Financial Data Analysis

Time Series Analysis

May 31, 2011

Time Series Analysis

- We discuss techniques for modeling univariate time series.
- These are used, for example, for short–term prediction of asset prices or to test market efficiency.
- We focus on linear **A**uto**R**egressive **M**oving **A**verage (ARMA) models, which is the most commonly used class of models.
- Financial data typically exhibit a more complex structure than can be captured by these processes, but they serve as a useful starting point.
- The concepts used to study these models are also employed in other contexts.

Time Series Analysis

- A time series is a stochastic sequence of random variables,

$$\{Y\}_{t=-\infty}^{\infty} = \{\ldots, Y_{-2}, Y_{-1}, Y_0, Y_1, Y_2, \ldots\}.$$

- The index t of Y_t refers to *time*.
- We will just write "time series Y_t" rather than $\{Y_t\}_{t\in\mathbb{Z}}$.
- In practice, we only observe a finite segment, e.g.,

$$\{Y_1, \ldots, Y_T\}, \tag{1}$$

of a **single realization** of a time series.

- Therefore, in order to have any chance of understanding the system, and predicting the future, we need to assume some *a priori* structure: *stationarity*.

- Stationarity may be viewed as a kind of "*statistical equilibrium*".[1]

[1] Box, Jenkins, and Reinsel (2008), *Time Series Analysis. Forecasting and Control*, 4e, John Wiley & Sons, p. 24.

Strict Stationarity

- The *joint distribution function* (cdf) of a finite set of random variables $\{Y_{t_1}, Y_{t_2}, \cdots, Y_{t_n}\}$ from the collection $\{Y_t\}_{t\in\mathbb{Z}}$ is defined by

$$
\begin{aligned}
& F_{Y_{t_1},Y_{t_2},\ldots,Y_{t_n}}(y_1, y_2, \ldots, y_n) && (2)\\
& \quad = \Pr(Y_{t_1} \le y_1, Y_{t_2} \le y_2, \ldots, Y_{t_n} \le y_n). && (3)
\end{aligned}
$$

  The system of finite–dimensional cdfs uniquely defines a stochastic process.

- A time series is called *strictly stationary* if

$$
\begin{aligned}
& F_{Y_{t_1},Y_{t_2},\ldots,Y_{t_n}}(y_1, y_2, \ldots, y_n) && (4)\\
& = F_{Y_{t_1+h},Y_{t_2+h},\ldots,Y_{t_n+h}}(y_1, y_2, \ldots, y_n),
\end{aligned}
$$

  where the equality must hold for all possible finite sets of indices $t_1, t_2, \ldots, t_n$ and $h \in \mathbb{Z}$ and all $(y_1, y_2, \ldots, y_n)$ in the range of the random variable Y_t.

- Note that the indices $t_1, t_2, \ldots, t_n$ are not necessarily consecutive.

Strict Stationarity

- For example, take $n = 1$.

- Then we simply have

$$F_{Y_s}(y) = F_{Y_t}(y) \quad \text{for all } s \text{ and } t \text{ (and all } y\text{)}.$$

- This means that the process has the same *marginal distribution* at each point of time, and we can hope to learn about the properties of the distribution of Y_t by treating the observed segment as a (although not independent) sample from the same distribution.

Strict Stationarity

- Now take $n = 2$.

- Then we have

$$F_{Y_s,Y_t}(y_1, y_2) = F_{Y_{s+h},Y_{t+h}}(y_1, y_2) \quad \text{for all } s \text{ and } t \text{ and } h.$$

- This means that the dependence structure between two variables of a time series (as embodied in their joint distribution) depends only on their "distance in time", i.e., $|s - t|$.

- It does not depend on h. That is, the dependence between the return today and the return tomorrow is the same as that between the return in two weeks and the return in two weeks plus one day; here $|s - t| = 1$ and $h = 10$ (ten trading days).

- This is useful since in general we are doing time series analysis because we are interested in *conditional distributions*, i.e., the distributions of future returns given the return path up to now.

Strict Stationarity

- More generally, strict stationarity thus implies that all the multivariate distributions for subsets of n variables in (4) must agree with their counterparts in the shifted set for all h.

- That is, their joint distribution depends only on the distance between the elements $t_1, t_2, \ldots, t_n \in \mathbb{Z}$, and not on their actual values.

Weak Stationarity

- A somewhat different concept of stationarity is **weak stationarity**, or **covariance stationarity**, or **wide–sense stationarity**, or **second–order stationarity**.

- This imposes conditions on the first two moments of the series.

- That is, time series Y_t is weakly stationary if

(1) the mean function is constant and finite, i.e.,

$$\mu_t := \mathsf{E}(Y_t) = \mu < \infty \quad \text{for all } t, \tag{5}$$

and

(2) the autocovariance function,

$$\gamma(s,t) = \mathsf{Cov}(Y_s, Y_t) = \mathsf{E}(Y_s Y_t) - \mathsf{E}(Y_s)\mathsf{E}(Y_t), \tag{6}$$

exists and depends only on the distance in time, $\tau = |s - t|$, between the two random variables.

- Condition (2) for $s - t = 0$ requires that the second moment (and hence the variance) is finite and does not depend on time, i.e.,

$$\mathsf{E}(Y_t^2) =: \gamma(0) < \infty \quad \text{for all } t. \tag{7}$$

- If (7) holds, we also have (Cauchy–Schwarz inequality)

$$|\mathsf{E}(Y_t Y_{t-\tau})| \leq \sqrt{\mathsf{E}(Y_t^2)\mathsf{E}(Y_{t-\tau}^2)} = \mathsf{E}(Y_t^2) < \infty.$$

Weak Stationarity

- For a weakly stationary process, we can thus define the autocovariance function at lag $\tau \in \mathbb{Z}$, $\gamma(\tau)$,

$$\begin{aligned}\gamma(\tau) := \mathsf{Cov}(Y_t, Y_{t-\tau}) &= \mathsf{E}(Y_t Y_{t-\tau}) - \mathsf{E}(Y_t)\mathsf{E}(Y_{t-\tau}) \\ &= \mathsf{E}(Y_t Y_{t-\tau}) - \mathsf{E}^2(Y_t),\end{aligned}$$

  and the **autocorrelation function**

$$\rho(\tau) := \frac{\mathsf{Cov}(Y_t, Y_{t-\tau})}{\mathsf{Var}(Y_t)} = \frac{\mathsf{E}(Y_t Y_{t-\tau}) - \mathsf{E}^2(Y_t)}{\mathsf{E}(Y_t^2) - \mathsf{E}^2(Y_t)} = \frac{\gamma(\tau)}{\gamma(0)}.$$

- Clearly $\gamma(\tau) = \gamma(-\tau)$. Thus, it suffices to consider $\tau \geq 0$.

- In the following, it is always understood that $\tau \geq 0$.

- The **sample autocorrelation function** (SACF), to be defined below, provides an indication of the extent to which it is possible to forecast a series from its own past.

Example: White Noise

- A (weakly) stationary time series $\{\epsilon_t\}$ is *white noise* if

$$\begin{aligned} \mu &= 0, \quad \text{and} \\ \gamma(\tau) &= \begin{cases} \sigma^2 & \text{for } \tau = 0 \\ 0 & \text{for } \tau \neq 0, \end{cases} \end{aligned}$$

  that is, it is an uncorrelated zero–mean process.

