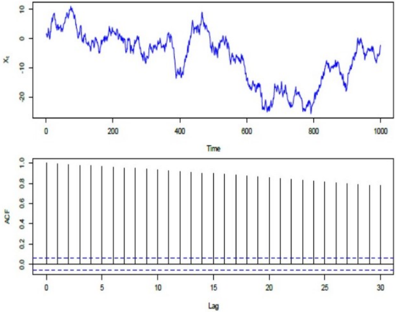
This is a random walk staring from Y_0. Assuming Y_0 is uncorrelated with ϵ_t's one has...

$$
\begin{aligned}
Var\left(Y_t\right) &= Var\left(Y_0 + \sum_{k=1}^{t} \epsilon_k\right) \\
&= Var\left(Y_0\right) + Var\left(\sum_{k=1}^{t} \epsilon_k\right) \\
&= Var\left(Y_0\right) + \sum_{k=1}^{t} Var\left(\epsilon_k\right) \\
&= Var\left(Y_0\right) + t\sigma^2
\end{aligned}
$$

which is either time-dependent or infinite if $Var(Y_0) = \infty$. Further, one can show

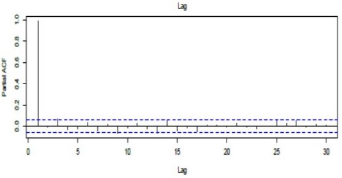
$$
\begin{aligned}
Cov\left(Y_{t+k}, Y_t\right) &= Cov\left(Y_0 + \sum_{j=1}^{t+k} \epsilon_j, Y_0 + \sum_{j=1}^{t} \epsilon_j\right) \\
&= Var\left(Y_0\right) + \min(t, t+l)\sigma^2
\end{aligned}
$$

Thus ARIMA(0,1,0) is clearly a non-stationary process. The same is true for all ARIMA(p,d,q) process when d is a positive integer.



A realizaiton of random walk.

```
N=1000; Wt=rnorm(N,0,1); Xt=cumsum(Wt);
par(mfrow=c(3,1));par(mar=c(4.5,4.5,.1,.1));
plot.ts(Xt,col="blue",ylab=expression(X[t]));
acf(Xt,type="correlation");acf(Xt, type="partial")
```



Example

Consider $AR(1)$ model

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t \,, \quad \epsilon \sim \mathcal{WN}(0, \sigma^2) \,.$$

Stationary if $|\phi_1| < 1$. But, if $\phi_1 = 1$:

$$Y_t = Y_{t-1} + \epsilon_t \,, \quad \textit{[random walk]}$$

which is non-stationary (exercise: prove using successive substitutions and using $Y_0 = 0$)

However, note the difference

$$Z_t := \nabla Y_t = \epsilon_t \sim \mathcal{WN}(0, \sigma^2) \,,$$

is stationary.

Example

If an $ARMA$ model has d-many AR roots $= 1$, then

$$\phi(B)(1 - B)^d Y_t = \theta(B)\epsilon_t \,.$$

Now define $Z_t := (1 - B)^d Y_t = \nabla^d Y_t$ (i.e. take dth difference of Y_t). Then Z_t is stationary.

Remark

- *In practice, 1st differencing is often found to be adequate to make a series stationary.*
- *Sometimes, 2nd differencing is required.*
- *3rd or higher differencing is not usually required.*

CAVEAT! Over-differencing will cause:

- an increase in variance
- an increase in the order of MA - any ACF cut-off lag will increase
- Over-differencing will cause non-invertibility! (MA poly. has unit root).

Example

Consider (again) random walk $Y_t = Y_{t-1} + \epsilon_t$*. Then first difference* $\nabla Y_t = \epsilon_t$ *is (stationary and) invertible. The second difference* $\nabla^2 Y_t = \nabla \epsilon_t = \epsilon_t - \epsilon_{t-1}$ *(is still stationary — c.f. Lemma 76) but is now not invertible — it is* $MA(1)$ *with* $\theta_1 = 1$.

We can now extend our interest from stationary *ARMA* models to any non-stationary model that can be 'transformed' into a stationary *ARMA* model by differencing.

