1 Introduction: Why Time Series Analysis

- Compare the OLS estimation of β in AR(1) model

$$y_t = \beta y_{t-1} + u_t, |\beta| < 1 \tag{1}$$

and the OLS estimation of β in the model

$$y_t = \beta y_{t-1} + u_t, \beta = 1, \tag{2}$$

where $t = 1, 2, \cdots, T$; $y_0 = 0$, and $\{u_t\}$ is a white noise process, i.e.

$$\begin{aligned} E(u_t) &= 0, \\ E(u_t^2) &= \sigma^2, \\ E(u_t u_\tau) &= 0, \ t \neq \tau. \end{aligned}$$

The OLS estimator $\hat{\beta}$ is

$$\hat{\beta} = \left(\sum_{t=1}^{T} y_{t-1}^2 \right)^{-1} \sum_{t=1}^{T} y_t y_{t-1} = \beta + \left(\sum_{t=1}^{T} y_{t-1}^2 \right)^{-1} \sum_{t=1}^{T} u_t y_{t-1}.$$

This estimator is biased ($E\hat{\beta} \neq \beta$) since u_t is not independent of $y_t, y_{t+1}, \cdots, y_T$ (even though u_t is independent of y_{t-1}). The conventional t and F tests can not be applied. For Model (1), as $T \to \infty$,

$$\sqrt{T}(\hat{\beta} - \beta) \to N(0, 1 - \beta^2).$$

However, for Model (2), as $T \to \infty$,

$$\sqrt{T}(\hat{\beta} - \beta) \to 0 \text{ in probability},$$

which is of no use in test. How about $T(\hat{\beta} - 1)$ for Model (2)? From Model (2),

$$y_t = u_1 + u_2 + \cdots + u_t,$$

and hence $y_t \sim N(0, t\sigma^2)$ or $\frac{y_t}{\sigma\sqrt{t}} \sim N(0, 1)$. Note that $y_t^2 = (y_{t-1} + u_t)^2 = y_{t-1}^2 + u_t^2 + 2u_t y_{t-1}$. Therefore,

$$\begin{aligned} & \frac{1}{\sigma^2 T} \sum_{t=1}^{T} u_t y_{t-1} = \frac{1}{2\sigma^2 T} \sum_{t=1}^{T} \left(y_t^2 - y_{t-1}^2 - u_t^2 \right) \\ = & \frac{1}{2\sigma^2 T} y_T^2 - \frac{1}{2\sigma^2 T} \sum_{t=1}^{T} u_t^2 = \frac{1}{2} \left(\frac{y_T}{\sigma\sqrt{T}} \right)^2 - \frac{1}{2\sigma^2 T} \sum_{t=1}^{T} u_t^2 \\ \to & \frac{1}{2} \left(\chi^2(1) - 1 \right). \end{aligned}$$

Since $y_{t-1} \sim N(0,(t-1)\sigma^2)$, we have $E(y_{t-1}^2) = (t-1)\sigma^2$ and

$$E\left(\sum_{t=1}^{T} y_{t-1}^2\right) = \sum_{t=1}^{T}(t-1)\sigma^2 = \frac{T(T-1)}{2}\sigma^2 = O(T^2).$$

To construct a random variable with convergence distribution for the estimator $\hat{\beta}$ in Model (2), we should study

$$T(\hat{\beta}-1) = \frac{(1/T)\sum_{t=1}^{T} u_t y_{t-1}}{(1/T^2)\sum_{t=1}^{T} y_{t-1}^2} \tag{3}$$

instead of studying $\sqrt{T}(\hat{\beta}-1)$. Obviously, the asymptotic distribution of $T(\hat{\beta}-1)$ is not the same as normality. It will be shown that

$$T(\hat{\beta}-1) \to \left(\int_0^1 W^2(r)dr\right)^{-1} \int_0^1 W(r)dW(r),$$

where $W(r)$ is the standard Brownian motion. Hence the problems of stationarity and unstationarity.

- Examine the following simple model:

$$y_t = a_0 + a_1 z_t + e_t, \tag{4}$$

where $\{y_t\}$ and $\{z_t\}$ are two independent random walk processes, i.e.

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_{yt} && (5)\\ z_t &= z_{t-1} + \varepsilon_{zt} && (6) \end{aligned}$$

with two independent white-noise processes ε_{yt} and ε_{zt}. Any relationship between y_t and z_t is meaningless since $\{y_t\}$ and $\{z_t\}$ are independent. Sometimes OLS can give **spurious** estimation, i.e. high R^2 and significant t-test for the coefficients, but no economic meaning. Why? The error term e_t in the regression equation is

$$\begin{aligned} e_t &= y_t - a_0 - a_1 z_t \\ &= -a_0 + \sum_{i=1}^{t} \varepsilon_{yi} - a_1 \sum_{i=1}^{t} \varepsilon_{zi}. \end{aligned}$$

Since $Var(e_t) = t(\sigma_{\varepsilon_y}^2 + a_1^2\sigma_{\varepsilon_z}^2)$ becomes infinitely large as t increases, and $E_t e_{t+i} = e_t$ for $i \geq 0$, the t-test, F-test or R^2 values are unreliable. Under the null: $a_1 = 0$, $y_t = a_0 + e_t$. This is inconsistent with the distributional theory in OLS, which requires that the error term be a white-noise. Phillips (1986) shows that, the large the sample, the more likely to falsely reject the null (i.e. the more significant for the test in OLS). Therefore, before estimation, we should pretest the unstationarity of the time series variables in the regression model (the unit root test). A further study on some cases in Model (4):

1. If $\{y_t\}$ and $\{z_t\}$ are both stationary, the regression model is appropriate.
2. If $\{y_t\}$ and $\{z_t\}$ are integrated of different orders, the regression is meaningless.
3. If $\{y_t\}$ and $\{z_t\}$ are integrated of the same order and the residual sequence is unstationary, the regression is spurious.
4. If $\{y_t\}$ and $\{z_t\}$ are integrated of the same order and the residual sequence is stationary, then they are cointegrated.

Example 1 (see ex1): Give the graphs for the time series variables of the AR(1) and the Random walk process, and conduct the OLS estimation.

The purpose of this course is to introduce basic theory and applications of time series econometrics. The course requires basic knowledge of probability and statistics. Students are require to perform computations using EViews.

An outline of the course (tentative):
CH1 Basic Regression with Time Series
CH2 Stationary Autoregressive Process
CH3 ARCH and GARCH
CH4 Unstationary Autoregressive Process
CH5 Vector Autoregression (VAR) Models
CH6 Cointegration

There are no texts for this course. The following materials will be useful for the course.

- Walter Enters (2003), "Applied Econometric Time Series" (Second Edition).
- Hamilton (1994), "Time Series Analysis", Princeton University Press.
- Wooldridge(2009), Introductory Econometrics: A Modern Approach (4th Edition)

2 Stationary Autoregressive Process

Methods to solve the difference equation

- **Iterative method**: Consider, for example, the model

$$y_t = a_0 + a_1 y_{t-1} + u_t. \tag{7}$$

If y_0 is known, by iterating, y_t is expressed as a function of t, the known y_0, and the forcing process $x_t = \sum_{i=0}^{t-1} a_1^i u_{t-i}$, i.e.

