Autocovariance-Stationary ARMA Models

DENG Yanfei
https://idengyf.github.io/

April 9, 2024

1 Exercises and Questions

1. Exercises:
Enders (2015, ch.1: E7, ch.2: E1)
2. Qusetions (learning objectives):
1) Why $\phi \neq 1$?
2) Lag Operators vs. Forward Operators ($LF = 1$)
3) Dynamic Multipliers vs. Impulse Responses
4) $c_i =$?

2 White Noise and Expectations

The basic building block of discrete stochastic time-series models is the white noise process:

$$y_t = \epsilon_t, \quad \epsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\epsilon^2)$$

i.e., y_t normal and independent over time. However, time series are typically not iid (e.g., if GDP today is unusually high, GDP tomorrow is also likely to be unusually high.)

Time-series consists of interesting parametric models for the **joint distribution** of $\{y_t\}$. The models impose structure, which we must evaluate to see if it captures the features we think are present in the data (stylized facts). In turn, they reduce the estimation problem to the estimation of a few parameters of the time-series model.

$$\text{A white noise process}\begin{cases}\mathbb{E}\epsilon_t=\mathbb{E}(\epsilon_t|\epsilon_{t-1},\epsilon_{t-2},\ldots)=\mathbb{E}(\epsilon_t|\text{all information at } t-1)=\mathbb{E}(\epsilon_{t-1})=0\\ \mathbb{E}\epsilon_t^2=\text{var}(\epsilon_t)=\text{var}(\epsilon_t|\epsilon_{t-1},\epsilon_{t-2},\ldots)=\text{var}(\epsilon_t|\text{all information at } t-1)=\mathbb{E}\epsilon_{t-1}^2=\sigma_\epsilon^2\\ \mathbb{E}(\epsilon_t\epsilon_\tau)=\mathbb{E}(\epsilon_t\epsilon_{t-j})=\mathbb{E}(\epsilon_{t-j}\epsilon_{t-j-s})=\text{cov}(\epsilon_{t-j},\epsilon_{t-j-s})=0,\quad \text{for } j\neq 0 \text{ or } t\neq\tau\ \leftarrow\perp\end{cases}$$

$$\text{An independent white noise process}\begin{cases}\mathbb{E}\epsilon_t=0\\ \mathbb{E}\epsilon_t^2=\sigma_\epsilon^2\\ \mathbb{E}(\epsilon_t\epsilon_\tau)=0,\quad \text{for } t\neq\tau\quad \text{cov}(\epsilon_t,\epsilon_\tau)=0 \overset{\text{uncorrelation}}{\Longleftrightarrow} \mathbb{E}(\epsilon_t\epsilon_\tau)=\mathbb{E}\epsilon_t\mathbb{E}\epsilon_\tau\\ \text{Pro}(\epsilon_t,\epsilon_\tau)=\text{Pro1}(\epsilon_t)\text{Pro2}(\epsilon_\tau)\text{ independent}\end{cases}$$

$$\text{The Gaussian white noise process}\begin{cases}\mathbb{E}\epsilon_t=0\\ \mathbb{E}\epsilon_t^2=\sigma_\epsilon^2\\ \mathbb{E}(\epsilon_t\epsilon_\tau)=0,\quad \text{for } t\neq\tau\\ \text{Pro}(\epsilon_t,\epsilon_\tau)=\text{Pro1}(\epsilon_t)\text{Pro2}(\epsilon_\tau)\\ \epsilon_t\sim(0,\sigma_\epsilon^2)\text{ normal distribution}\end{cases}$$

By itself, ϵ_t is a pretty boring process since it does not capture the interesting property of persistence that motivates the study of time series. Most of the time we will study a class of models created by taking linear combinations of ϵ_t:

$$\begin{cases}y_t=\epsilon_t \quad\leftarrow \text{white noise}\\ y_t=\epsilon_t+\theta\epsilon_{t-1}\quad\leftarrow \text{MA(1)}\\ y_t=\epsilon_t+\theta_1\epsilon_{t-1}+\cdots+\theta_q\epsilon_{t-q}\overset{\theta_0\equiv 1}{=}\sum_{i=0}^q\theta_i\epsilon_{t-i}\quad\leftarrow\text{MA(q)}\\ y_t=c+\phi y_{t-1}+\epsilon_t\quad\leftarrow\text{AR(1)}\rightarrow\text{if } c=(1-\phi)\bar{y}\Rightarrow(y_t-\bar{y})=\phi(y_{t-1}-\bar{y})+\epsilon_t\\ y_t=c+\phi_1y_{t-1}+\phi_2y_{t-2}+\cdots+\phi_py_{t-p}+\epsilon_t=c+\sum_{i=1}^p\phi_iy_{t-i}+\epsilon_t\quad\leftarrow\text{AR(p)}\\ y_t=c+\phi_1y_{t-1}+\phi_2y_{t-2}+\cdots+\phi_py_{t-p}+\epsilon_t+\theta_1\epsilon_{t-1}+\cdots+\theta_q\epsilon_{t-q}\quad\leftarrow\text{ARMA(p, q)}\end{cases}$$

$$\underbrace{A(L)y_t=\epsilon_t}_{\text{AR}}\left\{\begin{matrix}\text{AR(1)}:\ (1-\phi L)y_t=\epsilon_t\\ \text{AR(p)}:\ (1+\phi_1L+\cdots+\phi_pL^p)y_t=\epsilon_t\end{matrix}\right\}\overbrace{A(L)y_t=B(L)\epsilon_t}^{\text{ARMA}}\left\{\begin{matrix}\text{MA(1)}:\ y_t=(1+\theta L)\epsilon_t\\ \text{MA(q)}:\ y_t=(1+\theta_1L+\cdots+\theta_qL^q)\epsilon_t\end{matrix}\right\}\underbrace{y_t=B(L)\epsilon_t}_{\text{MA}}.$$

AR forms are the easiest to estimate, since the OLS assumptions still apply;
MA forms are the easiest to find variances and covariances.
1) AR(1) to MA(∞): $y_t=\phi y_{t-1}+\epsilon_t \xRightarrow{\text{by iteration or lag-operators}} y_t=\sum_{i=0}^\infty\phi^i\epsilon_{t-i}$;
2) AR(p)[1] to MA(∞);
3) MA(q) to AR(∞): $y_t=B(L)\epsilon_t \xRightarrow{\text{it's invertible}} B(L)^{-1}y_t=\epsilon_t$;
4) ARMA(p, q) $\xrightarrow{\text{the solution}}$ MA(∞).