Definition (Integrated autoregressive, moving average $ARIMA(p, d, q)$ process)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$ *and let the process* $\{(1 - B)^d Y_t\}$ *be an* $ARMA(p, q)$ *process of order* (p, q). *Then,* $\{Y_t\}$ *is an* *integrated autoregressive, moving average process* *of order* (p, d, q), *written* $ARIMA(p, d, q)$, *with model equation:*

$$\phi(B)(1 - B)^d Y_t = \theta(B)\epsilon_t \, .$$

Remark

If $\{Y_t\}$ *has to be differenced* d*-many times before it is a stationary* (p, q) *process, then* $\{Y_t\}$ *is an* $ARIMA(p, d, q)$ *process.*

Example ($ARIMA(0, 1, 0)$ or $I(1)$)

Recall random walk: $\nabla Y_t = \epsilon_t$, *with* $\epsilon_t \sim \mathcal{WN}(0, \sigma^2)$.

Example ($ARIMA(0,1,1)$ or $IMA(1,1)$)

Find autocorrelation of Y_t, *where* $(1-B)Y_t = (1+\theta_1 B)\epsilon_t$. *I.e.*

$$Y_t = Y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}\,.$$

Assuming $Y_0 = 0$,

$$\begin{aligned}
Y_1 &= \epsilon_1 + \theta_1\,\epsilon_0 \\
Y_2 &= Y_1 + \epsilon_2 + \theta_1\,\epsilon_1 = \epsilon_2 + \theta_1\,\epsilon_0 + (1+\theta_1)\,\epsilon_1 \\
Y_3 &= Y_2 + \epsilon_3 + \theta_1\,\epsilon_2 = \epsilon_3 + \theta_1\,\epsilon_0 + (1+\theta_1)\,\epsilon_2 + (1+\theta_1)\,\epsilon_1 \\
&\vdots \\
Y_t &= \epsilon_t + \theta_1\epsilon_0 + (1+\theta_1)\sum_{j=1}^{t-1}\epsilon_j
\end{aligned}$$

Now $\mathbb{E}(Y_t) = 0$. Hence $\text{cov}(Y_t, Y_{t+k}) = \mathbb{E}(Y_t Y_{t+k}) =$

$$\mathbb{E}\left(\left(\epsilon_t + \theta_1\epsilon_0 + (1+\theta_1)\sum_{j=1}^{t-1}\epsilon_j\right)\left(\epsilon_{t+k} + \theta_1\epsilon_0 + (1+\theta_1)\sum_{j=1}^{t+k-1}\epsilon_j\right)\right)$$

For $k \geq 0$:

$$\mathbb{E}\left(\left(\epsilon_t + \theta_1\epsilon_0 + (1+\theta_1)\ \sum_{j=1}^{t-1}\epsilon_j\right)\left(\epsilon_{t+k} + \theta_1\epsilon_0 + (1+\theta_1)\ \sum_{j=1}^{t+k-1}\epsilon_j\right)\right)$$

$$= \mathbb{E}(\epsilon_t\epsilon_{t+k}) + \theta_1^2\mathbb{E}(\epsilon_0^2) + (1+\theta_1)^2\sum_{j=1}^{t-1}\sum_{\ell=1}^{t+k-1}\mathbb{E}(\epsilon_j\epsilon_\ell) + (1+\theta_1)\mathbb{E}\left(\epsilon_t\sum_{\ell=1}^{t+k-1}\epsilon_\ell\right)$$

$$= \sigma^2\left(\delta_{0,k} + \theta_1^2 + (1+\theta_1)^2\sum_{j=1}^{t-1}1\right) + (1+\theta_1)\mathbb{E}\left(\epsilon_t\sum_{\ell=1}^{t+k-1}\epsilon_\ell\right)$$

$$= \sigma^2(\delta_{0,k} + \theta_1^2 + (1+\theta_1)^2(t-1)) + (1+\theta_1)\mathbb{E}\left(\epsilon_t\sum_{\ell=1}^{t+k-1}\epsilon_\ell\right)$$

$$= \begin{cases}\sigma^2\left(1+\theta_1^2+(1+\theta_1)^2(t-1)\right), & k=0\\ \sigma^2\left(\theta_1^2+(1+\theta_1)^2(t-1)+1+\theta_1\right), & k>0\end{cases}$$