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i u_{t-i}.$$

If y_0 is unkown, by further iterating, we have

$$\begin{aligned} y_t &= a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i u_{t-i} \\ &= a_0 \sum_{i=0}^{t+m} a_1^i + a_1^{t+m+1} y_{-m-1} + \sum_{i=0}^{t+m} a_1^i u_{t-i}. \end{aligned}$$

If $|a_1| < 1$, as $m \to \infty$,

$$\begin{aligned} y_t &= a_0 \sum_{i=0}^{\infty} a_1^i + \sum_{i=0}^{\infty} a_1^i u_{t-i} \\ &= \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i u_{t-i}. \end{aligned}$$

This is only a special solution to Model (7). The **general solution** is given by

$$y_t = A a_1^t + \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i u_{t-i}. \tag{8}$$

Choosing $A = y_0 - a_0/(1-a_1) - \sum_{i=0}^{\infty} a_1^i u_{-i}$, we can derive the special solution from the general one above. Note in (8) that Aa_1^t is the **homogeneous solution** of the homogeneous equation $y_t = a_1 y_{t-1}$ and the other part is a **particular solution** to the difference equation (7):

The general solution = the homogeneous solution + a particular solution.

Imposing the initial condition on the general solution gives a special solution satisfying the initial condition. If $|a_1| > 1$, given an initial condition y_0, y_t can be solved: $y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i u_{t-i}$. If no initial conditions are given, the sequence cannot be convergent. If $a_1 = 1$,

$$y_t = a_0 + y_{t-1} = a_0 t + y_0 + \sum_{i=1}^{t} u_i$$

and $\Delta y_t = a_0 + u_t$. This shows that each disturbance has a permanent non-decaying effect on the value of y_t.

- **How to determine the homogeneous solution?** The structure of the homogeneous equation is determined by the pattern of the characteristic roots. For AR(1) Model (7), the characteristic equation (root) is $\lambda = a_1$; hence the homogeneous solution is $y_t^h = Aa_1^t$, where A is an arbitray constant, which is interpreted as a deviation from long-run equilibrium. For AR(2) model $y_t = a_1 y_{t-1} + a_2 y_{t-2} + u_t$, the homogeneous equation is $y_t = a_1 y_{t-1} + a_2 y_{t-2}$. Inserting $y_t = A\lambda^t$ deduces that the characteristic equation is $\lambda^2 - a_1\lambda - a_2 = 0$. The two roots are

$$\lambda_1, \lambda_2 = \left(a_1 \pm \sqrt{a_1^2 + 4a_2}\right)/2.$$

Note that the linear combination of λ_1^t and λ_2^t also solves the homogeneous equation. There are three cases according as $a_1^2 + 4a_2 > 0, = 0$ and < 0. The homogeneous solutions are, respectively,

$$\begin{aligned} y_t^h &= A_1\lambda_1^t + A_2\lambda_2^t, \\ y_t^h &= (A_1 + A_2 t)\lambda^t, \ \lambda = \lambda_1 = \lambda_2 \\ y_t^h &= A_1 r^t \cos(\theta t + A_2), \ r = \sqrt{-a_2}, \theta = \arg tg\left(\sqrt{-a_1^2 - 4a_2}/a_1\right). \end{aligned}$$

For higher order homogeneous equation $y_t = \sum_{i=1}^{p} a_i y_{t-i}$, the characteristic equation is $\lambda^t - \sum_{i=1}^{p} a_i \lambda^{t-i} = 0$ or $\lambda^p - \sum_{i=1}^{p} a_i \lambda^{p-i} = 0$.

- **How to determine particular solutions?** Consider $y_t = \sum_{i=1}^{p} a_i y_{t-i} + x_t$. If x_t is deterministic, e.g. $x_t = 0; bd^{rt}; a_0 + bt^d$, setting $y_t^p = c, c_0 + c_1 d^{rt}, c_0 + c_1 t + \cdots + c_d t^d$,solve the constants by inserting the y_t^p into the equation. If $x_t = u_t$ is stochastic, set $y_t^s = \sum_{i=0}^{\infty} \alpha_i u_{t-i}$, insert y_t^s into the equation and solve the coefficients α_i by the method of undetermined coefficients.

- **Stable solution and stability conditions:** all the characteristic roots lie inside the unit circle. For $AR(2)$ model $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$, the stability conditions are

$$\begin{aligned} & a_2 + a_1 < 1 \\ & a_2 - a_1 < 1 \\ & a_2 > -1, a_2 < 0 \end{aligned}$$

- The general solution of the difference equation is $y_t = y_t^p + y_t^h$, where y_t^p is a particular solution of the difference equation; y_t^h is the general solution of its homogeneous equation. Further, y_t^p can be expressed as $y_t^d + y_t^s$, where y_t^d is the deterministic part and y_t^s is the stochastic part.

- y_t^d is determined according to the different cases of the deterministic part x_t in the difference equation, e.g. $x_t =$ constant, bd^{rt} or bt^d.

- y_t^s is determined by the stochastic part x_t in the difference equation, e.g. if $y_t = a_1 y_{t-1} + a_2 y_{t-2} + x_t$, and $x_t = \varepsilon_t$, set $y_t^s = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i}$.

- The coefficients of y_t^d and y_t^s in $y_t = y_t^p + y_t^h$ can be determined by substituting y_t into the original difference equation and using the method of undetermined coefficients.

- Lag operator is a linear operator, which is extensively applied in time series analysis. Note the application of $1/(1-aL)$, $|a| < 1$ or $|a| > 1$. If $|a| < 1$,

$$\frac{y_t}{1-aL} = \sum_{i=0}^{\infty} a^i L^i y_t = \sum_{i=0}^{\infty} a^i y_{t-i}.$$

  If $|a| > 1$,

$$\frac{y_t}{1-aL} = -(aL)^{-1} \frac{y_t}{1-(aL)^{-1}} = -(aL)^{-1} \sum_{i=0}^{\infty} (aL)^{-i} y_t = -(aL)^{-1} \sum_{i=0}^{\infty} a^{-i} y_{t+i}.$$

- $A(L) y_t = a_0 + B(L)\varepsilon_t$ has the particular solution $y_t = (a_0 + B(L)\varepsilon_t)/A(L)$. The stability condition is that the inverse characteristic roots (i.e. the roots of the inverse characteristic equation $A(L) = 0$) lie outside of the unit circle.

- Here we introduce three methods to express y_t as the sum of a function of time t and a moving average of the disturbance: Iterative Method, the Method of Undetermined Coefficients, and Lag Operator Approach. Whether such an expansion is convergent so that the original difference equation is stable is an important issue. It will be shown that, if y_t is expressed by a linear stochastic difference equation, the stability condition is a necessary condition for the time series $\{y_t\}$ to be stationary.

Stationary Processes

- $E_t y_{t+i} \equiv E[y_{t+i}|y_t, y_{t-1}, ..., y_1]$, the expectation value of y_{t+i} conditional on the observed values of $y_1, y_2, \cdots, y_t$.

- White-noise process: $\{\varepsilon_t\}$, $E[\varepsilon_t] = 0$, $Var(\varepsilon_t) = \sigma^2$ (constant), $E\varepsilon_t\varepsilon_{t-s} = E\varepsilon_{t-j}\varepsilon_{t-s-j}$ $= 0$, $\forall t$ and for all j, $s \neq 0$.