- For example, if $\{\epsilon_t\}$ is independent and identically distributed (iid), then it is *strict* (or independent).

- White noise processes are important building blocks to construct more complex time series.

Weak and Strict Stationarity

- The terminology suggests that weak stationarity is a weaker property than strict stationarity, or that strict stationarity implies weak stationarity.

- Due to the "$< \infty$" part of the weak stationarity condition, this may not be the case, however.

- For example, if Y_t is iid but follows a *Cauchy* distribution[2] with pdf

$$f(y) = \frac{1}{\pi(1+y^2)},$$

  then even the mean does not exist.

- This process is strictly but not weakly stationary.

- A strictly stationary process with finite first and second moments is weakly stationary.

[2] This is a Student's t distribution with one degree of freedom.

Weak and Strict Stationarity

- A process $\{Y_t\}$ is a Gaussian process if all the n–dimensional random vectors in (4) have multivariate normal distributions.

- A Gaussian white noise process is called *Gaussian white noise*.

The Partial Autocorrelation Function

- In addition to the ACF, the **partial autocorrelation function** turns out to be useful for the identification of time series processes.

- The ACF at lag τ measures the unconditional correlation between Y_t and $Y_{t-\tau}$ without taking the influence of the intervening variables $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-\tau+1}$ into account.

- The partial autocorrelation between Y_t and $Y_{t-\tau}$, denoted by $\pi(\tau)$, reflects the net association between Y_t and $Y_{t-\tau}$ over and above that part of the association that results from their mutual relationship with $Y_{t-1}, \ldots, Y_{t-\tau-1}$.

- More precisely, the partial autocorrelation between Y_t and $Y_{t-\tau}$ is the simple linear correlation between Y_t and $Y_{t-\tau}$ after removing the linear effects of variables $Y_1, \ldots, Y_{t-\tau}$.

The Partial Autocorrelation Function

- Suppose we attempt to linearly approximate Y_t by $Y_{t-1}, \ldots, Y_{t-\tau-1}$,

$$\widehat{Y}_t = \sum_{i=1}^{\tau-1} \alpha_i Y_{t-i}, \tag{8}$$

where the coefficients of the best linear approximation may be obtained by minimizing the mean squared error[3]

$$\mathsf{E}(Y_t - \widehat{Y}_t)^2 = \mathsf{E}\left\{\left(Y_t - \sum_{i=1}^{\tau-1} \alpha_i Y_{t-i}\right)^2\right\}. \tag{9}$$

[3] Think of this as a population version of linear regression.

The same can be done for $Y_{t-\tau}$,

$$\widehat{Y}_{t-\tau} = \sum_{i=1}^{\tau-1} \alpha_i Y_{t-\tau+i}, \tag{10}$$

and it turns out that the coefficients $\alpha_1, \ldots, \alpha_{\tau-1}$ in (8) and (10) are the same due to stationarity.

- The partial autocorrelation at lag τ is then defined as

$$\pi(1) = \rho(1),$$

and, for $\tau \geq 2$

$$\pi(\tau) = \text{Corr}\{(Y_t - \widehat{Y}_t), (Y_{t-\tau} - \widehat{Y}_{t-\tau})\},$$

i.e., it is the correlation between Y_t and $Y_{t-\tau}$ after removing the linear effects of $Y_{t-1}, \ldots, Y_{t-\tau+1}$.

- Several simple examples will be considered below.

Moving Average (MA) Processes

- A moving average process of order q, or MA(q) process, is defined by

$$Y_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t, \tag{11}$$

where $\{\epsilon_t\}$ is a white noise process.

- At time t, the process is a weighted average of $\epsilon_t, \ldots, \epsilon_{t-q}$, which moves through time; thus the name.

- Occasionally, it is convenient to define $\theta_0 = 1$, so that

$$Y_t = \mu + \sum_{i=0}^{q} \theta_i \epsilon_{t-i}. \tag{12}$$

- We can calculate

$$\mathsf{E}(Y_t) = \mu,$$

and

$$\begin{aligned}\gamma(\tau) &= \mathsf{E}\left[\left(\sum_{i=0}^{q}\theta_i\epsilon_{t-1}\right)\left(\sum_{j=0}^{q}\theta_j\epsilon_{t-\tau-j}\right)\right] \\ &= \sum_{i=0}^{n}\sum_{j=0}^{n}\theta_i\theta_j\mathsf{E}(\epsilon_{t-i}\epsilon_{t-\tau-j}).\end{aligned}$$

- Now

$$\mathsf{E}(\epsilon_{t-i}\epsilon_{t-\tau-j}) = \begin{cases}\sigma^2 & \text{for } t-i = t-\tau-j \Leftrightarrow i = \tau+j \\ 0 & i \neq \tau + j,\end{cases}$$

and so

$$\gamma(\tau) = \begin{cases}\sigma^2\sum_{j=0}^{q-\tau}\theta_j\theta_{j+\tau} & \text{for } \tau \leq q \\ 0 & \text{for } \tau > q.\end{cases}$$

- The variance is

$$\gamma(0) = \sigma^2\sum_{i=0}^{q}\theta_j^2,$$

and the ACF

$$\rho(\tau) = \begin{cases} \frac{\sum_{j=0}^{q-\tau} \theta_j \theta_{j+\tau}}{\sum_{j=0}^{q} \theta_j^2} & \text{for } \tau \leq q \\ 0 & \text{for } \tau > q. \end{cases} \tag{13}$$

- We observe that the ACF of an MA(q) process cuts off after lag q.
- Finite–order MA(q) processes are weakly stationary.
- If $\{\epsilon_t\}$ is Gaussian white noise, the marginal distribution of $\{Y_t\}$ is likewise normal and the process is also strictly stationary.

Moving Average (∞) Processes

- We also define stationary *infinite–order moving average*, or, in short, MA(∞), processes, i.e.,

$$Y_t = \mu + \epsilon_t + \sum_{i=1}^{\infty} \theta_i \epsilon_{t-i}, \tag{14}$$

provided that

$$\sum_{i=1}^{\infty} |\theta_i| < \infty,$$

i.e., the sequence of MA–coefficients, $\theta_i, i = 1, 2, \ldots$, is absolutely summable.

- In this case, we can compute the moments (proceeding in the same manner as for finite–order MA(∞) processes) as

$$\begin{aligned} \mathsf{E}(Y_t) &= \mu, & (15) \\ \gamma(0) &= \sigma^2 \sum_{i=0}^{\infty} \theta_i^2, & (16) \\ \gamma(\tau) &= \sigma^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+\tau}, & (17) \\ \rho(\tau) &= \frac{\sum_{i=0}^{\infty} \theta_i \theta_{i+\tau}}{\sum_{i=0}^{\infty} \theta_i^2}. & (18) \end{aligned}$$

- If $\{\epsilon_t\}$ is strict (independent) white noise, then process (14) is said to be a **linear time series**.

The Partial Autocorrelation Function of MA Processes

- Consider the MA(1) process,

$$Y_t = \theta \epsilon_{t-1} + \epsilon_t, \tag{19}$$

with

$$\gamma(\tau) = \begin{cases} (1+\theta^2)\sigma^2, & \tau = 0 \\ \theta\sigma^2 & \tau = 1 \\ 0, & \tau > 1, \end{cases}$$

and

$$\rho(\tau) = \frac{\theta}{1+\theta^2} \quad \text{for } \tau = 1 \tag{20}$$

and $\rho(\tau) = 0$ for $\tau > 1$.