3 Stationarity and Ergodicity

3.1 Strong stationarity (SS)

If the joint probability distribution (various moments including the first- and second- moment etc.) function of $y_{t-h},\ldots,y_t,\ldots,y_{t+h}$ is independent of t for all h, then the process $\{y_t\}$ is strongly/strictly stationary. SS is useful, e.g., a nonlinear function of a SS variable is SS.

[1]It can be expressed as a vector AR(1)

3.2 Weak stationarity (WS)

The unconditional covariances vs. The conditional covariance

Weak stationarity is often misunderstood. The definition merely requires that the unconditional covariances are not a function of time.

Eg1. If the conditional covariances of a series vary over time, as in ARCH models, the series can still be stationary.

Eg2. "A unit root" is one form of nonstationarity, but there are lots of others.

Eg3. If a series has breaks in trends, or if a series changed over time, then it may be "nonstationary" (when the trend break or structural shift occurs at one point in time no matter how history comes out), but it may be not (when breaks or shifts occur stochasticly→the unconditional covariances will no longer have a time index, then it still be weakly stationary).

AR lag polynomials are invertible & MA lag polynomials are square summable.

(1) If and only if the impluse-response function ($\sum_{h=0}^{\infty} \beta^h \frac{\partial y_{t+h}}{\partial \epsilon_t} = \sum_{h=0}^{\infty} \beta^h \mathbf{F}_{11}^h$) eventually decays exponentially.

⇕

(2) If the eigenvalues of $\mathbf{F}$ in AR(1) all lie inside the unit circle ($|\mathbf{F} - \boldsymbol{\lambda}\mathbf{I}| = \mathbf{0}$), or if all roots of the lag polynomial $A(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ lie outside the unit circle (i.e., the lag polynomial is **invertible**) then the original AR(p) model turns out to be convariance-stationary.[2]

⇕

(3) Weak stationarity does not require the MA polynomial $B(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ to be invertible.

⇕

(4) If neither the mean nor the variance depend on time t (i.e., they are finite[3]) and the autocovariances $\mathbb{E}(y_t y_{t-h})$ depend only on h but not t, then the stochastic process is said to be **covariance-stationary**[4](CovS, weakly stationary/2nd-order stationary/wide-sense stationary):

$$\begin{aligned}
\mathbb{E}y_t &= \mathbb{E}y_{t-h} = \mu, \\
\text{var}(y_t) &\equiv \mathbb{E}(y_t-\mu)^2 = \mathbb{E}(y_{t-h}-\mu)^2 \equiv \text{var}(y_{t-h}) = \sigma_\epsilon^2 \xleftarrow{\mu=c} \quad y_t = c + \epsilon_t, \\
\text{cov}(y_t, y_{t-h}) &\equiv \mathbb{E}[(y_t-\mu)(y_{t-h}-\mu)] = \mathbb{E}[(y_{t-j}-\mu)(y_{t-j-h}-\mu)] \equiv \text{cov}(y_{t-j}, y_{t-j-h}) = \gamma_h \\
\text{autocorrelation}(y_t, y_{t-h}) &\equiv \rho_h \equiv \frac{\gamma_h}{\gamma_0} = \frac{\mathbb{E}(y_t-\mu)(y_{t-h}-\mu)}{\mathbb{E}(y_t-\mu)(y_{t-0}-\mu)} = \frac{\text{cov}(y_t, y_{t-h})}{\text{var}(y_t)}.
\end{aligned}$$

---

[2] For a general ARMA(p, q) model, write it using lag operators so that

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) y_t = c + \sum_{i=0}^{q} \theta_i \epsilon_{t-i} \quad \Rightarrow \quad y_t^p = \frac{c + \sum_{i=0}^{q} \theta_i \epsilon_{t-i}}{1 - \sum_{i=1}^{p} \phi_i L^i}.$$

Notice that the particular solution is **convergent** so that the linear stochastic DE is **stable** (the stability condition is that the roots of $1 - \sum \phi_i L^i$ must lie outside the unit circle or the eigenvalues of $\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \cdots - \phi_p = 0$ all lie inside the unit circle).

[3] Note that a strongly stationary process need not have a finite mean and/or variance

[4] Autocovariance is the covariance between y_t and its own lags; Cross-covariance refers to the covariance between one series and another.