- $MA(q)$: a moving average of order q, $\{x_t\}$ satisfying $x_t = \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$, where $\beta_0 = 1$.If two or more of the coefficients β_i differ from 0, $\{x_t\}$ are not white-noise. (Consider $x_t = \varepsilon_t + 0.5\varepsilon_{t-1}$)

- $AR(p)$: p-order autoregressive, $y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t$

- $ARMA(p,q)$: (p,q)-order autoregressive moving-average process,

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}, \ \beta_0 = 1$$

  If at least one of the characteristic roots ≥ 1, $\{y_t\}$ is said to be an integrated process, called as $ARIMA(p,q)$ model.

- Moving-average representation of $ARMA(p,q)$ process $\{y_t\}$ (all the characteristic roots lie in the unit circle) in terms of ε_t:

$$y_t = \left(a_0 + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}\right) / \left(1 - \sum_{i=1}^{p} a_i L^i\right),$$

  which yields an $MA(\infty)$ process.

- $\{y_t\}$ **is (covariance) stationary** if, for any t, $t-s$,

$$\begin{aligned} Ey_t &= Ey_{t-s} = \mu, \\ Var(y_t) &= Var(y_{t-s}) = \sigma_y^2 < \infty, \\ Cov(y_t, y_{t-s}) &= Cov(y_{t-j}, y_{t-j-s}) \equiv \gamma_s \end{aligned}$$

are all constants, which are time-invariant (i.e. unaffected by a change of time origin). **Note the difference** between "(covariance) stationary" and "strictly stationary"($\forall j_1, j_2, \cdots, j_s$, the joint distribution of $(y_t, y_{t+j_1}, \cdots, y_{t+j_s})$ is determined by $j_1, j_2, \cdots, j_s$, but unaffected by a change of time origin; i.e. it is invariant to the time t in which the observations are made). The covariance stationary process requires that the mean and the covariance are time-invariant and finite while strictly stationary process requires that the mean, the covariance and the other higher moments be time-invariant, but not necessarily be finite. If the mean and the covariance are finite, the strong stationary process is covariance stationary and the strong stationarity is stronger.

- autocorrelation between y_t and y_{t-s}: $\rho_s \equiv \frac{\gamma_s}{\gamma_0} = \frac{Cov(y_t, y_{t-s})}{Cov(y_t, y_t)} = \frac{Cov(y_t, y_{t-s})}{\sigma_y^2}$, $\rho_0 = 1$. The autocorrelation coefficients ρ_s are time-invariant for the stationary sequence $\{y_t\}$.

- **Find stationarity conditions** for $AR(1)$:

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t,$$

where ε_t is white-noise. Note that, for any given initial condition y_0 (deterministic),

$$y_t = \frac{a_0(1 - a_1^t)}{1 - a_1} + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i}.$$

Since two means

$$Ey_t = \frac{a_0(1 - a_1^t)}{1 - a_1} + a_1^t y_0,$$
$$Ey_{t-s} = \frac{a_0(1 - a_1^{t-s})}{1 - a_1} + a_1^{t-s} y_0$$

are time dependent and $Ey_t \neq Ey_{t-s}$, the $\{y_t\}$ cannot be stationary. **Add restrictions:**

$$|a_1| < 1 \text{ and } \{y_t\} \text{ have been occurring for an infinitely long time.}$$

Then for any integer $m > 0$,

$$y_t = \frac{a_0(1 - a_1^{t+m+1})}{1 - a_1} + a_1^{t+m+1} y_{-m-1} + \sum_{i=0}^{t+m} a_1^i \varepsilon_{t-i}. \tag{9}$$

As $m \to \infty$, $y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$. Consider this $\{y_t\}$, the limit of (9). Since

$$Ey_t = \frac{a_0}{1-a_1} = Ey_{t-s},$$
$$Var(y_{t-s}) = \frac{\sigma^2}{1-a_1^2},$$
$$Cov(y_t, y_{t-s}) = \sigma^2 \sum_{j=0}^{\infty} a_1^{2j+s} \delta_{i,j+s} = \frac{\sigma^2 a_1^s}{1-a_1^2}$$

are all time-invariant, $\{y_t\}$ is stationary. Here we use $\delta_{ij} = 1$, if $i = j$; 0, otherwise. We have known that, if no initial condition were given, the general solution of AR(1) is (use the solution construction $y_t = y_t^p + y_t^h$)

$$y_t = \frac{a_0}{1-a_1} + \sum_{i=1}^{\infty} a_1^i \varepsilon_{t-i} + A a_1^t,$$

which cannot be stationary unless $A a_1^t \neq 0$. Therefore, for $\{y_t\}$ to be stationary, the initial condition cannot be deterministic and the sequence must have started infinitely long ago or the arbitrary constant A must be zero (no deviation from the long-run equilibrium). Therefore, **the stationarity conditions for $AR(1)$ sequence $\{y_t\}$ are:**

$$\begin{aligned} y_t^h &\equiv 0, \\ |a_1| &< 1. \end{aligned}$$

The former follows from the assumption: Either the sequence $\{y_t\}$ must have started infinitely far in the past or the process must always be in equilibrium (so that $A = 0$). The latter says that the characteristic root of AR(1) sequence must be less than unity in absolute value.

A hint from above is that if any portion of the homogeneous equation is present, the mean, variance, and all covariances will be time-dependent. Hence, for any ARMA(p, q) model, stationarity necessitates that the homogeneous solution be zero.

- **Find stationarity conditions** for $ARMA(2, 1)$:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1}.$$

From above, stationarity requires that y_t^h must be zero. It is only necessary to find a particular solution of $ARMA(2, 1)$. Now try a particular solution $y_t = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i}$ by using the method of undetermined coefficients:

$$\alpha_0 \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \sum_{i=2}^{\infty} \alpha_i \varepsilon_{t-i} = \varepsilon_t + (a_1 \alpha_0 + \beta_1) \varepsilon_{t-1} + \sum_{i=2}^{\infty} (a_1 \alpha_{i-1} + a_2 \alpha_{i-2}) \varepsilon_{t-i}$$

and hence

$$\alpha_0 = 1$$
$$\alpha_1 = a_1\alpha_0 + \beta_1 \Rightarrow \alpha_1 = a_1 + \beta_1$$
$$\alpha_i = a_1\alpha_{i-1} + a_2\alpha_{i-2},\ i \geq 2.$$

Therefore, $y_t = \varepsilon_t + (a_1 + \beta_1)\varepsilon_{t-1} + \sum_{i=2}^{\infty} \alpha_i\varepsilon_{t-i}$, where α_i are determined by the difference equation $\alpha_i = a_1\alpha_{i-1} + a_2\alpha_{i-2},\ i \geq 2$, with $\alpha_0 = 1$ and $\alpha_1 = a_1 + \beta_1$. **If the characteristic roots of $ARMA(2,1)$ model lie within the unit circle**, $\{\alpha_i\}$ constitute a convergent sequence, and $\{y_t\}$ become stationary. Check the stationarity conditions for $\{y_t\}$ generated by the $ARMA(2,1)$:

$$Ey_t = 0 = Ey_{t-s}, \forall t, s,$$

$$Var(y_t) = Var(y_{t-s}) = \sigma^2 \sum_{i=0}^{\infty} \alpha_i^2, \forall t, s,$$

$$Cov(y_t, y_{t-s}) = E\sum_{i,j=0}^{\infty} \alpha_i\alpha_j\varepsilon_{t-i}\varepsilon_{t-s-j} = \sigma^2 \sum_{i,j=0}^{\infty} \alpha_i\alpha_j\delta_{i,s+j} = \sigma^2 \sum_{j=0}^{\infty} \alpha_{s+j}\alpha_j.$$