- Now Y_t depends on ϵ_t and ϵ_{t-1}, and Y_{t-2} depends on ϵ_{t-2} and ϵ_{t-3}, so there is no correlation at all between Y_t and Y_{t-2}: no "raw" correlation, no partial correlation, no whatsoever correlation, right?

- However, when it comes to partialling out Y_{t-1}, we have

$$\dots, \overbrace{\epsilon_t, \epsilon_{t-1}}^{Y_t}, \overbrace{\epsilon_{t-2}, \epsilon_{t-3}}^{Y_{t-2}}, \dots, \qquad (21)$$

$$\dots, \epsilon_t, \underbrace{\epsilon_{t-1}, \epsilon_{t-2}}_{Y_{t-1}}, \epsilon_{t-3}, \dots,$$

so that we may actually generate some dependence when involving Y_{t-1}.

- The coefficient α of the best linear approximation is determined by

$$\begin{aligned} \min_{\alpha} \mathsf{E}(Y_t - \alpha Y_{t-1})^2 &= (1+\alpha^2)\mathsf{E}(Y_t^2) - 2\alpha \mathsf{E}(Y_t Y_{t-1}) \\ &\Rightarrow \alpha = \frac{\mathsf{E}(Y_t Y_{t-1})}{\mathsf{E}(Y_t^2)} = \frac{\theta}{1+\theta^2} = \rho(1), \end{aligned}$$

and by stationarity the same result is obtained for the approximation of Y_{t-2}.

- Then the second–order partial covariance between Y_t and Y_{t-2} is

$$\mathsf{PCov}(Y_t, Y_{t-2}) = \mathsf{E}\{(Y_t - \alpha Y_{t-1})(Y_{t-2} - \alpha Y_{t-1})\} = -\frac{\mathsf{E}^2(Y_t Y_{t-1})}{\mathsf{E}(Y_t^2)} = \sigma^2 \frac{-\theta^2}{1+\theta^2},$$

and thus, since

$$\text{Var}(Y_t - \alpha Y_{t-1}) = \text{Var}(Y_{t-2} - \alpha Y_{t-1}) = \sigma^2 \frac{1+\theta^2+\theta^4}{1+\theta^2},$$

the second–order partial autocorrelation of the MA(1) process,

$$\pi(2) = \frac{-\theta^2}{1+\theta^2+\theta^4}.$$

In general, for the MA(1) process,

$$\pi(\tau) = \frac{(-1)^{\tau+1}\theta^\tau}{1+\theta^2+\theta^4+\cdots+\theta^{2\tau}} = \frac{(-1)^{\tau+1}\theta^\tau(1-\theta^2)}{1-\theta^{2(\tau+1)}}.$$

- The **general pattern** for MA(q) processes is
  - the autocorrelation function cuts off after lag q,
  - the partial autocorrelation function tails off.

Autoregressive Processes

- An autoregressive process of order p, abbreviated AR(p), is of the form

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t, \tag{22}$$

where ϵ_t is white noise with mean zero and variance σ^2.

As written in (22), the mean of the process is zero. If the mean μ is not zero, we may write

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t, \tag{23}$$

or alternatively

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t, \tag{24}$$

where

$$c = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p). \tag{25}$$

Lag Operators

- The lag operator L is defined as

$$Ly_t = y_{t-1}. \tag{26}$$

We have

(i) $L^2 y_t = L(Ly_t) = Ly_{t-1} = y_{t-2}$, $L^q y_t = y_{t-q}$
(ii) For constant c, $Lc = c$, and $L(cy_t) = c(Ly_t) = cy_{t-1}$.
(iii) $L(y_t + x_t) = Ly_t + Lx_t = y_{t-1} + x_{t-1}$.
(iv) An example of a *lag polynomial* is

$$\begin{aligned}(1-\lambda_1 L)(1-\lambda_2 L)y_t &= \{1-(\lambda_1+\lambda_2)L+\lambda_1\lambda_2 L^2\}y_t \\ &= y_t-(\lambda_1+\lambda_2)y_{t-1}+\lambda_1\lambda_2 y_{t-2}.\end{aligned}$$

- Consider the first–order autoregressive (AR(1)) process, which upon repeated substitution yields

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \epsilon_t \\ &= c(1+\phi) + \phi\epsilon_{t-1} + \epsilon_t + \phi^2 y_{t-2} \\ &\vdots \\ &= c\sum_{i=0}^{\tau-1}\phi^i + \sum_{i=0}^{\tau-1}\phi^i\epsilon_{t-i} + \phi^\tau y_{t-\tau}. \end{aligned} \tag{27}$$

If $|\phi| < 1$ and $\tau \rightarrow \infty$, we obtain an MA(∞) process,[4]

$$y_t = \frac{c}{1-\phi} + \sum_{i=0}^{\infty}\phi^i\epsilon_{t-i}. \tag{29}$$

[4] Recall that

$$\sum_{i=0}^{\tau}\phi^i = \frac{1-\phi^{\tau+1}}{1-\phi}, \quad \lim_{\tau\rightarrow\infty}\sum_{i=0}^{\tau}\phi^i = \sum_{i=0}^{\infty}\phi^i = \lim_{\tau\rightarrow\infty}\frac{1-\phi^{\tau+1}}{1-\phi} = \frac{1}{1-\phi} \text{ if } |\phi| < 1. \tag{28}$$

On the other hand, using lag operator notation, (27) can be written

$$y_t - \phi y_{t-1} = (1 - \phi L)y_t = c + \epsilon_t. \tag{30}$$

- Comparison of (29) and (30) suggests that, if $|\phi| < 1$, we can invert and expand the lag polynomial according to a geometric series,

$$(1 - \phi L)^{-1} = \frac{1}{1 - \phi L} = \sum_{i=1}^{\infty} \phi^i L^i, \tag{31}$$