3.3 The relationship between strong- and weak- stationarity

$$\begin{cases} \text{1. SS does not imply WS but SS } + \mathbb{E}y_t, \mathbb{E}y_t^2 < \infty \Rightarrow \text{WS;} \\ \text{2. WS does not imply SS but WS + normality} \Rightarrow \text{SS.} \end{cases}$$

3.4 Weak stationarity restrictions

1) MA(∞) MA(q) is a special case

$$y_t = \sum_{i=0}^{\infty} \theta_i \epsilon_{t-i} \begin{cases} \{\epsilon_t\} \text{ is a white-noise process} \\ \{y_t\} \text{ is } \textbf{not} \text{ a white-noise process} \begin{cases} \mathbb{E}y_t = 0 \quad (\checkmark) \\ \text{var}(y_t) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i^2 \quad (\checkmark) \\ \text{cov}(y_t, y_{t-h}) = (\theta_h + \theta_1\theta_{h+1} + \theta_2\theta_{h+2} + \cdots)\sigma_\epsilon^2 \neq 0 \quad (\times) \end{cases} \\ \{y_t\} \text{ is CovS} \begin{cases} \mathbb{E}y_t = \mathbb{E}y_{t-h} = \mu = 0 \quad (\checkmark) \\ \text{var}(y_t) = \text{var}(y_{t-h}) = \sigma_\epsilon^2 \sum\limits_{i=0}^{\infty} \theta_i^2 \begin{cases} \sum \text{ finite is required for } i = 1, \dots, q \quad (\checkmark) \\ \sum \text{ finite is required for } i = 1, \dots, \infty \quad (\checkmark) \end{cases} \\ \text{cov}(y_t, y_{t-h}) = \mathbb{E}[(y_t - 0)(y_{t-h} - 0)] = \mathbb{E}(y_t y_{t-h}) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i \theta_{h+i} \begin{cases} \sum \ (\checkmark) \\ \sum \ (\checkmark) \end{cases} \end{cases} \end{cases}$$

The iff conditions for any MA process to be covariance stationary are for the $\sum \theta_i^2$ and $\sum \theta_i \theta_{h+i}$ to be finite.

2) AR(1)[5]

(1) with an initial condition y_0

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t,$$

$$\Rightarrow y_t = c \sum_{i=0}^{t-1} \phi^i + y_0 \phi^t + \sum_{i=0}^{t-1} \phi^i \epsilon_{t-i} \quad \leftarrow \text{with an initial condition}$$

$$\Rightarrow \mathbb{E}y_t = c \sum_{i=0}^{t-1} \phi^i + y_0 \phi^t \xLeftrightarrow[\text{the sequence cannot be stationary}]{\mathbb{E}y_t \neq \mathbb{E}y_{t+h} \text{(both means are time dependent)}} \mathbb{E}y_{t+h} = c \sum_{i=0}^{t+h-1} \phi^i + y_0 \phi^{t+h};$$

$$\overset{\text{or}}{\Rightarrow} \lim y_t \overset{t\to\infty}{=} \frac{c}{1-\phi} + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} \quad \left(\text{vs.} \quad y_t = A\phi^t + \frac{c}{1-\phi} + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} \text{ without an initial condition}\right)$$

$$\Rightarrow \mathbb{E} \lim y_t = \frac{c}{1-\phi} \xLeftrightarrow[\text{one of stationarity conditions}]{\mathbb{E}y_t = \mathbb{E}y_{t+h} \text{(both means are finite and time independent)}} \mathbb{E} \lim y_{t+h} = \frac{c}{1-\phi},$$

$$\Rightarrow \text{var}(y_t) = \mathbb{E}(y_t - \mu)^2 = \mathbb{E}(\lim y_t - \mathbb{E} \lim y_t)^2 = \mathbb{E}\left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}\right)^2 = (1 + \phi^2 + \phi^4 + \cdots)\sigma_\epsilon^2 = \frac{\sigma_\epsilon^2}{1-\phi^2},$$

$$\Rightarrow \text{cov}(y_t, y_{t-h}) = \mathbb{E}[(y_t - \mu)(y_{t-h} - \mu)] = \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-h-i}\right)\right] = (1 + \phi^2 + \cdots)\phi^h \sigma_\epsilon^2 = \frac{\phi^h \sigma_\epsilon^2}{1-\phi^2}.$$

[5] Notice that the solution of $\{y_t\}$ converges but $\{y_t\}$ may not be stationary, except $t \to \infty$

(2) without the inital value y_0,

$$y_t = A\phi^t + \frac{c}{1-\phi} + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i},$$
$$\mathbb{E}y_t = A\phi^t + \frac{c}{1-\phi} = f(t).$$

It's clear that $\{y_t\}$ cannot be stationary unless $y_t^h = A\phi^t = 0 \to \begin{cases} A = 0 \text{ (i.e.,}\{y_t\} \text{ always in equilibrium)} \\ |\phi| < 1 \& t \to \infty \end{cases}$

3) AR(p)

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \begin{cases} \text{the homogenous equation} y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p} = 0; \\ y_t^p = \frac{c}{1-\sum_{i=1}^{p}\phi_i} + \sum_{i=0}^{\infty} x_i \epsilon_{t-i} \begin{cases} x_0 - 1 = 0 \Rightarrow x_0 = 1, \\ x_1 - ? = 0, \\ \vdots \\ x_i - \phi_1 x_{i-1} - \phi_2 x_{i-2} - \cdots - \phi_p x_{i-p} = 0. \end{cases} \\ \Downarrow \\ \begin{cases} \mathbb{E}y_t^p = \mathbb{E}y_{t-h}^p = \underbrace{\frac{c}{1-\sum\phi_i > 0}}_{} \overline{\because |\phi_i| < 1 \text{ is a necessary condition for } \lambda_i < 1} \therefore \text{finite \& time invariant}, \\ \text{var}(y_t^p) = \text{var}(y_{t-h}^p) = \mathbb{E}(\sum_{i=0}^{\infty} x_i \epsilon_{t-i})^2 = \sigma_\epsilon^2 \sum x_i^2 \begin{cases} \sum \text{is finite for } i = 1, \ldots, p \ (\checkmark) \\ \sum \text{is finite for } i = 1, \ldots, \infty \text{ if } |x_i| < 1 \ (\end{cases} \\ \text{cov}(y_t^p, y_{t-h}^p) = \sigma_\epsilon^2 (x_h + x_1 x_{h+1} + x_2 x_{h+2} + \cdots + x_p x_{h+p}) = \sigma_\epsilon^2 \sum_{i=0}^{p} x_i x_{h+i} \begin{cases} \sum \ (\checkmark) \\ \sum \ (\checkmark) \end{cases} \end{cases} \end{cases}$$