- **Find stationarity conditions** for $MA(\infty) : x_t = \sum_{i=0}^{\infty} \beta_i\varepsilon_{t-i},\ \beta_0 = 1.\ \forall t, s,$

$$\begin{aligned} Ex_t &= 0 = Ex_{t-s} \\ Var(x_t) &= \sigma^2 \sum_{i=0}^{\infty} \beta_i^2 = Var(x_{t-s}) \\ Cov(x_t, x_{t-s}) &= Ex_t x_{t-s} = \sigma^2 \sum_{i=0}^{\infty} \beta_i\beta_{i+s} \end{aligned}$$

If $\sum_{i=0}^{\infty} \beta_i\beta_{i+s} < \infty\ \forall s$, $MA(\infty)$ will be stationary. A direct implication is that $MA(q)$ is always stationary for any finite q.

- **Find stationarity conditions** for $AR(p)$:

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t.$$

If the characteristic roots of the homogeneous equation all lie inside the unit circle (and hence $1 - \sum_{i=1}^{p} a_i \neq 0$), the particular solution

$$y_t = \frac{a_0}{1 - \sum_{i=1}^{p} a_i} + \sum_{i=0}^{\infty} \alpha_i\varepsilon_{t-i}$$

is convergent, since the series $\{\alpha_i\}$ solve the difference equation $\alpha_i - \sum_{j=1}^{p} a_j \alpha_{i-j} = 0$. Check the stationary conditions:

$$\begin{aligned}
Ey_t &= Ey_{t-s} = \frac{a_0}{1-\sum_{i=1}^{p} a_i}, \\
Var(y_t) &= E\left(\sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}\right) = \sum_{i,j=0}^{\infty} \alpha_i \alpha_j E(\varepsilon_{t-i}\varepsilon_{t-j}) \\
&= \sum_{i,j=0}^{\infty} \alpha_i \alpha_j \sigma^2 \delta_{i,j} = \sigma^2 \sum_{i=0}^{\infty} \alpha_i^2 < \infty, \\
Con(y_t, y_{t-s}) &= E\left(\sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-s-j}\right) = \sigma^2 \sum_{i,j=0}^{\infty} \alpha_i \alpha_j \delta_{i,s+j} \\
&= \sigma^2 \sum_{j=0}^{\infty} \alpha_{j+s} \alpha_j < \infty
\end{aligned}$$

are all time-invariant for t.

- **Find stationarity conditions** for $ARMA(p,q)$:

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}, \;\; \beta_0 = 1.$$

Since $\{\sum_{i=0}^{q} \beta_i \varepsilon_{t-i}\}$ is stationary for any finite q, only the characteristic roots of the autoregressive portion of the $ARMA(p,q)$ process determine whether the $\{y_t\}$ is stationary. Therefore, **if the roots of the inverse characteristic equation** $1 - a_1 L - a_2 L^2 - \cdots - a_p L^p = 0$ **lie outside of the unit circle**, then $\{y_t\}$ is stationary.

Autocorrelation function (ACF): $\rho_s \equiv \gamma_s/\gamma_0$

- The autocorrelation function $\rho_s = \frac{\gamma_s}{\gamma_0} = \frac{Cov(y_t, y_{t-s})}{Var(y_t)}$, the plot of which against s, serves as a useful tool to identify and estimate time-series models (the Box-Jenkins Approach).

- **ACF for** $AR(1)$ **process**: $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$. Since $y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$, we have

$$\begin{aligned}
\gamma_0 &= Var(y_t) = \sigma^2/(1-a_1^2), \\
\gamma_s &= Cov(y_t, y_{t-s}) = \sigma^2 \sum_{j=0}^{\infty} a_1^{2j+s} \delta_{i,j+s} = \frac{\sigma^2 a_1^s}{1-a_1^2}.
\end{aligned}$$

Therefore $\rho_0 = 1$, $\rho_s = a_1^s$ for $s \geq 1$. The ACF ρ_s converge to 0 geometrically (directly or oscillatorily according as $a_1 > 0$ and $a_1 < 0$) as $s \to \infty$, provided $|a_1| < 1$.

- **ACF for $AR(2)$ process**: $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$. Stationarity requires that the roots of $(1 - a_1 L - a_2 L^2)$ be outside the unit circle. Note that $y_t = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i}$, $E y_t \varepsilon_t = \sigma^2$, and $E y_{t-s} \varepsilon_t = 0$, $\forall s \geq 1$. **Yule-Walker technique**: multiply $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$ by y_t, $y_{t-1}, \cdots, y_{t-s}$, respectively, and take expectation,

$$\begin{aligned}
E y_t y_t &= a_1 E y_{t-1} y_t + a_2 E y_{t-2} y_t + E y_t \varepsilon_t, \\
E y_t y_{t-1} &= a_1 E y_{t-1} y_{t-1} + a_2 E y_{t-2} y_{t-1} + E y_{t-1} \varepsilon_t, \\
E y_t y_{t-2} &= a_1 E y_{t-1} y_{t-2} + a_2 E y_{t-2} y_{t-2} + E y_{t-2} \varepsilon_t, \\
&\vdots \\
E y_t y_{t-s} &= a_1 E y_{t-1} y_{t-s} + a_2 E y_{t-2} y_{t-s} + E y_{t-s} \varepsilon_t,
\end{aligned}$$

we have $\gamma_0 = a_1 \gamma_1 + a_2 \gamma_2 + \sigma^2$, $\gamma_s = a_1 \gamma_{s-1} + a_2 \gamma_{s-2}$, $s \geq 1$. Therefore,

$$\begin{aligned}
\rho_0 &= 1, \ \rho_1 = a_1/(1 - a_2), \\
\rho_s &= a_1 \rho_{s-1} + a_2 \rho_{s-2}, \ s \geq 2.
\end{aligned}$$

The key point is that ρ_s satisfy the difference equation (note that $\rho_s = \rho_{-s}$): $\rho_s = a_1 \rho_{s-1} + a_2 \rho_{s-2}$, $s > 0$, which is stationary since the characteristic roots lie inside the unit circle. The ACF converge to 0 (directly or oscillatorily) as $s \to \infty$.

- **ACF for $MA(1)$ process**: $y_t = \varepsilon_t + \beta \varepsilon_{t-1}$.

$$\begin{aligned}
\gamma_0 &= Var(y_t) = (1 + \beta^2)\sigma^2, \\
\gamma_1 &= Cov(y_t, y_{t-1}) = \beta \sigma^2, \\
\gamma_s &= Cov(y_t, y_{t-s}) = 0, \ s \geq 2.
\end{aligned}$$

Hence

$$\begin{aligned}
\rho_0 &= 1, \\
\rho_1 &= \beta/(1 + \beta^2), \\
\rho_s &= 0, \ s \geq 2.
\end{aligned}$$

The ACF ρ_s $(s = 1, 2, ...)$ has one spike $(\rho_1 \neq 0)$ and then cuts to 0.