I.e., for $k > 0$:

$$\mathrm{cov}(Y_t, Y_{t+k}) = \mathrm{var}(Y_t) + \theta_1\sigma^2\,.$$

$\mathrm{cov}(Y_t, Y_{t+k}) = \mathrm{var}(Y_t) + \theta_1\sigma^2$. Now,

$$\mathrm{corr}(Y_t, Y_{t+k}) = \frac{\mathrm{cov}(Y_t, Y_{t+k})}{\sqrt{\mathrm{var}(Y_t)\,\mathrm{var}(Y_{t+k})}} = \frac{\mathrm{var}(Y_t) + \theta_1\sigma^2}{\sqrt{\mathrm{var}(Y_t)\,\mathrm{var}(Y_{t+k})}}\,.$$

Remark

Note

- $\mathrm{var}(Y_t)$ *is a function of* $t \Rightarrow \{Y_t\}$ *is non-stationary.*
- *for* t *large and* k *small,* $\mathrm{corr}(Y_t, Y_{t+k}) \approx 1$*. (ACF very slow decay)*

Example ($ARIMA(1,1,0)$ or $ARI(1,1)$)

Find autocorrelation of Y_t*, where* $(1-\phi_1 B)(1-B)Y_t = \epsilon_t$*. More straightforward than previous example(!) (homework exercise).*

Non-zero Mean

To allow $\{Y_t\}$ to have (possibly) non-zero mean...

Definition ($ARMA(p,q)$ process with constant)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. Then, $\{Y_t\}$, with $\mu := \mathbb{E}(Y_t)$, is an autoregressive, moving average process of order (p, q), written $ARMA(p, q)$, if it contains p-many AR terms and q-many MA terms:

$$Y_t - \mu = \sum_{j=1}^{p} \phi_j (Y_{t-j} - \mu) + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \, .$$

Remark

I.e., for $\mathbb{E}(Y_t) = \mu$, the $ARMA$ model can be written

$$\phi(B)(Y_t - \mu) = \theta(B)\epsilon_t \, ,$$

and the $ARIMA$ model is

$$\phi(B)(1 - B)^d (Y_t - \mu) = \theta(B)\epsilon_t \, ,$$

Proposition

Let $\{X_t\}$ *be a stationary process. Let*

$$Y_t := \mu(t; d) + X_t \,,$$

where

$$\mu(t; d) := \sum_{j=0}^{d} m_j t^j \,, \quad m_j \in \mathbb{R} \,.$$

(i.e. the mean of Y_t *is a deterministic polynomial of degree* d*).*
Then,

$$\nabla^d Y_t = d! m_d + \nabla^d X_t \,,$$

and $\mathbb{E}(\nabla^d Y_t) = d! m_d$ *(a constant).*

Proof Homework exercise.

Corollary

Let the d*-differenced process* $\{Z_t := \nabla^d Y_t\}$ *be a stationary* (p, q) *process with non-zero, constant, mean. Then,* $\{Y_t\}$ *contains a polynomial trend term of degree* d*.*

Seasonal Autoregressive Moving Average models (SARIMA)

Often, time series contain seasonal (or periodic) components. For example, consider the following 'economic' data

Quarterly 'economic' data

Quarter	1	2	3	4
Year 1	y_1	y_2	y_3	y_4
Year 2	y_5	y_6	y_7	y_8
Year 3	y_9	y_{10}	y_{11}	y_{12}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

It might be reasonable to assume that correlations exist between quarters 1 and 5; 2 and 6; and so on. Namely

$$Y_t = \phi_1 Y_{t-4} + \epsilon_t \,, \quad \epsilon_t \sim \mathcal{WN}(0, \sigma^2) \,.$$

This is a seasonal *AR* model.

If the data had been sampled at monthly intervals, we **might** expect:

$$Y_t = \phi_1 Y_{t-12} + \epsilon_t \,, \quad \epsilon_t \sim \mathcal{WN}(0, \sigma^2) \,.$$

Likewise, the noise terms could contain perdiodic components.