4) ARMA(2, 1) (from the previous analysis, we know that $y_t^h = 0$)

$$\begin{cases} y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta\epsilon_{t-1} \xRightarrow{\text{undetermined coefficients}} y_t^p = k_0 + k_1 t + \sum_{i=0}^{\infty} x_i \epsilon_{t-i}; \\ y_t = 0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta\epsilon_{t-1} \xRightarrow{\text{undetermined coefficients}} y_t^p = \sum_{i=0}^{\infty} x_i \epsilon_{t-i} \begin{cases} x_0 = 1, \\ x_1 = \phi_1 x_0 + \theta = \phi_1 + \theta, \\ x_2 = \phi_1 x_1 + \phi_2 x_0, \\ x_3 = \phi_1 x_2 + \phi_2 x_1, \\ \vdots \\ x_i = \phi_1 x_{i-1} + \phi_2 x_{i-2}. \end{cases} \end{cases}$$

Q: If the characteristic roots of the ARMA(2, 1) process are within the unit circle, the $\{x_i\}$ must constitute a convergent sequence?

A: Notice that $x_i = \phi_1 x_{i-1} + \phi_2 x_{i-2}$ for all $i \geq 2$, this is the same homogeneous form with the original process, i.e., $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2}$.

$$\mathbb{E}y_t = \mathbb{E}y_t^h + \mathbb{E}y_t^p = 0 + \mathbb{E}y_t^p = 0 + 0 = 0 = \mathbb{E}y_{t-h} \quad \forall t \& h;$$
$$\text{var}(y_t) = \mathbb{E}(x_0\epsilon_t + x_1\epsilon_{t-1} + x_2\epsilon_{t-2} + \cdots)^2 = \sigma_\epsilon^2 \sum_{i=0}^{\infty} x_i^2 = \text{var}(y_{t-h});$$
$$\text{cov}(y_t, y_{t-h}) = \sigma_\epsilon^2 (x_h + x_{h+1}x_1 + x_{h+2}x_2 + \cdots) \xLeftarrow{\text{cov}(y_t, y_{t-1}), \text{cov}(y_t, y_{t-2}), \ldots} .$$

Conversely, if the characteristic roots do not lie within the unit circle, the $\{x_i\}$ sequence will not be convergent. As such, the $\{y_t\}$ sequence cannot be convergent.

5) ARMA(p, q)

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=0}^{q} \theta_i \epsilon_{t-i} \begin{cases} y_t^h = \begin{cases} \sum_{i=1}^{p} A_i \lambda_i^t & \text{distint and real roots;} \\ \lambda^t \sum_{i=1}^{m} A_i t^{i-1} + \sum_{i=m+1}^{p} A_i \lambda_i^t & \text{m repeated real roots.} \end{cases} \\ y_t^p = \frac{c}{1-\sum_{i=1}^{p} \phi_i} + \frac{\epsilon_t}{1-\sum_{i=1}^{p} \phi_i L^i} + \frac{\theta_1 \epsilon_{t-1}}{1-\sum_{i=1}^{p} \phi_i L^i} + \frac{\theta_2 \epsilon_{t-2}}{1-\sum_{i=1}^{p} \phi_i L^i} + \cdots \begin{cases} \mathbb{E} y_t^p & = (\checkmark) \\ \text{var}(y_t^p) & = (\checkmark) \\ \text{cov}(y_t^p, y_{t-h}^p) & = (\checkmark) \end{cases} \end{cases}$$

Note that each of expressions on the RHS of the y_t^p is stationary as long as the roots of $1 - \sum \phi_i L^i$ are outside the unit circle and given the MA(q) is stationary, only the roots of the autoregressive portion of MA(q) determine whether the $\{y_t\}$ sequence is stationary.

3.5 Ergodicity

For many applications, **stationarity** and **ergodicity** turn out to amount to the same requirements. For purposes of clarifying the concepts of stationarity and ergodicity, however, it may be helpful to consider an example of a process that is stationary but **not** ergodic, $\cdots$ cf. Hamilton (1994, ch.3.1)

4 The Autocorrelation Function (ACF) or Correlogram

4.1 Autocovariances/autocorrelations of ARMA(p, q)

The autocovariances and autocorrelations serve as useful tools to identifying and estimating time-series models.

1) Gaussian white noise

$$\begin{aligned} y_t &= \epsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\epsilon^2), \\ \mu \equiv \mathbb{E} y_t &= \mathbb{E} \epsilon_t = 0, \\ \gamma_0 \equiv \text{var}(y_t) &= \text{var}(\epsilon_t) = \sigma_\epsilon^2, \\ \gamma_h \equiv \text{cov}(y_t, y_{t-h}) &= \text{cov}(\epsilon_t, \epsilon_{t-h}) = 0, \\ \rho_h = \frac{\gamma_h}{\gamma_0} \equiv \text{corr}(y_t, y_{t-h}) &= \frac{\text{cov}(\epsilon_t, \epsilon_{t-h})}{\text{var}(\epsilon_t)} = \begin{cases} 1, & h = 0, \\ 0, & h \neq 0. \end{cases} \end{aligned}$$