- **ACF for $MA(2)$ process**: $y_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2}$.

$$\begin{aligned}\gamma_0 &= Var(y_t) = (1+\beta_1^2+\beta_2^2)\sigma^2, \\ \gamma_1 &= Cov(y_t, y_{t-1}) = (\beta_1+\beta_1\beta_2)\sigma^2, \\ \gamma_2 &= Cov(y_t, y_{t-2}) = \beta_2\sigma^2, \\ \gamma_s &= Cov(y_t, y_{t-s}) = 0, \ s \geq 3.\end{aligned}$$

  The ACF has two spikes and then cuts to $0 : \rho_s = 0, \ s \geq 3$. Generally, for $MA(q) : y_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \cdots + \beta_q\varepsilon_{t-q}$, its ACF has q spikes and then cuts to $0 : \rho_s = 0, \ s \geq q+1$.

- **ACF for $ARMA(1,1)$ process**: $y_t = a_1 y_{t-1} + \varepsilon_t + \beta_1\varepsilon_{t-1}$. Note that $Ey_t\varepsilon_t = \sigma^2$, $Ey_t\varepsilon_{t-1} = (a_1+\beta_1)\sigma^2$ and $Ey_{t-s}\varepsilon_t = 0$, $\forall s > 0$.

$$\begin{aligned}\gamma_0 &= Ey_ty_t = a_1\gamma_1 + \sigma^2 + \beta_1(a_1+\beta_1)\sigma^2 \\ \gamma_1 &= Ey_ty_{t-1} = a_1\gamma_0 + \beta_1\sigma^2 \\ \gamma_s &= Ey_ty_{t-s} = a_1\gamma_{s-1}, \ s > 1.\end{aligned}$$

  Hence $\gamma_0 = \frac{1+\beta_1^2+2a_1\beta_1}{1-a_1^2}\sigma^2$ and $\gamma_1 = \frac{(1+a_1\beta_1)(a_1+\beta_1)}{1-a_1^2}\sigma^2$. Therefore,

$$\begin{aligned}\rho_1 &= \frac{(1+a_1\beta_1)(a_1+\beta_1)}{1+\beta_1^2+2a_1\beta_1} \\ \rho_s &= a_1\rho_{s-1}, \ s \geq 2,\end{aligned}$$

  which can be solved from the initial condition ρ_1.The ACF ρ_s converge to 0 geometrically (directly or oscillatorily according as $a_1 > 0$ and $a_1 < 0$), as $s \to \infty$, provided $|a_1| < 1$.

- **ACF for $ARMA(p,q)$ process**: $y_t = a_1y_{t-1}+\cdots+a_py_{t-p}+\varepsilon_t+\beta_1\varepsilon_{t-1}+\cdots+\beta_q\varepsilon_{t-q}$. The ACF ρ_s $(s = 1, 2, \cdots, q)$ calculation is complicated, thus omitted here. The ACF ρ_s $(s > q)$ satisfy (Note that $\rho_s = \rho_{-s}$):

$$\rho_s = a_1\rho_{s-1} + \cdots + a_p\rho_{s-p}, \ s \geq q+1.$$

  Under the stationarity restriction (all the characteristic roots of the model are within the unit circle), the ACF converge to 0 as $s \to \infty$.

Partial Autocorrelation Function (PACF): ϕ_{ss}

- PACF between y_t and y_{t-s} eliminates the effects of the intervening values y_{t-1}, $y_{t-2}, \cdots, y_{t-s+1}$.

- How to do? Set $y_t^* = y_t - Ey_t$ and construct

$$y_t^* = \phi_{11} y_{t-1}^* + e_t,$$

then $\phi_{11} = PACF$ between y_t and $y_{t-1.}$ Construct

$$y_t^* = \phi_{21} y_{t-1}^* + \phi_{22} y_{t-2}^* + e_t,$$

then $\phi_{22} = PACF$ between y_t and $y_{t-2.}$ The same arguments for $\phi_{33}, \phi_{44}, ...$

- AR(P): no direct correlation between y_t and y_{t-s} for $s > p$, i.e. $\phi_{ss} = 0$ for $s \geq p + 1$.

- MA(q) results in infinite-order AR representation. The PACF exhibit a decay.

- ARMA(p,q): ACF begin to decay after lag q (since the ACF for MA(q) cut to 0 after lag q and the ACF for AR(p) decay) while PACF begin to decay after lag p (since the PACF for AR(p) cut to 0 after lag p and the PACF for MA(q) exhibit a decay).

- A rule to select models is used by comparing the graphs of the ACF and PACF to the theoretical patterns. For example, if the ACF exhibited a single spike and the PACF exhibited monotonic decay, try to select an MA(1) model; however, if the ACF exhibited monotonic decay and the PACF exhibited a single spike, try to select an AR(1) model. If the ACF exhibited monotonic decay and the PACF exhibited two spikes, try to select an AR(2) model. If the PACF exhibited monotonic decay with no spikes, try to select an ARMA or MA model. .... (see ex2 for the graphs of ACF and PACF for some simpe ARMA models).

Sample ACF and Sample PACF and Model Selection

- **Sample ACF and Sample PACF** of $\{y_t\}_{t=1}^T$: The sample ACF and the sample PACF help identify the true model of the data generating process. Define $\bar{y} = (1/T)\sum_{t=1}^T y_t$, $\hat{\sigma}^2 = (1/T)\sum_{t=1}^T (y_t - \bar{y})^2$. For $s = 1, 2, \cdots$, define the sample ACF

$$r_s = \frac{\sum_{t=s+1}^T (y_t - \bar{y})(y_{t-s} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2},$$

which is the sample analog for the ACF $\rho_s = Cov(y_t, y_{t-s})/Var(y_t)$. The sample PACF $\hat{\phi}_{ss}$ is the estimator of ϕ_{ss} in $y_t^* = \phi_{s1} y_{t-1}^* + \cdots + \phi_{ss} y_{t-s}^* + e_t$. The sample partial autocorrelation at lag s is recursively determined by

$$\hat{\phi}_{ss} = \begin{cases} r_1, \text{ for } s = 1 \\ \left(r_s - \sum_{j=1}^{s-1} \hat{\phi}_{s-1,j} r_{s-j}\right) / \left(1 - \sum_{j=1}^{s-1} \hat{\phi}_{s-1,j} r_j\right), \text{ for } s \geq 2, \end{cases}$$

where $\hat{\phi}_{s,j} = \hat{\phi}_{s-1,,j} - \hat{\phi}_{ss}\hat{\phi}_{s-1,s-j}$, $j = 1, 2, \cdots, s-1$. $\hat{\phi}_{ss}$ is a consistent approximation of the PACF. For example, if the true value of r_s is zero (i.e. $\rho_s = 0$), there is no autoregression part in the process and hence the process is $MA(s-1)$; if the true value of $\hat{\phi}_{ss}$ is zero (i.e. $\phi_{ss} = 0$), there is no moving average part in the process and hence the process is $AR(s-1)$.