so that

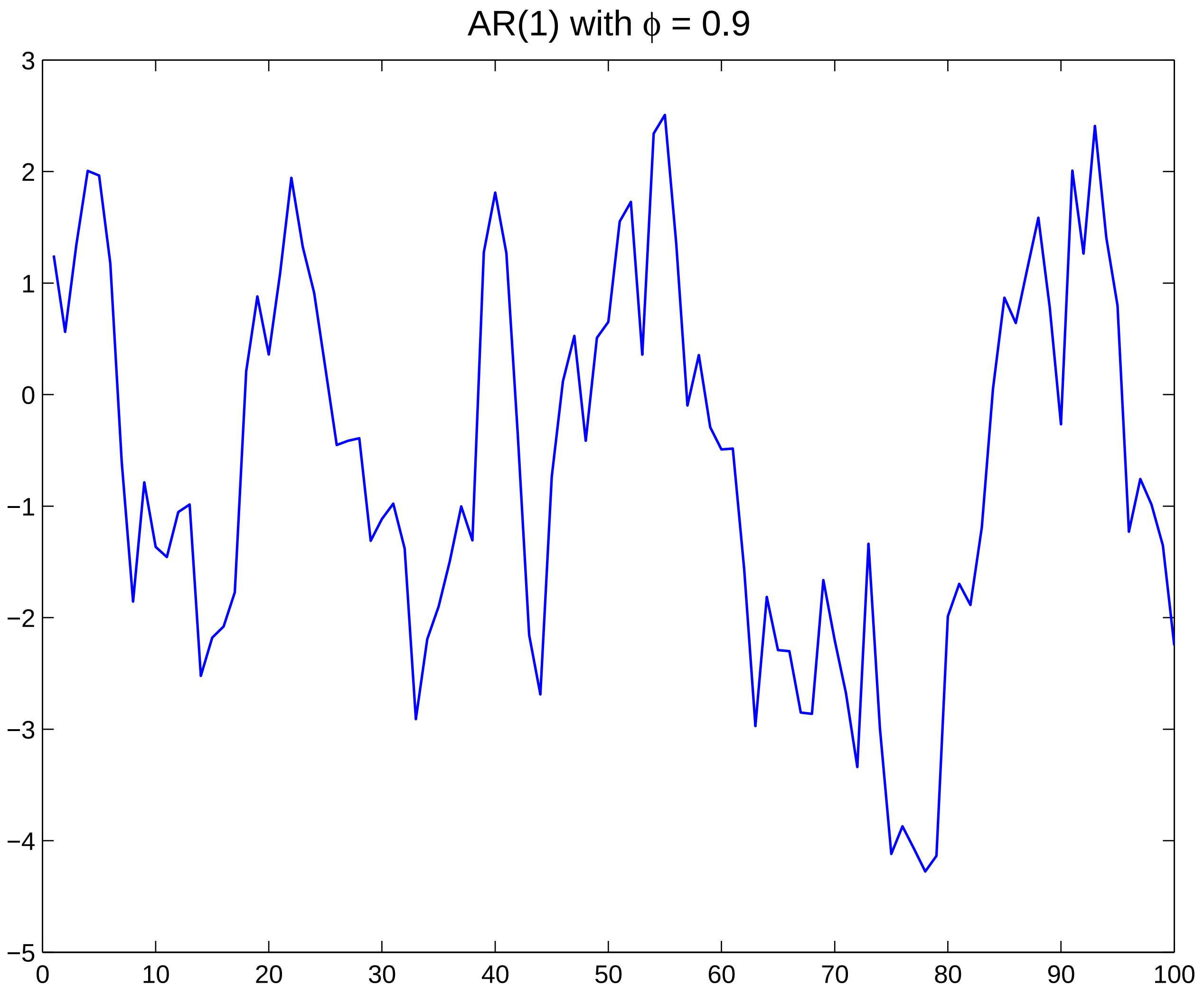
$$\begin{aligned}
(1 - \phi L)y_t &= c + \epsilon_t \\
\Rightarrow y_t &= \frac{c}{1 - \phi L} + \frac{\epsilon_t}{1 - \phi L} \\
&= \frac{c}{1 - \phi} + \sum_{i=0}^{\infty} \phi^i (L^i \epsilon_t) \\
&= \frac{c}{1 - \phi} + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}.
\end{aligned}$$

- For our purposes, the operator L can formally be manipulated as if it were a number with absolute value 1.

- The MA(∞) process has absolutely summable coefficients and is the stationary solution of the AR(1) process with $\phi < 1$. The moments of this process follow from (15)–(18), namely,

$$\begin{aligned}
\mathsf{E}(Y_t) &= \frac{c}{1-\phi}, && (32)\\
\gamma(0) &= \sigma^2 \sum_{i=0}^{\infty} \phi^{2i} = \frac{\sigma^2}{1-\phi^2}, && (33)\\
\gamma(\tau) &= \sigma^2 \sum_{i=0}^{\infty} \phi^{2i+\tau} = \frac{\sigma^2 \phi^\tau}{1-\phi^2}, && (34)\\
\rho(\tau) &= \phi^\tau. && (35)
\end{aligned}$$

- Note that the ACF decays to zero geometrically with rate ϕ, which thus can be viewed as a measure of the memory or persistence of the process.



Observations close together in time are positively correlated with $\phi = 0.9$.

AR(p) processes

- For the AR(p) process,

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

we can write

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) Y_t = \epsilon_t,$$

or, defining the lag polynomial $\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$,

$$\phi(L) Y_t = \epsilon_t. \tag{36}$$

- Let the roots of the *characteristic polynomial*

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \tag{37}$$

be denoted by $z_1, \ldots, z_p$, and the roots of the reverse characteristic polynomial

$$z^p \phi(z^{-1}) = z^p - \phi_1 z^{p-1} - \cdots - \phi_{p-1} z - \phi_p \tag{38}$$

be denoted by $\lambda_1, \ldots, \lambda_p$, where $\lambda_i = 1/z_i$, $i = 1, \ldots, p$.

- Assume that

$$|z_i| > 1, \quad i = 1, \ldots, p, \text{ i.e., } \phi(z) = 0 \Rightarrow |z| > 1, \tag{39}$$

i.e., all the roots of $\phi(z)$ are outside the unit circle, and thus those of the reverse polynomial are inside the unit circle.

- We can then invert the lag polynomial to get a stationary MA(∞) process.

- Factorizing the characteristic polynomial,

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \tag{40}$$

$$= (-\phi_p) \prod_{i=1}^{p} (z - z_i). \tag{41}$$

Since $\phi(0) = 1$, we have

$$(-\phi_p)(-1)^p \prod_{i=1}^{p} z_i = 1, \tag{42}$$

and so

$$\phi(z) = (-\phi_p) \prod_{i=1}^{p} (z - z_i) = (-\phi_p)(-1)^p \prod_{i=1}^{p} z_i \prod_{i=1}^{p} \left(1 - \frac{z}{z_i}\right)$$

$$= \prod_{i=1}^{p} (1 - \lambda_i z) \, . \tag{43}$$

- We can then write Y_t as a stationary MA(∞) process,

$$\begin{aligned} Y_t &= \phi(1)^{-1}c + \phi(L)^{-1}\epsilon_t && (44) \\ &= \frac{c}{1-\phi_1-\phi_2-\cdots-\phi_p} + \frac{\epsilon_t}{(1-\lambda_1 L)(1-\lambda_2 L)\cdots(1-\lambda_p L)} \\ &= \frac{c}{1-\phi_1-\phi_2-\cdots-\phi_p} + \sum_{i=1}^{\infty}\theta_i\epsilon_{t-i}. && (45) \end{aligned}$$

- Condition

$$|z_i| > 1, \quad i = 1, \ldots, p, \text{ i.e., } \phi(z) = 0 \Rightarrow |z| > 1, \tag{46}$$

is also referred to as the **stationarity condition** for the AR(p) process. If it is satisfied, a stationary solution exists and is given by the MA(∞) representation.

- It is also the stability condition for the difference equation associated with the deterministic part of the autoregressive equation.