2) MA(1)

$$
\begin{aligned}
y_t &= \epsilon_t + \theta\epsilon_{t-1},\\
\mu &= \mathbb{E}y_t = 0,\\
\gamma_0 &= \text{var}(y_t) = \mathbb{E}(y_t y_t) = \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_t + \theta\epsilon_{t-1})] = (1+\theta^2)\sigma_\epsilon^2,\\
\gamma_1 &= \text{cov}(y_t, y_{t-1}) = \mathbb{E}(y_t y_{t-1}) = \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2})] = \theta\sigma_\epsilon^2,\\
\gamma_2 &= \text{cov}(y_t, y_{t-2}) = \mathbb{E}(y_t y_{t-2}) = \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-2} + \theta\epsilon_{t-3})] = 0,\\
&\vdots\\
\gamma_h &= \text{cov}(y_t, y_{t-h}) = \mathbb{E}(y_t y_{t-h}) = \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-h} + \theta\epsilon_{t-h-1})] = 0.\\
&\Rightarrow \begin{cases}
\rho_0 = \text{corr}(y_t, y_t) = \frac{\gamma_0}{\gamma_0} = 1,\\
\rho_1 = \text{corr}(y_t, y_{t-1}) = \frac{\gamma_1}{\gamma_0} = \frac{\theta}{1+\theta^2},\\
\rho_2 = \text{corr}(y_t, y_{t-2}) = \frac{\gamma_2}{\gamma_0} = 0,\\
\vdots\\
\rho_h = \text{corr}(y_t, y_{t-h}) = \frac{\gamma_h}{\gamma_0} = 0.
\end{cases}
\end{aligned}
$$

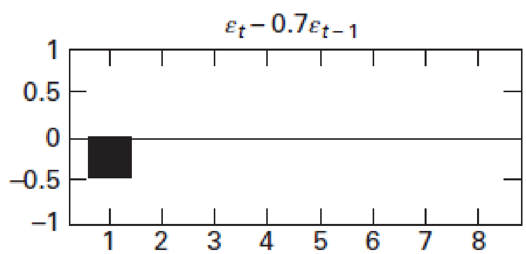

Figure 1: Theoretical ACF Patterns of MA(1) with $-1 < \theta < 0$ [1]

2) MA(2)

$$
\begin{aligned}
y_t &= \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2},\\
\mu \equiv \mathbb{E}y_t &= \mathbb{E}\epsilon_t + \theta_1\mathbb{E}\epsilon_{t-1} + \theta_2\mathbb{E}\epsilon_{t-2} = 0,\\
\gamma_0 \equiv \text{var}(y_t) &= \mathbb{E}[(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2})(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2})] = (1+\theta_1^2+\theta_2^2)\sigma_\epsilon^2,\\
\gamma_1 \equiv \text{cov}(y_t, y_{t-1}) &= \mathbb{E}[(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2})(\epsilon_{t-1} + \theta_1\epsilon_{t-2} + \theta_2\epsilon_{t-3})] = (\theta_1+\theta_1\theta_2)\sigma_\epsilon^2,\\
\gamma_2 \equiv \text{cov}(y_t, y_{t-2}) &= \mathbb{E}[(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2})(\epsilon_{t-2} + \theta_1\epsilon_{t-3} + \theta_2\epsilon_{t-4})] = \theta_2\sigma_\epsilon^2,\\
\gamma_3 \equiv \text{cov}(y_t, y_{t-3}) &= \mathbb{E}[(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2})(\epsilon_{t-3} + \theta_1\epsilon_{t-4} + \theta_2\epsilon_{t-5})] = 0,\\
&\vdots\\
\gamma_h \equiv \text{corr}(y_t, y_{t-h}) &= 0 \quad \text{for } h \geq 3,\\
&\Rightarrow \begin{cases}
\rho_0 = \frac{\gamma_0}{\gamma_0} \equiv \text{corr}(y_t, y_t) = 1,\\
\rho_1 = \frac{\gamma_1}{\gamma_0} \equiv \text{corr}(y_t, y_{t-1}) = \frac{\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2},\\
\rho_2 = \frac{\gamma_2}{\gamma_0} \equiv \text{corr}(y_t, y_{t-2}) = \frac{\theta_2}{1+\theta_1^2+\theta_2^2},\\
\rho_3 = \frac{\gamma_3}{\gamma_0} \equiv \text{corr}(y_t, y_{t-2}) = 0,\\
\vdots\\
\rho_h = \frac{\gamma_h}{\gamma_0} \equiv \text{corr}(y_t, y_{t-h}) = 0 \quad \text{for } h \geq 3.
\end{cases}
\end{aligned}
$$

[1]Source: Enders (2015, p.61)

4) AR(1)

$$
\begin{aligned}
y_t &= c + \phi y_{t-1} + \epsilon_t, \\
\Rightarrow y_t &= \frac{c}{1-\phi} + \frac{\epsilon_t}{1-\phi L}, \\
&= \frac{c}{1-\phi} + (1 + \phi L + \phi^2 L^2 + \cdots)\epsilon_t, \quad |\phi| < 1 \\
&= \frac{c}{1-\phi} + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}, \\
\Rightarrow \mu = \mathbb{E} y_t &= \frac{c}{1-\phi},
\end{aligned}
$$

$$
\Rightarrow \gamma_0 = \mathrm{cov}(y_t, y_{t-0}) = \mathrm{var}(y_t) = \mathbb{E}(y_t - \mathbb{E}y_t)^2 = \left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}\right)^2 = \frac{\sigma_\epsilon^2}{1-\phi^2},
$$

$$
\Rightarrow \gamma_h = \mathrm{cov}(y_t, y_{t-h}) = \mathbb{E}[(y_t - \mu)(y_{t-h} - \mu)] = \left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-h-i}\right) = \frac{\phi^h}{1-\phi^2}\sigma_\epsilon^2.
$$

$$
\Rightarrow \begin{cases}
\rho_0 = \mathrm{corr}(y_t, y_t) = \frac{\gamma_0}{\gamma_0} = 1, \\
\rho_1 = \mathrm{corr}(y_t, y_{t-1}) = \frac{\gamma_1}{\gamma_0} = \phi, \\
\rho_2 = \mathrm{corr}(y_t, y_{t-2}) = \frac{\gamma_2}{\gamma_0} = \phi^2, \\
\vdots \\
\rho_h = \mathrm{corr}(y_t, y_{t-h}) = \frac{\gamma_h}{\gamma_0} = \phi^h.
\end{cases}
$$