Definition (seasonal moving average $SMA(1)_s$ model)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. Then $\{Y_t\}$ is a seasonal moving average process of order 1, with period s, written as $SMA(1)_s$, if

$$Y_t = \epsilon_t + \Theta_1 \epsilon_{t-s}, \quad s \geq 2$$

Note $\mathbb{E}(Y_t) = 0$. Multiply both sides by Y_{t-k}, take $\mathbb{E}$:

$$\begin{aligned}
\gamma(k) &= \mathbb{E}(Y_t Y_{t-k}) = \mathbb{E}\big((\epsilon_t + \Theta_1 \epsilon_{t-s})(\epsilon_{t-k} + \Theta_1 \epsilon_{t-k-s})\big) \\
&= \mathbb{E}(\epsilon_t \epsilon_{t-k}) + \phi_1^2 \mathbb{E}(\epsilon_{t-s}\epsilon_{t-k-s}) + \Theta_1 \mathbb{E}(\epsilon_t \epsilon_{t-k-s}) + \Theta_1 \mathbb{E}(\epsilon_{t-s}\epsilon_{t-k}) \\
&= \sigma^2 \delta_{0,k} + \sigma^2 \Theta_1^2 \delta_{0,k} + \sigma^2 \Theta_1 \delta_{-s,k} + \sigma^2 \Theta_1 \delta_{s,k} \\
&= \sigma^2 \left(\delta_{0,k}(1 + \Theta_1^2) + \Theta_1 \delta_{s,|k|} \right) \\
&= \begin{cases} \sigma^2 (1 + \Theta_1^2), & k = 0 \\ \sigma^2 \Theta_1, & |k| = s \\ 0, & \text{oth.} \end{cases}
\end{aligned}$$

$$\gamma(k) = \begin{cases} \sigma^2(1+\Theta_1^2)\,, & k = 0 \\ \sigma^2\Theta_1\,, & |k| = s \\ 0\,, & \text{oth.} \end{cases}$$

And, $\rho(k) = \gamma(k)/\gamma(0) \Rightarrow$

$$\rho(k) = \begin{cases} 1\,, & k = 0 \\ \dfrac{\Theta_1}{1+\Theta_1^2}\,, & |k| = s \\ 0\,, & \text{oth.} \end{cases}$$

Remark

The $SMA(1)_s$ process $Y_t = \epsilon_t + \Theta_1\epsilon_{t-s}$ can be written, in backshift notation, as $Y_t = (1+\Theta_1 B^s)\epsilon_t$.

Definition (seasonal moving average $SMA(Q)_s$ model)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0,\sigma^2)$. Then $\{Y_t\}$ is a seasonal moving average process of order Q, with period s, written as $SMA(Q)_s$, if

$$Y_t = \epsilon_t + \Theta_1\epsilon_{t-s} + \Theta_2\epsilon_{t-2s} + \ldots + \Theta_Q\epsilon_{t-Qs}\,, \quad s \geq 2$$

Definition

Denote by $\Theta(B)_{Q,s}$ *the* *seasonal MA characteristic polynomial. Then the* $SMA(Q)_s$ *process* $Y_t = \epsilon_t + \sum_{j=1}^{Q} \Theta_j \epsilon_{t-js}$ *can be written, in backshift notation, as* $Y_t = (1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_Q B^{Qs})\epsilon_t$ *. I.e.*

$$Y_t = \left(1 + \sum_{j=1}^{Q} \Theta_j B^{js}\right) \epsilon_t =: \Theta(B)_{Q,s} \epsilon_t$$

Remark

A $SMA(Q)_s$ *process is always stationary. It is invertible iff roots of* $\Theta(B)_{Q,s}$ *lie outside unit circle. (c.f.* $MA(q)$ *process.)*

Remark

$$\rho(ks) = \begin{cases} \frac{+\Theta_k + \sum_{j=1}^{Q-k} \Theta_j \Theta_{j+k}}{1 + \sum_{j=1}^{Q} \Theta_j^2}, & k = 1, 2, \ldots, Q \\ 0, & oth. \end{cases}$$

I.e. ACF is only non-zero at $k = 0, s, 2s, \ldots, Qs$.