– For example, for the AR(2) with $c = 0$,

$$\begin{aligned} Y_t &= \frac{\epsilon_t}{(1-\lambda_1 L)(1-\lambda_2 L)} = \frac{1}{\lambda_1 - \lambda_2}\left[\frac{\lambda_1}{1-\lambda_1 L} - \frac{\lambda_2}{1-\lambda_2 L}\right]\epsilon_t \\ &= \frac{1}{\lambda_1 - \lambda_2}\sum_{i=0}^{\infty}(\lambda_1^{i+1} - \lambda_2^{i+1})L^i \epsilon_t \\ &= \sum_{i=0}^{\infty}\underbrace{\frac{(\lambda_1^{i+1} - \lambda_2^{i+1})}{\lambda_1 - \lambda_2}}_{=\theta_i}\epsilon_{t-i}. \end{aligned}$$

– For more concreteness, consider the process

$$Y_t = 0.7Y_{t-1} - 0.1Y_{t-2} + \epsilon_t, \quad \text{or} \quad (1 - 0.7L + 0.1L^2)Y_t = \epsilon_t. \tag{47}$$

– The (reverse) characteristic equation of this process is

$$\lambda^2 - 0.7\lambda + 0.1 = 0 \Rightarrow \lambda_{1/2} = \frac{0.7 \pm \sqrt{0.49 - 0.4}}{2} = \frac{0.7 \pm 0.3}{2}, \tag{48}$$

or $\lambda_1 = 0.5$ and $\lambda_2 = 0.2$, and the inverse roots are $z_1 = 2$ and $z_2 = 5$. As λ_1 and λ_2 are both smaller than one in magnitude (hence z_1 and z_2 are larger than one in magnitude), the process is stationary. To find the stationary MA(∞) representation, we factorize the lag polynomial as

$$\begin{aligned} 1 - 0.7L + 0.1L^2 &= 0.1(L - z_1)(L - z_2) = 0.1 z_1 z_2 \left(1 - \frac{L}{z_1}\right)\left(1 - \frac{L}{z_2}\right) \\ &= 0.1 \times 5 \times 2(1 - \lambda_1 L)(1 - \lambda_2 L) = (1 - \lambda_1 L)(1 - \lambda_2 L) \\ &= (1 - 0.5L)(1 - 0.2L), \end{aligned}$$

i.e.,

$$(1 - 0.5L)(1 - 0.2L)Y_t = \epsilon_t,$$

and inverting and expanding the lag polynomial gives the MA(∞)

representation as

$$
\begin{aligned}
Y_t &= \frac{\epsilon_t}{(1-0.5L)(1-0.2L)} = \frac{1}{0.5-0.2}\left(\frac{0.5}{1-0.5L} - \frac{0.2}{1-0.2L}\right)\epsilon_t \\
&= \frac{1}{0.3}\sum_{i=0}^{\infty}\left(0.5^{i+1} - 0.2^{i+1}\right)L^i\epsilon_t = \frac{1}{0.3}\sum_{i=0}^{\infty}\left(0.5^{i+1} - 0.2^{i+1}\right)\epsilon_{t-i} \\
&= \sum_{i=0}^{\infty}\theta_i\epsilon_{t-i}, \quad \text{where} \quad \theta_i = \frac{0.5^{i+1} - 0.2^{i+1}}{0.3}, \quad i = 0, 1, 2, \ldots
\end{aligned}
$$

– Note that Y_t depends only on current and past values of the white noise process $\{\epsilon_t\}$, so that

$$
\mathsf{Cov}(Y_t, \epsilon_{t+\tau}) = \mathsf{E}(Y_t\epsilon_{t+\tau}) = 0, \quad \tau > 0. \tag{49}
$$

- If the stability condition is satisfied and the process is allowed to be initialized in the infinite past, or is assumed to be initialized with its stationary distribution (moments), it is stationary.
- Otherwise, (46) guarantees that it converges rapidly to stationarity, and is termed *asymptotically stationary*.

- If the innovations ϵ_t are Gaussian white noise, the stationary distribution is likewise Gaussian.

- The coefficients θ_i in the MA(∞) representation $Y_t = \epsilon_t + \sum_{i=0}^{\infty} \theta_i \epsilon_{t-i}$ are also called the **impulse response coefficients** of the process, since θ_i shows the impact of a shock (impulse) in period $t - i$ on the system's output in period t.

Calculating and characterizing the autocorrelations of a stationary AR(p) process: Yule–Walker equations

- In principle, the moment structure of an AR(p) process could be calculated from its MA(∞) representation, but this is very cumbersome.

- Alternatively, we can use the *Yule–Walker* equations.

- The process is

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

- Multiply by $Y_{t-\tau}$, $\tau \geq 0$ and take expectations,

$$\mathsf{E}(Y_t Y_{t-\tau}) = \phi_1 \mathsf{E}(Y_{t-1} Y_{t-\tau}) + \cdots + \phi_p \mathsf{E}(Y_{t-p} Y_{t-\tau}) + \mathsf{E}(\epsilon_t Y_{t-\tau}).$$

- That is, for $\tau = 0$,

$$\gamma(0) = \sum_{i=1}^{p} \phi_i \gamma(i) + \sigma^2, \tag{50}$$

and for $\tau > 0$,

$$\gamma(\tau) = \sum_{i=1}^{p} \phi_i \gamma(\tau - i), \tag{51}$$

which carries over to the autocorrelations, i.e., for $\tau > 0$,

$$\rho(\tau) = \phi_1 \rho(\tau - 1) + \phi_2 \rho(\tau - 2) + \cdots + \phi_p \rho(\tau - p), \tag{52}$$

and for $\tau \geq p$, the ACF follows a pth order linear difference equation.

- For constants $c_1, \ldots, c_p$ to be determined from the first p autocorrelations $\rho(0), \rho(1), \ldots, \rho(p-1)$, the solution of this difference equation can be written as[5]

$$\rho(\tau) = c_1 \lambda_1^\tau + \cdots + c_p \lambda_p^\tau, \quad \tau = 0, 1, \ldots, \tag{53}$$

where $\lambda_1, \ldots, \lambda_p$ are the roots of the reverse characteristic equation

$$z^p \phi(z^{-1}) = z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \cdots - \phi_{p-1} z - \phi_p = 0, \tag{54}$$

[5] Here we assume that the roots are all distinct. In case there are multiple roots, (53) has to be modified slightly, but the central message remains unaffected.

and, by the stationarity condition, $|\lambda_i| < 1$, $i = 1, \ldots, p$, so

$$\lim_{\tau \to \infty} \rho(\tau) = 0. \tag{55}$$

- Thus we observe that the ACF is described by a mixture of damped exponentials (real roots) and damped sine waves (complex roots).
- In particular, the ACF of AR processes dies out gradually and, in contrast to MA(q) processes, does not cut off after a specific lag.
- The speed of convergence in (55) is governed by $\max_{1 \leq i \leq p}\{|\lambda_i|\}$, so the largest root in magnitude of the (reverse) characteristic polynomial may be taken as a measure for the memory of the process.

- Example 1: Consider the AR(2) process

$$Y_t = 1.4Y_{t-1} - 0.48Y_{t-2} + \epsilon_t. \tag{56}$$

- The roots of the polynomial $z^2 - \phi_1 z - \phi_2$ are given by 0.8 and 0.6, so the solution is of the form

$$\rho(\tau) = c_1 0.8^\tau + c_2 0.6^\tau, \quad \tau \geq 0. \tag{57}$$

- To find coefficients c_1 and c_2, we first note that $\rho(0) = 1$.