Note that $\gamma_h = \gamma_{-h}$ for a scalar process. We will see that $\Gamma_h \neq \Gamma_{-h}$ for a vector process at Lec 3.

The plot of ρ_h against h should converge to 0 geometrically if the series is stationary (i.e.,$|\phi| < 1$).

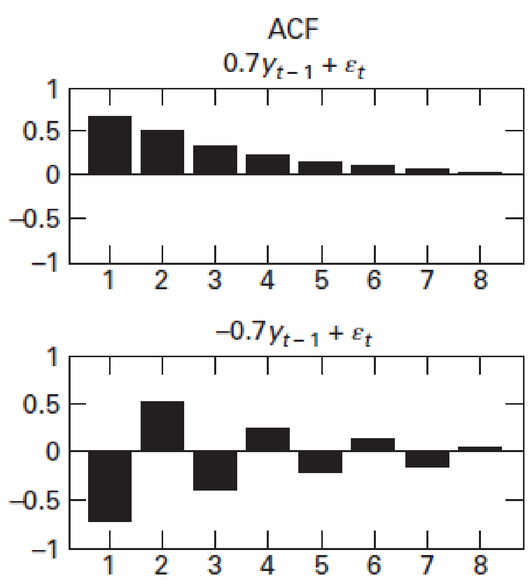

Figure 2: Theoretical ACF Patterns of AR(1) with $0 < \phi < 1$ and $-1 < \phi < 0$ [1]

[1] Source: Enders (2015, p.61)

5) AR(2)

$y_t = 0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \leftarrow$ omit c since it has no effect on the ACF and note that $\mathbb{E}(\epsilon_t y_{t-h}) = \begin{cases} \sigma_\epsilon^2 & \text{for } h = 0 \\ 0 & \text{for } h \neq 0 \end{cases}$.

$$\mu = \mathbb{E}y_t = \frac{c}{1 - \phi_1 L - \phi_2 L^2} = 0,$$
$$\gamma_0 = \mathbb{E}(y_t y_t) = \mathbb{E}[(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t) y_t] = \phi_1 \mathbb{E}(y_{t-1} y_t) + \phi_2 \mathbb{E}(y_{t-2} y_t) + \mathbb{E}(\epsilon_t y_t) = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\epsilon^2,$$
$$\gamma_1 = \mathbb{E}(y_t y_{t-1}) = \mathbb{E}[(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t) y_{t-1}] = \phi_1 \mathbb{E}(y_{t-1} y_{t-1}) + \phi_2 \mathbb{E}(y_{t-2} y_{t-1}) + \mathbb{E}(\epsilon_t y_{t-1}) = \phi_1 \gamma_0 + \phi_2 \gamma_1,$$
$$\gamma_2 = \mathbb{E}(y_t y_{t-2}) = \mathbb{E}[(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t) y_{t-2}] = \phi_1 \mathbb{E}(y_{t-1} y_{t-2}) + \phi_2 \mathbb{E}(y_{t-2} y_{t-2}) + \mathbb{E}(\epsilon_t y_{t-2}) = \phi_1 \gamma_1 + \phi_2 \gamma_0,$$
$$\vdots$$
$$\gamma_h = \underbrace{\mathbb{E}(y_t y_{t-h}) = \mathbb{E}[(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t) y_{t-h}] = \phi_1 \mathbb{E}(y_{t-1} y_{t-h}) + \phi_2 \mathbb{E}(y_{t-2} y_{t-h}) + \mathbb{E}(\epsilon_t y_{t-h}) = \phi_1 \gamma_{h-1} + \phi_2 \gamma_{h-2}}_{\text{The Yule-Walker equations}}.$$

$$\Rightarrow \begin{cases} \rho_0 = \text{corr}(y_t, y_t) = \frac{\gamma_0}{\gamma_0} = 1, \\ \rho_1 = \text{corr}(y_t, y_{t-1}) = \frac{\gamma_1}{\gamma_0} = \phi_1 \rho_0 + \phi_2 \rho_1 \quad \Rightarrow \quad \rho_1 = \frac{\phi_1}{1-\phi_2}, \\ \rho_2 = \text{corr}(y_t, y_{t-2}) = \frac{\gamma_2}{\gamma_0} = \phi_1 \rho_1 + \phi_2 \rho_0 \quad \Rightarrow \quad \rho_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2, \\ \vdots \\ \rho_h = \text{corr}(y_t, y_{t-h}) = \frac{\gamma_h}{\gamma_0} = \phi_1 \rho_{h-1} + \phi_2 \rho_{h-2} \leftarrow \text{a 2nd-order DE with two initial condtions } \rho_0 \text{ and } \rho_1. \end{cases}$$

Note that

$$\gamma_0 = \text{var}(y_t) = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\epsilon^2 \Rightarrow \gamma_0 = \phi_1(\rho_1 \gamma_0) + \phi_2(\rho_2 \gamma_0) + \sigma_\epsilon^2 \Rightarrow \gamma_0(1 - \phi_1 \rho_1 - \phi_2 \rho_2) = \sigma_\epsilon^2 \Rightarrow \gamma_0 = \cdots$$

Although ρ_h are cumbersome to derive, we can easily characterize their properties by resorting to the 2nd-order DE with initial values ρ_0 and ρ_1. Note that the stationarity condition for y_t necessitates that the characteristic roots of the 2nd-order DE lie inside the unit circle which let the ρ_h sequence will be convergent.