Definition (seasonal autoregressive $SAR(1)_s$ model)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. Then $\{Y_t\}$ is a seasonal autoregressive process of order 1, with period s, written as $SAR(1)_s$, if

$$Y_t = \Phi_1 Y_{t-s} + \epsilon_t$$

Which can be written, in backshift notation, as $(1 - \Phi_1 B^s) Y_t = \epsilon_t$.

Note $\mathbb{E}(Y_t) = 0$. For $k \geq 1$, multiply both sides by Y_{t-k}, take $\mathbb{E}$:

$$\begin{aligned}\gamma(k) &= \mathbb{E}(Y_t Y_{t-k}) = \mathbb{E}(Y_{t-s} Y_{t-k}) + \cancel{\mathbb{E}(\epsilon_t Y_{t-k})} \rightarrow 0,\ k \geq 1 \\ &= \Phi_1 \gamma(k - s)\,,\end{aligned}$$

Now, divide both sides by $\gamma(0)$:

$$\rho(k) = \Phi_1 \rho(k - s), \quad k \geq 1\,.$$

$$\begin{aligned}\text{At } k = s: \quad \rho(s) &= \Phi_1 \rho(0) = \Phi_1 \\ \text{At } k = 2s: \quad \rho(2s) &= \Phi_1 \rho(s) = \Phi_1^2 \\ \vdots \qquad & \vdots \\ \text{At } k = \ell s: \quad \rho(\ell s) &= \Phi_1 \rho((\ell - 1)s) = \Phi_1^\ell\,,\ \ell \in \mathbb{N}\end{aligned}$$

Recall

$$\rho(k) = \Phi_1 \rho(k-s), \quad k \geq 1\,.$$

At $k = s - m$, where $1 \leq m \leq s - 1$:

$$\rho(s-m) = \Phi_1 \rho(-m) = \Phi_1 \rho(m)\,. \tag{0.17}$$

But at $k = m$:

$$\rho(m) = \Phi_1 \rho(m-s) = \Phi_1 \rho(s-m) = \Phi_1^2 \rho(m), \qquad \text{[from (0.17)]}$$

But, $\{Y_t\}$ stationary $\Rightarrow |\Phi_1| < 1$, i.e. $\Phi_1 \neq 1 \Rightarrow \rho(m) = 0$ for $m = 1, \ldots, s-1$.

Similarly, $\rho(2s - m) = \Phi_1 \rho(s-m) = 0$ [from (0.17)]. Likewise, in general: $\rho(\ell s - m) = 0 \, \forall \ell \in \mathbb{Z}$, and

$$\rho(ks) = \begin{cases} \Phi_1^k\,, & k = 0, \pm s, \pm 2s, \pm 3s, \ldots \\ 0\,, & \text{oth.} \end{cases}$$

I.e., the non-zeros at lags $k = 0, \pm s, \pm 2s, \ldots$ decay exponentially.

Definition (seasonal autoregressive $SAR(P)_s$ model)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. Then $\{Y_t\}$ is a seasonal autoregressive process of order P, with period s, written as $SAR(P)_s$, if

$$Y_t = \Phi_1 Y_{t-s} + \Phi_2 Y_{t-2} + \ldots + \Phi_P Y_{t-P} + \epsilon_t$$

Denote by $\Phi(B)_{P,s}$ the seasonal AR characteristic polynomial and then the $SAR(P)_s$ process $Y_t = \epsilon_t + \sum_{j=1}^{P} \Phi_j Y_{t-js}$ can be written, in backshift notation, as

$$\Phi(B)_{P,s} Y_t := \left(1 - \sum_{j=0}^{P} \Phi_j B^{js}\right) Y_t = \epsilon_t$$

Remark

A $SAR(P)_s$ process is stationary iff all roots of $\Phi(B)_{P,s}$ lie outside unit circle. (c.f. (p) process.)