- Moreover, the Yule–Walker equation for $\tau = 1$,

$$\rho(1) = \phi_1 \underbrace{\rho(0)}_{=1} + \phi_2 \rho(1), \tag{58}$$

gives

$$\rho(1) = \frac{\phi_1}{1 - \phi_2} = \frac{1.4}{1 + .48} = 0.946. \tag{59}$$

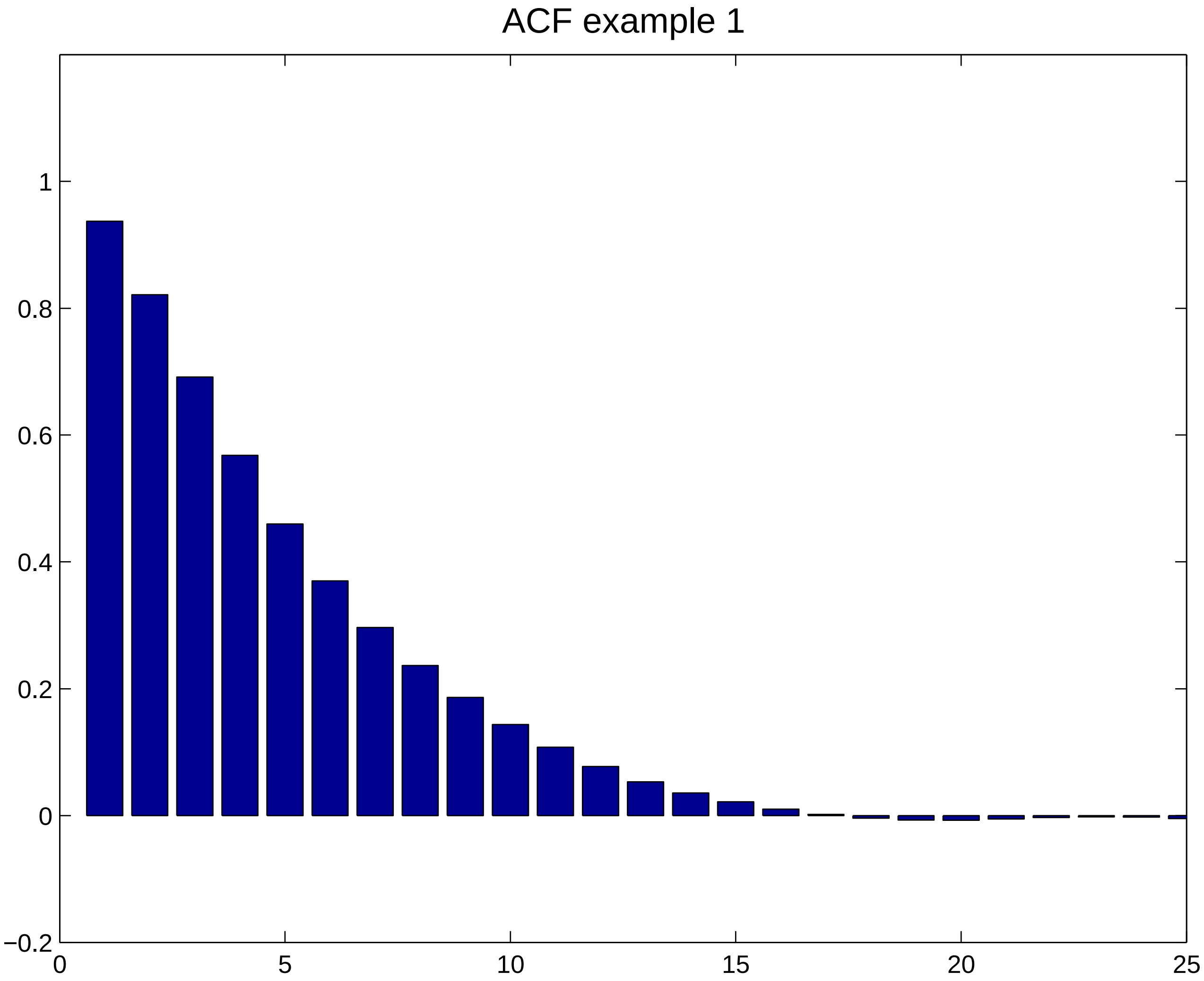
- We then solve

$$c_1 + c_2 = 1 (= \rho(0)) \tag{60}$$

$$0.8c_1 + 0.6c_2 = 0.946 (= \rho(1)), \tag{61}$$

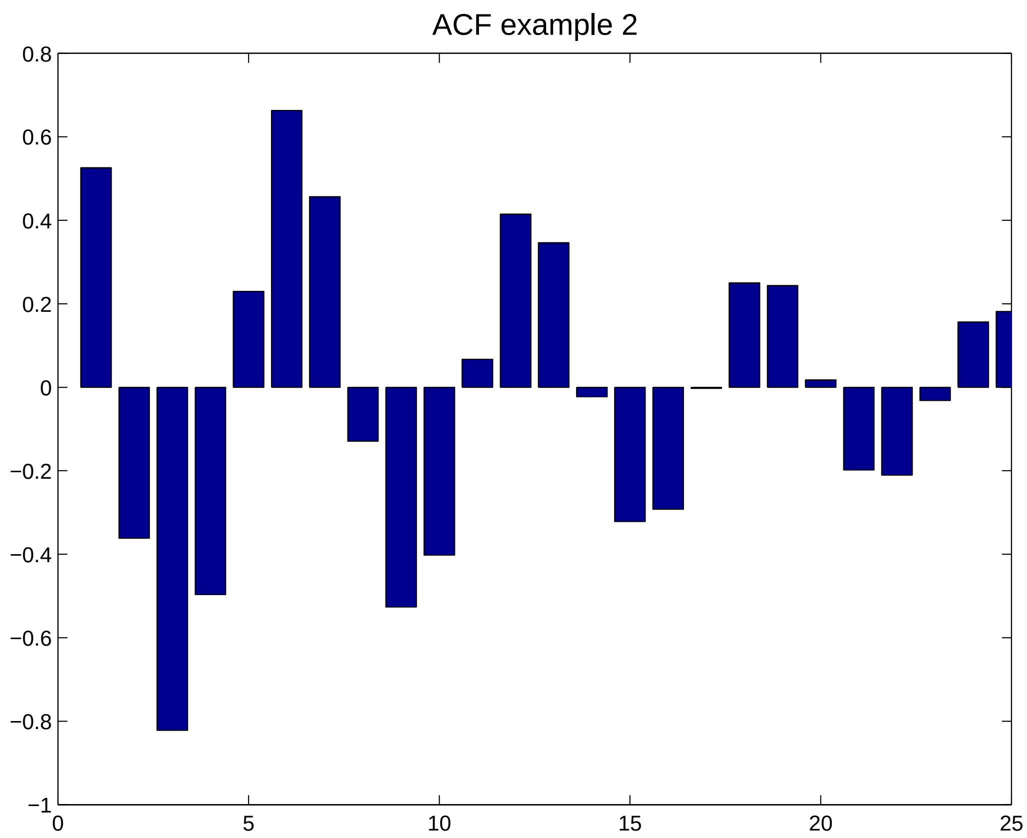
which gives $c_1 = 1.73$ and $c_2 = -0.73$; thus, the ACF of process (56) is

$$\rho(\tau) = 1.73 \times 0.8^\tau - 0.73 \times 0.6^\tau. \tag{62}$$


ACF example 1
1
0.8
0.6
0.4
0.2
0
−0.2
0
5
10
15
20
25

- Example 2: $Y_t = Y_{t-1} - 0.89Y_{t-2} + \epsilon_t$
  This is a process with complex roots $z_{1/2} = 0.5 \pm i0.8$.

- The correlogram of this process has a distinctive cyclical pattern associated with complex roots. This illustrates how AR processes can generate time series with moderately regular cycles.