- Under the null: $y_t \sim MA(s-1)$ (i.e. $\rho_s = 0$) with normally distributed errors, $r_s \sim N(0, Var(r_s))$ asymptotically, where

$$Var(r_s) = \begin{cases} T^{-1}, & s = 1 \\ T^{-1}\left(1 + 2\sum_{i=1}^{s-1} r_j^2\right), & s > 1. \end{cases}$$

  Under the null: $y_t \sim AR(p)$ (i.e. $\phi_{p+i,p+i} = 0$, $i > 0$), the variance $Var(\hat{\phi}_{p+i,p+i})$ is approximately $1/T$. In EViews, the dotted lines in the plots of the ACF and PACF are the approximate two standard error bounds of r_1 or $\hat{\phi}_{11}$ computed as $\pm 2/\sqrt{T}$. If the value of the ACF or PACF is within these bounds, it is not significantly different from zero at (approximately) the 5% significance level.

- **Two kinds of test: (1) t-test:** From the sample ACF, construct t-ratio: $t = r_s/\sqrt{Var(r_s)}$ for the significance of s-order autocorrelation for some $s > 0$ ($H_0 : \rho_s = 0$ or $y_t \sim MA(s-1)$). From the sample PACF, construct t-ratio: $t = \sqrt{T}\hat{\phi}_{p+i,p+i}$ for the significance of p-order autoregression ($H_0 : \phi_{p+i,p+i} = 0$ or $y_t \sim AR(p)$). **(2) Q-statistic** (Box and Pierce (1970)): test whether a group of autocorrelation is significantly different from zero. It shows that $Q = T\sum_{k=1}^{s} r_k^2 \sim \chi^2(s)$ under the null hypothesis that all $r_k = 0$ for $k = 1, 2, ..., s$. Rejecting the null hypothesis means that at least one autocorrelation is not zero. For a white-noise process, $Q = 0$. If the calculated value of Q exceeds the critical point of $\chi^2(s)$, reject the null hypothesis, meaning that at least one autocorrelation is not zero and there are some autoregressive terms in the model. (this statistic works poorly). **Modified Q-statistic** (Ljung and Box (1978)): $Q = T(T+2)\sum_{k=1}^{s} r_k^2/(T-k) \sim \chi^2(s)$. In EViews, the Ljung and Box Q-statistic and the P-values are presented.

- Q-statistic can be used to check if the residuals from an estimated $ARMA(p, q)$ model behave as a white-noise process. Note that in this case the degrees of freedom is $s - p - q$, and hence $Q \sim \chi^2(s - p - q)$. If a constant is included in $ARMA(p, q)$, $Q \sim \chi^2(s - p - q - 1)$.

- There is a tradeoff between reducing the estimated SSR (from adding lags for p and/or q) and increasing the degrees of freedom. In application, try to find a more parsimonious model for the estimation.

- **Model Selection Criteria——select the most appropriate model by using the AIC and the SBC:**

$$\begin{aligned} AIC &= T\ln(SSR) + 2n \\ SBC &= T\ln(SSR) + n\ln T \end{aligned}$$

where n is the number of parameters in the estimation ($p + q+$ constant term), T is the number of usable observations (fixed. Here, the same sample period for different models should be used). In EViews,

$$\begin{aligned} AIC &= (-2\ln L + 2n)/T, \\ SBC &= (-2\ln L + n\ln T)/T. \end{aligned}$$

The different methods of calculating the AIC or the SBC will necessarily select the same model.

- **The smaller the AIC and the SBC, the better is the selected model:** model A fits better than model B if the AIC or the SBC for model A is smaller than for model B. Since $\ln T > 2$, the SBC will always select a more parsimonious model than the AIC. It is wonderful if both the AIC and the SBC select the same model; if not, be cautious. The AIC can select an overparameterized model, so t-test on all the coefficients should be significant when we estimate the model selected from the AIC. See the examples about the estimation of AR(1), AR(2), and ARMA(1,1) models on pages 70 through 75.

- **Three-stage Model Selection Method** (Box-Jenkins(1976))**:** Identification, estimation, and diagnostic checking.

  1. Identification: Examine the time plot of the series (data), the ACF, and the PACF visually. Stationary or nonstationary? Nonstationary variables may have a pronounced trend or appear to meander without a constant long-run mean or variance (hence some test approaches for nonstationarity are necessary). Standard practice was to first-difference nonstationary series and make them stationary. Under stationarity, a comparison of the sample ACF and PACF to those of various theoretical ARMA processes may suggest plausible ARMA models.

  2. Estimation: Fit the suggested models under stationarity and examine the estimates of a_i and β_i in the ARMA models. Select the model with parsimony, where Q-statistic, AIC, and SBC are used for model selection. There is a

tradeoff: more coefficients or more degrees of freedom? Three points to Note: **1) Parsimony of $\mathbf{ARMR(p,q)}$.** A parsimonious model fits the data well without incorporating any needless parameters. Avoid the common factor problem (e.g. use $AR(1)$ model $y_t = 0.5y_{t-1} + \varepsilon_t$ instead of $AR(2,1)$ model $y_t = 0.25y_{t-2} + \varepsilon_t + 0.5\varepsilon_{t-1}$). The coefficients should not be strongly correlated with each other. Coefficient estimates in the model should be significant (t-test or F-test). **2) Stationarity and Invertibility.** Be cautious: the estimated coefficient for $AR(1)$ model is close to 1 and the characteristic roots of the estimated polynomial $1 - a_1L - \cdots - a_pL^p$ for $AR(p)$ lie inside but close to the unit circle. Invertibility of the moving average part is required for the estimation of an $ARMA(p,q)$ model even though there is nothing improper about a non-invertible model. Consider the MLE of MA(1): $y_t = \varepsilon_t + \beta\varepsilon_{t-1}$, where $\{y_t\}_{t=1}^T$ is the observed series and $\varepsilon_0 = 0$. Suppose that $\{\varepsilon_t\}$ is a white-noise sequence drawn from a normal distribution $N(0,\sigma^2)$, i.e. the likelihood of ε_t is $1/\sqrt{2\pi}\exp\left(-\varepsilon_t^2/(2\sigma^2)\right)$. The log likelihood of the joint realizations $\{\varepsilon_t\}_{t=1}^T$ is $\frac{-T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^T \varepsilon_t^2$. If $|\beta| < 1$,

$$\varepsilon_t = y_t/(1+\beta L) = \sum_{i=0}^{\infty}(-\beta)^i y_{t-i} = \sum_{i=0}^{t-1}(-\beta)^i y_{t-i},$$

which shows that the values of ε_t represent a convergent process (the MA process is invertible). The log likelihood function is

$$\ln L = \frac{-T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}(-\beta)^i y_{t-i}\right)^2.$$

It is possible for computers to use search algorithms to select the values of β and σ^2 that maximize the value of $\ln L$. However, if $|\beta| > 1$, $\{\varepsilon_t\}$ cannot be represented in terms of the observed $\{y_t\}$ series, and the MA process is not invertible. The MLE is invalid in this case. **3) Goodness of fit.** The AIC and/or SBC are more approprite measures of the overall fit of the model. The common R^2 and the average of the residual sum of squares are not good goodness-of-fit measures in the estimation of ARMA models. Two reasons: one is that the fit necessarily improves as more parameters are included in the model; the other is that, in the nonlinear search algorithms for the estimation of ARMA models, if the search fails to converge rapidly, the estimated parameters may be unstable and adding one or two additional observations can greatly alter the estimates.