Generally, processes can have short-term correlations (like *ARMA* processes) as well as seasonal correlations (like *SMA* and *SAR*).

Definition (multiplicative seasonal *ARMA* models (SARMA))

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. Then $\{Y_t\}$ is a multiplicative seasonal autoregressive, moving average process of order $(p, q) \times (P, Q)_s$, with period s, written as $SARMA(p, q) \times (P, Q)_s$, if

$$\phi(B)\,\Phi(B)_{P,s}\,Y_t = \theta(B)\,\Theta(B)_{Q,s}\,\epsilon_t$$

where

$$\begin{aligned}
\phi(x) &= 1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p \\
\Phi(x)_{P,s} &= 1 - \Phi_1 x^s - \Phi_2 x^{2s} - \ldots - \Phi_P x^{Ps} \\
\theta(x) &= 1 + \theta_1 x + \theta_2 x^2 + \ldots + \theta_q x^q \\
\Theta(x)_{Q,s} &= 1 + \Theta_1 x^s + \Theta_2 x^{2s} + \ldots + \Theta_Q x^{Qs}
\end{aligned}$$

Definition

Let $\{Y_t\}$ *be some process. Then (recall that) the difference operator* ∇ *is defined by*

$$\nabla Y_t = Y_t - Y_{t-1} \, .$$

The d*th difference operator* ∇^d *is defined as*

$$\nabla^d Y_t = (1 - B)^d Y_t \, .$$

Definition

Let $\{Y_t\}$ *be some process. Then the seasonal difference operator* ∇_s *is defined by*

$$\nabla_s Y_t = Y_t - Y_{t-s} \, .$$

The D*th seasonal difference operator* ∇_s^D *is defined as*

$$\nabla_s^D Y_t = (1 - B^s)^D \, .$$

Remark

Note $\nabla_s^2 = \nabla_s \nabla_s$, $\nabla_s^3 = \nabla_s \nabla_s \nabla_s$, *etc.*

Definition (Integrated, seasonal, autoregressive, moving average process: *SARIMA*)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$ and let the process $\{\nabla^d \nabla_s^D Y_t\}$ be a $SARMA(p,q) \times (P,Q)_s$ process of order $(p,q) \times (P,Q)_s$ with period s. Then, $\{Y_t\}$ is an integrated, seasonal, autoregressive, moving average process of order $(p,d,q) \times (P,D,Q)_s$, with period s, written $SARIMA(p,d,q) \times (P,D,Q)_s$, with model equation:

$$\phi(B)\,\Phi(B)\,\nabla^d\,\nabla_s^D\,Y_t = \theta(B)\,\Theta(B)\,\epsilon_t$$

Remark

If $\{Y_t\}$ has to be (non-seasonally) differenced d-many times and seasonally differenced D-many times before it is a stationary $SARMA(p,q) \times (P,Q)_s$ process, then $\{Y_t\}$ is a $SARIMA(p,d,q) \times (P,D,Q)_s$ process.