The properties of the various ρ_h follow directly from the homogeneous equation

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = 0.$$

If the discriminant is negative, i.e., $d \equiv \sqrt{\phi_1^2 + 4\phi_2} < 0$, the characteristic roots ($\lambda_1, \lambda_2 = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$) are imaginary so that the solution oscillates. $R = \sqrt{\left(\frac{\phi_1}{2}\right)^2 + \left(\frac{\sqrt{-(\phi_1^2 + 4\phi_2)}}{2}\right)^2} = \sqrt{-\phi_2}$, then, $|R| < 1$ is also a stable condition.

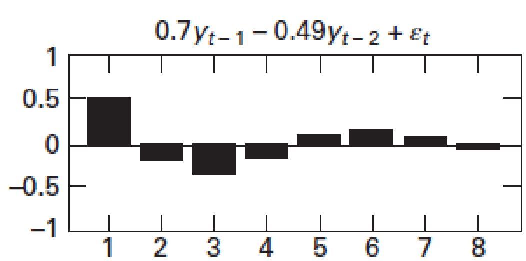

Figure 3: Theoretical ACF Patterns of AR(2) with $0 < \phi_1 < 1$ and $-1 < \phi_2 < 0$ [1]

[1] Source: Enders (2015, p.61)

6) ARMA(1, 1)

$$
\begin{aligned}
y_t &= 0 + \phi y_{t-1} + \epsilon_t + \theta\epsilon_{t-1},\\
\mu &= \mathbb{E}y_t = 0,\\
\gamma_0 &= \mathbb{E}(y_t y_t) = \mathrm{E}[(\phi y_{t-1} + \epsilon_t + \theta\epsilon_{t-1})y_t] = \phi\mathbb{E}(y_{t-1}y_t) + \mathbb{E}(\epsilon_t y_t) + \theta\mathbb{E}(\epsilon_{t-1}y_t) = \phi_1\gamma_1 + \sigma_\epsilon^2 + \theta(\phi+\theta)\sigma_\epsilon^2,\\
\gamma_1 &= \mathbb{E}(y_t y_{t-1}) = \mathrm{E}[(\phi y_{t-1} + \epsilon_t + \theta\epsilon_{t-1})y_{t-1}] = \phi\mathbb{E}(y_{t-1}y_{t-1}) + \mathbb{E}(\epsilon_t y_{t-1}) + \theta\mathbb{E}(\epsilon_{t-1}y_{t-1}) = \phi\gamma_0 + \theta\sigma_\epsilon^2,\\
\gamma_2 &= \mathbb{E}(y_t y_{t-2}) = \mathrm{E}[(\phi y_{t-1} + \epsilon_t + \theta\epsilon_{t-1})y_{t-2}] = \phi\mathbb{E}(y_{t-1}y_{t-2}) + \mathbb{E}(\epsilon_t y_{t-2}) + \theta\mathbb{E}(\epsilon_{t-1}y_{t-2}) = \phi\gamma_1,\\
&\vdots\\
\gamma_h &= \mathbb{E}(y_t y_{t-h}) = \mathrm{E}[(\phi y_{t-1} + \epsilon_t + \theta\epsilon_{t-1})y_{t-h}] = \phi\mathbb{E}(y_{t-1}y_{t-h}) + \mathbb{E}(\epsilon_t y_{t-h}) + \theta\mathbb{E}(\epsilon_{t-1}y_{t-h}) = \phi\gamma_{h-1}.
\end{aligned}
$$

$$
\Rightarrow \begin{cases}
\rho_0 = \mathrm{corr}(y_t, y_t) = \frac{\gamma_0}{\gamma_0} = 1,\\
\rho_1 = \mathrm{corr}(y_t, y_{t-1}) = \frac{\gamma_1}{\gamma_0} = f(\phi,\theta) \Leftarrow \begin{cases}\gamma_0 = \phi\gamma_1 + \sigma_\epsilon^2 + \theta(\phi+\theta)\sigma_\epsilon^2\\ \gamma_1 = \phi\gamma_0 + \theta\sigma_\epsilon^2\end{cases}\\
\rho_2 = \mathrm{corr}(y_t, y_{t-2}) = \frac{\gamma_2}{\gamma_0} = \frac{\phi\gamma_1}{\gamma_0} = \phi\rho_1\\
\vdots\\
\rho_h = \mathrm{corr}(y_t, y_{t-h}) = \frac{\gamma_h}{\gamma_0} = \frac{\phi\gamma_{h-1}}{\gamma_0} = \phi\rho_{h-1} \quad \text{with an initial value } \rho_1.
\end{cases}
$$

Thus, the ACF for an ARMA(1, 1) process is such that the magnitue of ρ_1 depends on both ϕ and θ. Beginning with ρ_1, the ACF of an ARMA(1, 1) process looks like that of the AR(1) process.