3. Diagnosis: Plot the residuals from the estimated model to look for outliers and for evidence of periods in which the model does not fit the data well. Construct the sample ACF and the PACF of the residuals and conduct the t-test and Q-test (in the above) to see whether any one or all of the residual autocorrelations or partial autocorrelation are statistically significant or not. If several residual correlations are marginally significant (from the t-test) and a Q-statistic test is not significant at 10% level, be wary: it is possible to form a better performing model. If there are sufficient observations, fit the same ARMA model to each of two subsamples (The standard F-test can be applied to test whether the data-generating process is unchanging).

Notes: (1) if all the plausible ARMA models estimated above show evidence of a poor fit during a reasonably long portion of the sample, consider multivariate estimation methods; (2) if the variance of the residuals is increasing or has some tendency to change, use a logarithmic transformation or ARCH techniques.

Forecast

- $AR(1): y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$. Given the actual data-generating process (i.e. a_0 and a_1 are known) and the current and past realizations of the $\{\varepsilon_t\}$ and $\{y_t\}$, by forward iteration,

$$\begin{cases} y_{t+1} = a_0 + a_1 y_t + \varepsilon_{t+1} \\ E_t y_{t+1} = a_0 + a_1 y_t, \end{cases}$$

$$\begin{cases} y_{t+2} = a_0(1 + a_1) + a_1^2 y_t + \varepsilon_{t+2} + a_1 \varepsilon_{t+1} \\ E_t y_{t+2} = a_0(1 + a_1) + a_1^2 y_t, \end{cases}$$

$$\vdots$$

$$\begin{cases} y_{t+j} = a_0(1 + a_1 + \cdots + a_1^{j-1}) + a_1^j y_t + \varepsilon_{t+j} + a_1 \varepsilon_{t+j-1} + \cdots + a_1^{j-1} \varepsilon_{t+1} \\ E_t y_{t+j} = a_0(1 + a_1 + \cdots + a_1^{j-1}) + a_1^j y_t. \end{cases}$$

The j-step-ahead forecast (**forecast function**) $\{E_t y_{t+j}\}_{j=1}^{\infty}$ is a function of the information set in period t, satisfying

$$\lim_{j \to \infty} E_t y_{t+j} = \frac{a_0}{1 - a_1} \text{ if } |a_1| < 1.$$

For the stationary AR(1) process, the conditional forecast of y_{t+j} converges to the unconditional mean as $j \to \infty$. Note that the j-step-ahead forecast error is

$$e_t(j) = y_{t+j} - E_t y_{t+j} = \varepsilon_{t+j} + a_1 \varepsilon_{t+j-1} + \cdots + a_1^{j-1} \varepsilon_{t+1}$$

with $E_t e_t(j) = 0$ and $Var[e_t(j)] = \sigma^2[1+a_1^2+\cdots+a_1^{2(j-1)}] = \sigma^2(1-a_1^{2j})/(1-a_1^2) \to \sigma^2/(1-a_1^2)$ as $j \to \infty$. The forecasts are unbiased, but the variance of the forecast errors is an increasing function of j, implying that the quality of the forecasts declines as we forecast further out into the future. If we assume that ε_t is normally distributed, then the 95% confidence interval for the one-step-ahead forecast of y_{t+1} is $a_0 + a_1 y_t \pm 1.96\sigma$ and the 95% confidence interval for j-step-ahead forecast of y_{t+j} is

$$\frac{a_0(1-a_1^j)}{1-a_1} + a_1^j y_t \pm 1.96\sigma \left(\frac{1-a_1^{2j}}{1-a_1^2}\right)^{1/2}.$$

- $ARMA(2,1): y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1}.$ Assume that all the coefficients are known, all the variables subscripted $t, t-1, \cdots$ are known at t, and $E_t \varepsilon_{t+j} = 0$ for $j > 0$.

$$\begin{cases} y_{t+1} = a_0 + a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1} + \beta_1 \varepsilon_{t.} \\ E_t y_{t+1} = a_0 + a_1 y_t + a_2 y_{t-1} + \beta_1 \varepsilon_t, \end{cases}$$

$$\begin{cases} y_{t+2} = a_0 + a_1 y_{t+1} + a_2 y_t + \varepsilon_{t+2} + \beta_1 \varepsilon_{t+1} \\ E_t y_{t+2} = a_0 + a_1 E_t y_{t+1} + a_2 y_t \\ = a_0(1+a_1) + (a_1^2 + a_2) y_t + a_1 a_2 y_{t-1} + a_1 \beta_1 \varepsilon_t \end{cases}$$

$$\begin{cases} y_{t+3} = a_0 + a_1 y_{t+2} + a_2 y_{t+1} + \varepsilon_{t+3} + \beta_1 \varepsilon_{t+2} \\ E_t y_{t+3} = a_0 + a_1 E_t y_{t+2} + a_2 E_t y_{t+1} \end{cases}$$

$$\vdots$$

$$\begin{cases} y_{t+j} = a_0 + a_1 y_{t+j-1} + a_2 y_{t+j-2} + \varepsilon_{t+j} + \beta_1 \varepsilon_{t+j-1} \\ E_t y_{t+j} = a_0 + a_1 E_t y_{t+j-1} + a_2 E_t y_{t+j-2}, \ j \geq 2. \end{cases}$$

Given the sample size T and the estimated coefficients $\hat{a}_0, \hat{a}_1, \hat{a}_2$ and $\hat{\beta}_1$, the estimated ARMA(2,1) model is

$$y_t = \hat{a}_0 + \hat{a}_1 y_{t-1} + \hat{a}_2 y_{t-2} + \hat{\varepsilon}_t + \hat{\beta}_1 \hat{\varepsilon}_{t-1.}$$

The out-of-sample forecasts can be easily constructed as follows:

$$E_T y_{T+1} = \hat{a}_0 + \hat{a}_1 y_T + \hat{a}_2 y_{T-1} + \hat{\beta}_1 \hat{\varepsilon}_{T.}$$

$$E_T y_{T+2} = \hat{a}_0 + \hat{a}_1 E_T y_{T+1} + \hat{a}_2 y_T$$

$$E_T y_{T+3} = \hat{a}_0 + \hat{a}_1 E_T y_{T+2} + \hat{a}_2 E_T y_{T+1}$$

$$E_T y_{T+j} = \hat{a}_0 + \hat{a}_1 E_T y_{T+j-1} + \hat{a}_2 E_T y_{T+j-2}, \quad j \geq 2.$$

However, the confidence intervals for the forecasts are difficult to construct.

- $ARMA(p,q): y_t = a_0 + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q}$.

$$\begin{aligned}
E_t y_{t+1} &= a_0 + a_1 y_t + \cdots + a_p y_{t+1-p} + \beta_1 \varepsilon_t + \cdots + \beta_q \varepsilon_{t+1-q} \\
E_t y_{t+2} &= a_0 + a_1 E_t y_{t+1} + \cdots + a_p y_{t+2-p} + \beta_2 \varepsilon_t + \cdots + \beta_q \varepsilon_{t+2-q} \\
&\vdots \\
E_t y_{t+q} &= a_0 + a_1 E_t y_{t+q-1} + \cdots + a_p E_t y_{t+q-p} + \beta_q \varepsilon_t \\
E_t y_{t+j} &= a_0 + a_1 E_t y_{t+j-1} + \cdots + a_p E_t y_{t+j-p}, \quad j > q.
\end{aligned}$$

The same argument as in ARMA(2,1) can be applied to construct the out-of-sample forecasts.