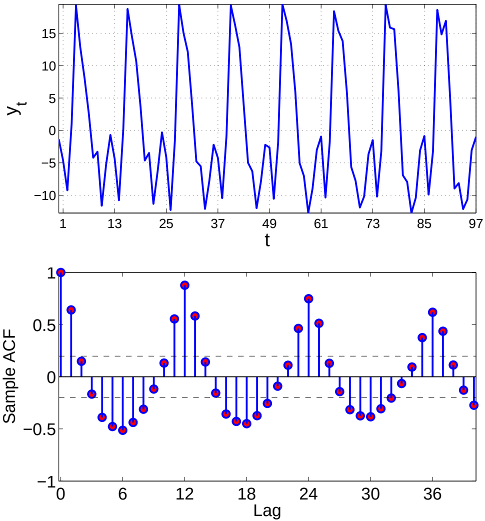


Figure: Simulated $SARIMA(0,0,1) \times (0,1,0)_{12}$, with ACF

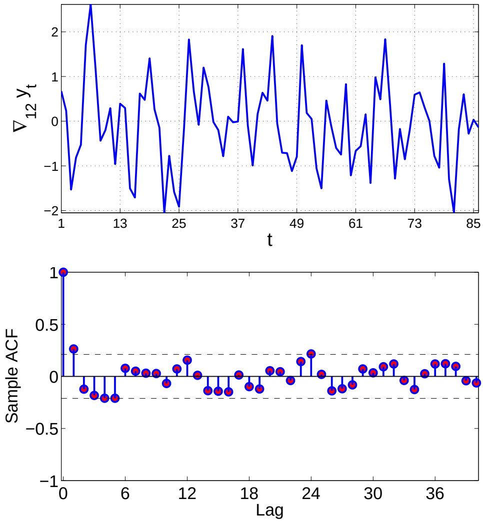


Figure: Seasonally differenced $SARIMA(0, 0, 1) \times (0, 1, 0)_{12}$, with ACF

Model Identification with the ACF

Model	ACF
$AR(1)$	$\rho(k) = \phi_1^k$: exponential decay for $0 < \phi_1 < 1$ (alternating exponential decay if $-1 < \phi_1 < 0$)
$AR(p)$	exponential decay or damped sinusoid
$SAR(1)_s$	zeros at $k \neq \ell s$; non-zeros at lags ℓs, $\ell \in \mathbb{Z}$ decay exponentially
$MA(1)$	$\rho(1) = \frac{-\theta_1}{1+\theta_1^2}$: 'spike' at lag 1, then 0 for lags ≥ 2 (spike is positive if $\theta_1 < 0$ and negative if $\theta_1 > 0$)
$MA(q)$	spikes at lags 1 to q and 0 for lags $\geq q + 1$
$SMA(1)_s$	$\rho(ks) = \frac{-\Theta_1}{1+\Theta_1^2}$: 'spike' at lag s, and 0 otherwise
$ARMA(p, q)$	exponential decay or damped sinusoid (for lags $> q$)
$SARMA$	'periodically extended' version of non-seasonal case

Example: (Hyndman et al 2018) Forecasting: Principles and Practice (2nd ed)

Describe the seasonal ARIMA modelling procedure using quarterly European retail trade data from 1996 to 2011.

- Quarterly retail trade index in the Euro area (17 countries), 1996–2011, covering wholesale and retail trade, and the repair of motor vehicles and motorcycles. (Index: 2005 = 100).

```
autoplot(euretail) + ylab("Retail index") + xlab("Year")
```

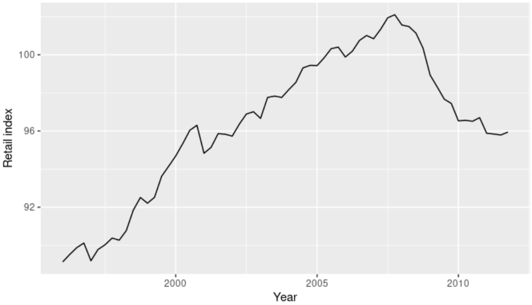


Figure: Source: Rob J. Hyndman - Forecasting: Principles and Practice (2nd ed)

The data are clearly non-stationary, with some seasonality, so we will first take a seasonal difference.

Seasonally differenced European retail trade index.

```
euretail %>% diff(lag=4) %>% ggtsdisplay()
```

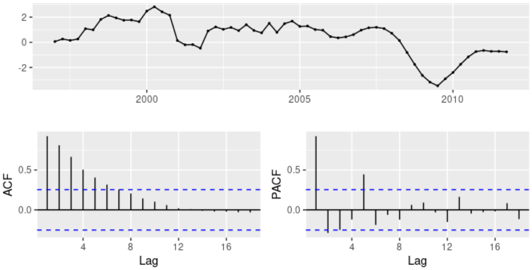


Figure: Source: Rob J. Hyndman - Forecasting: Principles and Practice (2nd ed)

These also appear to be non-stationary, so we take an additional first difference.