Partial autocorrelations of AR processes

- For the AR(1) process $Y_t = \phi Y_{t-1} + \epsilon_t$, the second–order partial autocovariance is

$$\begin{aligned} \mathsf{E}(Y_t - \phi Y_{t-1})(Y_{t-2} - \phi Y_{t-1}) &= \gamma(2) - 2\phi\gamma(1) + \phi^2\gamma(0) \\ &= \gamma(0)(\phi^2 - 2\phi^2 + \phi^2) \\ &= 0. \end{aligned}$$

- The **general pattern** for AR(q) processes is
  - the autocorrelation function tails off,
  - the partial autocorrelation function cuts off after lag p.

- Compare this with the MA(q) case.

Autoregressive Moving Average (ARMA) Time Series Models

- ARMA(p, q) is given by

$$\phi(L)Y_t = \theta(L)\epsilon_t, \tag{63}$$

where the lag polynomials

$$\begin{aligned} \phi(L) &= 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p & (64) \\ \theta(L) &= 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q. & (65) \end{aligned}$$

- Assume that $\phi(z)$ and $\theta(z)$ have no roots in common.

- Mixing MA and AR parts often leads to more flexible models with less parameters than using pure MA or AR models.

- As MA processes are stationary, the stationarity of an ARMA model depends on the autoregressive polynomial $\phi(z)$, and the stationarity condition is identical to (46).

- If the stationarity condition holds, we can invert the lag polynomial $\phi(L)$ to obtain an MA(∞) process,

$$Y_t = \frac{\theta(L)}{\phi(L)}\epsilon_t. \tag{66}$$

- For example, consider the ARMA(1,1) process,

$$Y_t = \phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t, \tag{67}$$

which can be inverted

$$\begin{aligned} Y_t &= \frac{1+\theta L}{1-\phi L}\epsilon_t = (1+\theta L)\sum_{i=0}^{\infty} \phi^i L^i \epsilon_t \\ &= \epsilon_t + (\theta + \phi)\sum_{i=1}^{\infty} \phi^{i-1}\epsilon_{t-i}. \end{aligned} \tag{68}$$

Autocorrelations of ARMA processes

- The autocorrelations could in principle be obtained from the MA(∞) representation.
- An important device in time series analysis is to write higher–order models in first–order vector form.
- This simplifies things greatly for the ARMA (p, q) model and is also useful in other contexts, such as the derivation of forecasts.
- Write the ARMA(p, q) process as

$$X_t = AX_{t-1} + U_t, \tag{69}$$

where[6] $X_t = (Y_t, Y_{t-1}, \ldots, Y_{t-p+1}, \epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-q+1})$,

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}_{(p+q)\times(p+q)}, \quad U_t = \begin{pmatrix} \epsilon_t \\ 0_{(p-1)\times 1} \\ \epsilon_t \\ 0_{(q-1)\times 1} \end{pmatrix} \tag{70}$$

[6] For processes with nonzero mean μ, replace all the Y_t by $Y_t - \mu$, Y_{t-1} by $Y_{t-1} - \mu$ and so on.

$$A_{11} = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{p\times p}, \quad A_{12} = \begin{pmatrix} \theta_1 & \theta_2 & \cdots & \theta_q \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{p\times q},$$

$$A_{21} = \mathbf{0}_{q\times p}, \quad A_{22} = \begin{pmatrix} \mathbf{0}_{1\times(q-1)} & 0 \\ \boldsymbol{I}_{q-1} & \mathbf{0}_{(q-1)\times 1} \end{pmatrix}_{q\times q}, \tag{71}$$

where $\boldsymbol{I}_{q-1}$ is the identity matrix of dimension $q-1$.

- For a pure AR(p) process, $X_t = (Y_t, Y_{t-1}, \ldots, Y_{t-p+1})'$, and $A = A_{11}$.

- Note that, by block triangularity, the characteristic equation of matrix A is

$$\det(\lambda I_{p+q} - A) = \det(\lambda I_p - A_{11}) \det(\lambda I_q - A_{22}) = \lambda^q \det(\lambda I_p - A_{11}), \tag{72}$$

i.e., the nonzero eigenvalues of A are the eigenvalues of A_{11}.

- It turns out that the eigenvalues of matrix A_{11} are identical to the roots of the reverse characteristic equation[7]

$$\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0, \tag{74}$$

so the stationarity (stability) condition can equivalently be phrased in terms of the eigenvalues of matrix A_{11} (which have to be smaller than unity in magnitude).

- The powers of matrix A go to zero geometrically if and only if the maximal eigenvalue of A is smaller than one in magnitude.

- Consider the AR(p) process

$$X_t = A_{11} X_{t-1} + U_t, \tag{75}$$

where $U_t = (\epsilon_t, 0, \ldots, 0)'$. Iterating this, we can obtain the MA(∞)

[7] E.g., for $p = 2$,

$$A_{11} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}, \quad \det \begin{pmatrix} \lambda - \phi_1 & -\phi_2 \\ -1 & \lambda \end{pmatrix} = \lambda^2 - \phi_1 \lambda - \phi_2. \tag{73}$$

representation

$$
\begin{aligned}
X_t &= A_{11}^2 X_{t-2} + A_{11}U_{t-1} + U_t = A_{11}^3 X_{t-3} + A_{11}^2 U_{t-2} + A_{11}U_{t-1} + U_t \\
&= A_{11}^\tau X_{t-\tau} + \sum_{i=0}^{\tau-1} A_{11}^i U_{t-i} \\
&= \sum_{i=0}^{\infty} A_{11}^i U_{t-i},
\end{aligned}
$$

and from the structure of U_t it follows that the coefficient of ϵ_{t-i} in the MA(∞) representation of Y_t (i.e., θ_i) is just the top left element of the matrix A_{11}^i.

- A similar argument can be made for mixed ARMA(p, q) processes.

- For an $m \times n$ matrix A and an $p \times q$ matrix B, this is defined as the $mp \times nq$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}.$$

- Important rule in time series analysis:

$$\text{vec}(ABC) = (C' \otimes A)\text{vec}(B),$$

where the vec operator stacks the elements of an $m \times n$ matrix A columnwise into an mn column vector, e.g.,

$$\text{vec}\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 2 & 1 \end{pmatrix} = (1, 2, 3, 4, 5, 2, 7, 8, 1)'. \qquad (76)$$

- From (69),

$$X_t X_t' = A X_{t-1} X_{t-1}' A' + U_t U_t' + A X_{t-1} U_t' + U_t X_{t-1}' A'. \tag{77}$$

Note that $\mathsf{E}(X_{t-1}U_t') = 0$, and so, by stationarity, $(E(X_t X_t') = E(X_{t-1}X_{t-1}'))$,

$$\mathsf{E}(X_t X_t') = A\mathsf{E}(X_t X_t)A' + \mathsf{E}(U_t U_t'), \tag{78}$$

where the covariance matrix of U_t,

$$\Sigma_U := \mathsf{E}(U_t U_t') = \sigma^2 \begin{pmatrix} 1 \\ 0_{(p-1)\times 1} \\ 1 \\ 0_{(q-1)\times 1} \end{pmatrix} \begin{pmatrix} 1 \\ 0_{(p-1)\times 1} \\ 1 \\ 0_{(q-1)\times 1} \end{pmatrix}'. \tag{79}$$

Vectorizing, and solving,

$$\text{vec}(\mathsf{E}(X_t X_t')) = (\boldsymbol{I}_{(p+q)\times(p+q)} - A \otimes A)^{-1} \Sigma_U, \tag{80}$$

and the top left element of $\mathsf{E}(X_t X_t')$ is $\mathsf{E}(Y_t^2)$.