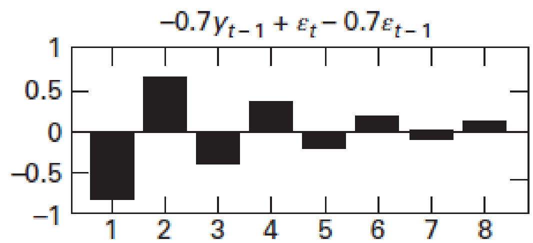

Figure 4: Theoretical ACF Patterns of ARMA(1, 1) with $-1 < \phi < 0$ and $-1 < \theta < 0$ [1]

7) AMRA(p, q)
Beginning after lag q, the values of the ρ_i will satisfy

$$\rho_i = \phi_1\rho_{i-1} + \phi_2\rho_{i-2} + \cdots + \phi_p\rho_{i-p}.$$

The previous p values can be treated as initial condtions that satisfy the Yule-Walker equations. For these lags, the shape of the ACF is determined by the characteristic equation.

4.2 Admissible autocorrelation function

Obviously, $|\rho_i| < 1$ is a necessary condition. But it's not sufficient for $\{\rho_i\}$ to be the autocorrelation function of an ARMA process.

We can find a stronger requirement than $|\rho_i| < 1$ under the extra condition that the variance of any random variable is positive (cf. Cochrane, 2005, pp.27-29).

[1]Source: Enders (2015, p.61)

5 PACF and Sample Autocorrelations

The partial autocorrelation function (PACF) and the sample autocorrelations of ARMA models (cf. Enders, 2015, pp.64-67).

The pth PACF is related to the coefficient on x_{t-p} of a regression of x_t on $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$. Thus for an AR(p), the (p+1)th and higher partial autocorrelations are 0. In fact, the PACF behaves in an exactly symmetrical fashion to the ACF: the PACF of an MA(q) is damped sines and exponentials after q (cf. Cochrane, 2005).

In an AR(1) process, the autocorrelation between y_t and y_{t-2}

$$
\begin{aligned}
y_t &= \phi y_{t-1} + \epsilon_t, \\
y_{t-1} &= \phi y_{t-2} + \epsilon_{t-1}, \\
&\vdots \\
\Rightarrow \rho_2 &= \text{corr}(y_t, y_{t-2}) = \text{corr}(y_t, y_{t-1})\text{corr}(y_{t-1}, y_{t-2}) = \rho_1^2.
\end{aligned}
$$

However, the partial autocorrelation between y_t and y_{t-2} eliminates the effects of the intervening values y_{t-1}; the partial autocorrelation between y_t and y_{t-3} eliminates the effects of the intervening values y_{t-1}, y_{t-2}, and so on.

One can form the partial autocorrelations from the autocorrelations

$$
\begin{aligned}
\rho_1^p &= \rho_1, \\
\rho_2^p &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \\
&\vdots \\
\rho_h^p &= \frac{\rho_h - \sum_{i=1}^{h-1} \phi_{h-1,i}\rho_{h-i}}{1 - \sum_{i=1}^{h-1} \phi_{h-1,i}\rho_i}, \quad h = 3, 4, 5, \ldots.
\end{aligned}
$$

For stationary processes, the key points to note are the follwing: (see Enders 2015, p.66).

Suppose that a researcher collected **sample data** and plotted the ACF and PACF. If the actual patterns compared favorably to the theoretical patterns, the researcher might try to estimate data using this theory.

Given that a series is stationary, we can use the sample mean ($\bar{y}$), the sample variance ($\hat{\sigma}^2$), and the sample autocorrelations ($\hat{\rho}$) to estimate the parameters of the actual data-generating process:

$$
\begin{aligned}
\bar{y} &= \frac{\sum_{t=1}^{T} y_t}{T}, \\
\hat{\sigma}^2 &= \frac{\sum_{t=1}^{T} (y_t - \bar{y}^2)}{T}, \\
\hat{\rho}_i &= \frac{\sum_{i+1}^{T} (y_t - \bar{y})(y_{t-i} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}.
\end{aligned}
$$

I postpone the following contents untill Lec 6.

5.1 Specification

A strategy for appropriate model selection:

Before an ARMA(p, q) model can be estimated we need to select the order p and q of the ARMA (cf. MIT-TSA)

ACF and PACF "identify" the appropriate "parsimonious" AR, MA, or ARMA process.

AIC, SBC "identify"the appropriate p and/or q.

Again, I refer the reader to Enders (2015, p.69) to find details.

5.2 Estimation

It's very uncommon to estimate moving average terms.

An AR process are easy to estimate since the OLS assumption still apply, whereas MA terms have to be estimated by maximum likelihood since every MA has an AR(∞) representation, pure AR can approximate vector MA processes.

In this stage, the goal is to select a stationary and parsimonious model that has a good fit which in turn examin the various coefficients (ϕ_i, θ_i).

Enders(2015, p.70), MIT-TSA, etc.

5.3 Diagnostic checking

(1) In this stage, the goal is to ensure that the residuals from the estimated model mimic a white-noise process.
(2) Incorporating additional coefficients will necessarily increase fitness ($R^2 \uparrow$) at a cost of reducing degrees of freedom. Box and Jenkins (1976) argue that parsimonious models fits the data well without incorporating any needless coefficients and it can produce better forecasts than overparameterized models (e.g., AR(1) with ony one coefficient $\Leftrightarrow$ MA(∞) with many many coefficients).

$$(1-\phi_1 L-\phi_2 L^2)y_t = (1+\theta_1 L+\theta_2 L^2+\theta_3 L^3)\epsilon_t \xLeftrightarrow{\text{if, a common factor}} (1+\alpha L)(1+\phi L)y_t = (1+\alpha L)(1+\theta_1 L+\theta_2 L^2)\epsilon_t.$$

6 The 1st of 3 Fundamental Representations

If two processes have the same autocovariance function, they are the same process. Matching fundamental representations is one of the most common tricks in manipulating time series (see Cochrane 2005, pp.26-27).

7 Autocovariance-Generating Functions

See Lec 7: Spectral Analysis (cf. Hamilton 1994, chaper 3.6, chaper 10.3, and chapter 6)