- **Forecast Evaluation:** Fit the best $\neq$ Forecast the best. Two aspects of uncertainty: the forecast error and the estimated coefficients result in bad forecasts. How to know the model with the best forecasting performance? Need enough observations. Some methods:

  1. Regression-based method: (1) Apart the sample $\{y_t\}_{t=1}^T$ into two parts $\{y_t\}_{t=1}^{T_0}$ and $\{y_t\}_{t=T_0+1}^T$, the first of which is used for estimation and the second for forecasts; (2) Construct one-step-ahead or j-step-ahead forecasts $\{f_t\}_{t=T_0+1}^T$; (3) Regress y_t on a constant and f_t for $t = T_0+1, \cdots, T$, i.e. $y_t = a_0 + a_1 f_t + v_t$, and apply the F-test to test the null $a_0 = 0$ and $a_1 = 1$. Rejecting the null means that the forecast is poor. If the significance levels from the F-tests of different models are similar, select the model with the smallest residual variance $Var(v_t)$.
  2. MSPE-based method: Construct $MSPE = \sum_{i=1}^H e_i^2$ for different models, where H is the number of observations in the holdback period (the second part of the sample), e_i is the forecast error. Take the larger MSPE of the two models in the numerator and construct the F-test

$$F \equiv \frac{MSPE_1}{MSPE_2} = \frac{\sum_{i=1}^H e_{1i}^2}{\sum_{i=1}^H e_{2i}^2} \sim F(H,H).$$

  The assumptions for the F-distribution are: $e_t \sim N(0, \delta^2)$, $Ee_t e_{t-s} = 0 (s \neq 0)$ and $Ee_{1t}e_{2t} = 0$. The violation of any one of the assumptions will lead to the failure of the F-distribution.

3. The Granger-Newbold(1976) test: ($Ee_{1t}e_{2t} = 0$ is violated). Set

$$\begin{aligned} x_t &= e_{1t} + e_{2t},\ z_t = e_{1t} - e_{2t} \\ \rho_{xz} &= Ex_tz_t = Ee_{1t}^2 - Ee_{2t}^2 \begin{cases} > 0, & \text{model 1 has a larger MSPE} \\ < 0, & \text{model 2 has a larger MSPE} \\ = 0, & \text{models 1,2 have the same MSPE} \end{cases} \end{aligned}$$

Under the null of equal forecast accuracy for the two models, $\rho_{xz} = Ee_{1t}^2 - Ee_{2t}^2 = 0$, i.e. x_t and z_t are uncorrelated. Let r_{xz} is the sample version of ρ_{xz}, then $r_{xz}/\sqrt{(1-r_{xz}^2)/(H-1)} \sim t(H-1)$. Examine the sign of this t-statistic and the significance of the t-test.

4. The Diebold-Mariano(1995) test: (Even the first two assumptions $e_t \sim N(0, \delta^2)$ and $Ee_te_{t-s} = 0 (s \neq 0)$ are not required). Use a more general loss function of the forecast error $g(e_i)$ instead of the quadratic one e_i^2. Let

$$\bar{d} = \frac{1}{H}\sum_{i=1}^{H} d_i = \frac{1}{H}\sum_{i=1}^{H}(g(e_{1i}) - g(e_{2i})),$$

which from CLT is asymptotically normally distributed: $\bar{d}/\sqrt{var(\bar{d})} \sim N(0,1)$ under the hypothsis that there is equal forecast accuracy. If $\{d_i\}$ is serially uncorrelated (conduct CDF, PACF and Q-statistic test to $\{d_i\}$),

$$\frac{\bar{d}}{\sqrt{\gamma_0}} \equiv \frac{\bar{d}}{\sqrt{\sum_{i=1}^{H}(d_i - \bar{d})^2/(H-1)}} \sim t(H-1).$$

If $\{d_i\}$ is serially correlated and $(\gamma_1, \cdots, \gamma_q) \neq 0$, where γ_i is the i-th sample autocovariance of $\{d_t\}$, then (Harvey et al (1998))

$$DM \equiv \bar{d}/\sqrt{(\gamma_0 + 2\gamma_1 + \cdots + 2\gamma_q)/(H-1)} \sim t(H-1), \quad \text{(1-step-ahead)}$$

$$\begin{aligned} DM &\equiv \bar{d}/\sqrt{(\gamma_0 + 2\gamma_1 + \cdots + 2\gamma_q)/(H+1-2j+H^{-1}j(j-1))} \sim t(H-1) \\ &\sim t(H-1), (j\text{-step-ahead}). \end{aligned}$$

- **Seasonality**: Forecasts that ignore seasonality will have a high variance, even in using the deseasonalized or seasonally adjusted data. In practice, the seasonal pattern will interact with the nonseasonal pattern in the data, making identification difficult. The ACF and PACF for a combined seasonal/nonseasonal process will reflect both elements. There are two methods to introduce the seasonal effect: additive seasonality and multiplicative seasonality. For example, the followings are additive specifications

$$y_t = a_1 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_4 \varepsilon_{t-4}$$
$$y_t = a_1 y_{t-1} + a_4 y_{t-4} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$$

while

$$(1 - a_1 L) y_t = (1 + \beta_1 L)(1 + \beta_4 L^4) \varepsilon_t$$
$$(1 - a_1 L)(1 - a_4 L^4) y_t = (1 + \beta_1 L) \varepsilon_t$$

are multiplicative specifications. Oftentimes strong seasonality and nonstationarity are found in the economic data. The ACF for the data is similar to that with no seasonality, but the spikes at lags s, $2s$, ...do not exhibit rapid decay.

1. First, seasonally difference the data and check the ACF of the resultant series. If the ACF shows a nonstationary process, the seasonally differenced data need to be further first differenced, i.e. apply the first difference to the seasonally differenced data, e.g.
$$(1 - L)(1 - L^4) y_t = (1 - L)(y_t - y_{t-4}) = (y_t - y_{t-4}) - (y_{t-1} - y_{t-5}).$$
2. Second, use the ACF and PACF to identify potential models. Try to estimate models with low-order nonseasonal ARMA coefficients. Allow the appropriate form of seasonality (additive or multiplicative) to be determined by the various diagnostic statistics.

- $ARIMA(p, d, q)(P, D, Q)_s : p, q =$the order of the nonseasonal ARMA coefficients,

$$\begin{aligned} d &= \text{number of nonseasonal differences} \\ P &= \text{number of multiplicative autoregressive coefficients} \\ D &= \text{number of seasonal difference} \\ Q &= \text{number of multiplicative moving-average coefficients} \\ s &= \text{seasonal period} \end{aligned}$$

  e.g. $m_t = a_0 + a_1 m_{t-1} + (1+\beta_1 L)(1+\beta_4 L^4)\varepsilon_t$ is an $ARIMA(1, 1, 0)(0, 1, 1)_4$ about y_t; $m_t = (1 + \beta_1 L)(1 + \beta_4 L^4)\varepsilon_t$ is an $ARIMA(0, 1, 1)(0, 1, 1)_4$ about y_t, where $m_t = y_t - y_{t-1}$.

An empirical example: ARMA Model Selection for the Producer Price Index (you are required to finish exercises 11 and 12 in Walter Enders(Chapter 2)) (see ex3 in notes).