- $\mathsf{E}(X_t X_t')$ also has the first $p-1$ cross products $\mathsf{E}(Y_t Y_{t-1}), \ldots, \mathsf{E}(Y_t Y_{t-p+1})$.

- All autocorrelations can be calculated by the Yule–Walker equations, for $\tau \geq 1$,

$$E(X_t X_{t-\tau}') = AE(X_{t-1} X_{t-\tau}') + \underbrace{E(U_t X_{t-\tau}')}_{=0},$$

so

$$\begin{aligned}
\mathsf{E}(X_t X_{t-1}') &= A\mathsf{E}(X_t X_t') \\
\mathsf{E}(X_t X_{t-2}') &= A \underbrace{\mathsf{E}(X_{t-1} X_{t-2}')}_{=\mathsf{E}(X_t X_{t-1}')} = A^2 \mathsf{E}(X_t X_t') \\
&\vdots \\
\mathsf{E}(X_t X_{t-\tau}') &= A^\tau \mathsf{E}(X_t X_t'),
\end{aligned}$$

and $\mathsf{E}(X_t X_t')$ has been calculated above.

- **For an ARMA(p, q) process, both the ACF and the PACF gradually die out (i.e., do not cut off)**.

Invertibility

- Consider the MA(1) process $Y_t = \theta\epsilon_{t-1} + \epsilon_t$, $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$, with lag–one ACF
$$\rho(1) = \frac{\theta}{1+\theta^2}.$$

- Now consider the MA(1) process $Y_t = \tilde{\theta}\tilde{\epsilon}_{t-1} + \tilde{\epsilon}_t$, $\tilde{\epsilon}_t \sim \mathrm{N}(0, \tilde{\sigma}^2)$.

- Assume

$$\begin{aligned} \tilde{\theta} &= \theta^{-1} \\ \tilde{\sigma}^2 &= \theta^2\sigma^2. \end{aligned}$$

- Then for the $\tilde{Y}_t$–process

$$\begin{aligned} \tilde{\rho}(1) &= \frac{\tilde{\theta}}{1+\tilde{\theta}^2} = \frac{\theta^{-1}}{1+\theta^{-2}} = \frac{\theta}{1+\theta^2}, \\ \tilde{\gamma}(0) &= (1+\tilde{\theta}^2)\tilde{\sigma}^2 = \left(1+\frac{1}{\theta^2}\right)\theta^2\sigma^2 = (1+\theta^2)\sigma^2, \end{aligned}$$

i.e., both processes have the same moments, so there is a lack of identification.

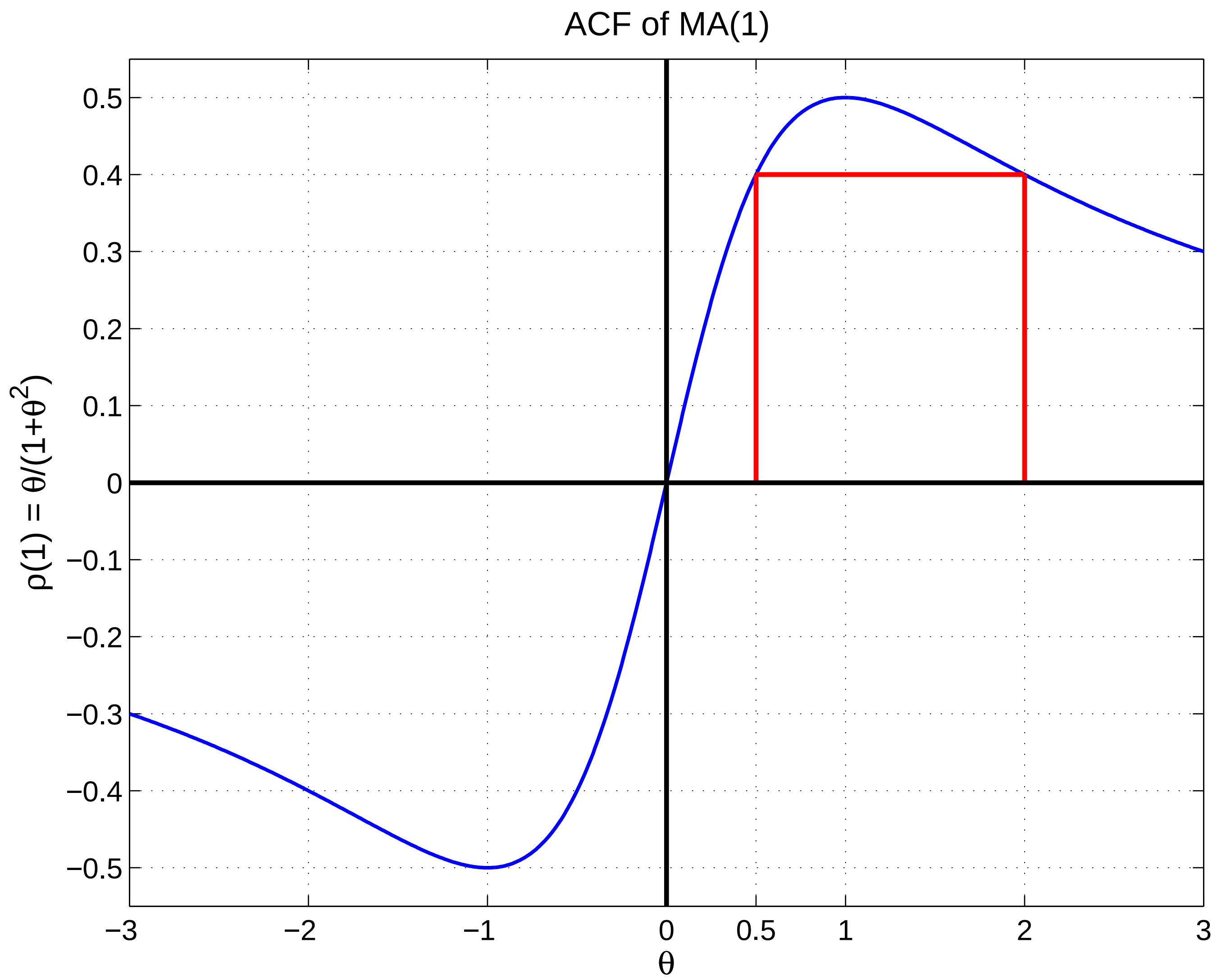
- To avoid such problems, we focus on models with $|\theta| < 1$, or, more generally, on models where the MA polynomial $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ has all roots outside the unit circle.

- The MA polynomial can then be inverted (thus the term **invertibility**), and the ARMA(p, q) process written as an AR(∞).

- E.g., for the MA(1),[8]

$$\frac{Y_t}{1+\theta L} = Y_t - \theta Y_{t-1} + \theta^2 Y_{t-2} - \cdots = \epsilon_t.$$

[8] For $|\theta| < 1$, $(1+\theta L)^{-1} = \sum_{i=0}^{\infty} (-\theta)^i L^i = 1 - \theta L + \theta^2 L^2 - \cdots$


ACF of MA(1)
ρ(1) = θ/(1+θ^2)
θ
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
−0.5
−3
−2
−1
0
0.5
1
2
3