Double differenced European retail trade index.

```
euretail % > % diff(lag=4) % > % diff() % > % ggtsdisplay()
```

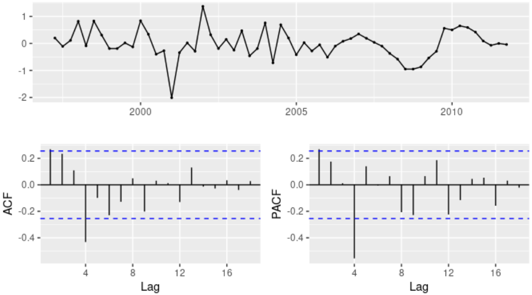


Figure: Source: Rob J. Hyndman - Forecasting: Principles and Practice (2nd ed)

Our aim now is to find an appropriate ARIMA model based on the ACF and PACF shown previously.

- The significant spike at lag 1 in the ACF suggests a non-seasonal MA(1) component, and the significant spike at lag 4 in the ACF suggests a seasonal MA(1) component.
- Consequently, we begin with an ARIMA$(0, 1, 1)(0, 1, 1)_4$ model, indicating a first and seasonal difference, and non-seasonal and seasonal MA(1) components.

The residuals for the fitted model

```
euretail % > %
Arima(order=c(0,1,1), seasonal=c(0,1,1)) % > %
residuals() % > % ggtsdisplay()
```

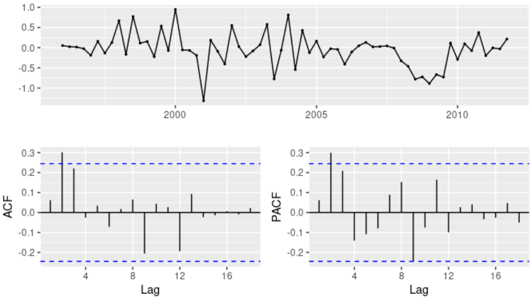


Figure: Source: Rob J. Hyndman - Forecasting: Principles and Practice (2nd ed)

Both the ACF and PACF show significant spikes at lag 2, and almost significant spikes at lag 3, indicating that some additional non-seasonal terms need to be included in the model.

Note:

- The AICc of the ARIMA$(0,1,2)(0,1,1)_4$ model is 74.36, while that for the ARIMA$(0,1,3)(0,1,1)_4$ model is 68.53.
- When other models with AR terms as well, but none that gave a smaller AICc value.
- Consequently, we choose the ARIMA$(0,1,3)(0,1,1)_4$ model.

The residuals for the fitted model

```
fit3 <- Arima(euretail, order=c(0,1,3), seasonal=c(0,1,1))

checkresiduals(fit3)
```

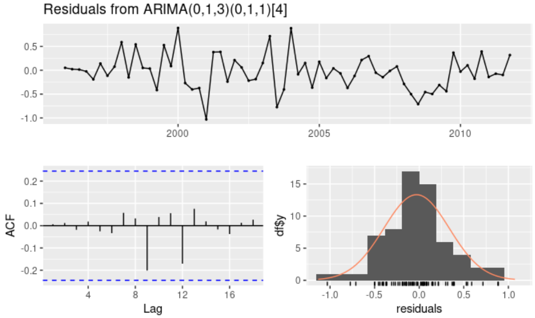


Figure: Source: Rob J. Hyndman - Forecasting: Principles and Practice (2nd ed)

The residuals for the fitted model

```
> Ljung-Box test
> data: Residuals from ARIMA(0,1,3)(0,1,1)[4]
> Q* = 0.51, df = 4, p-value = 1
> Model df: 4. Total lags used: 8
```

Thus, we now have a seasonal ARIMA model that passes the required checks and is ready for forecasting.

```
fit3 % > % forecast(h=12) % > % autoplot()
```

Multiplicative SARIMA Models

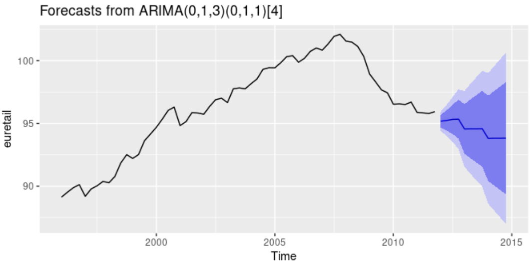


Figure: Source: Rob J. Hyndman - Forecasting: Principles and Practice (2nd ed)