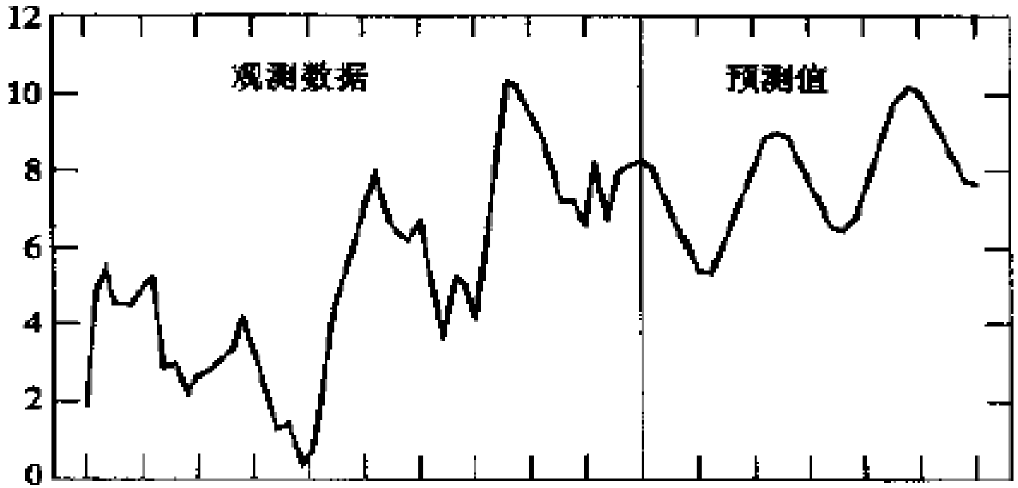
CH1 : Stationary Autoregressive Process

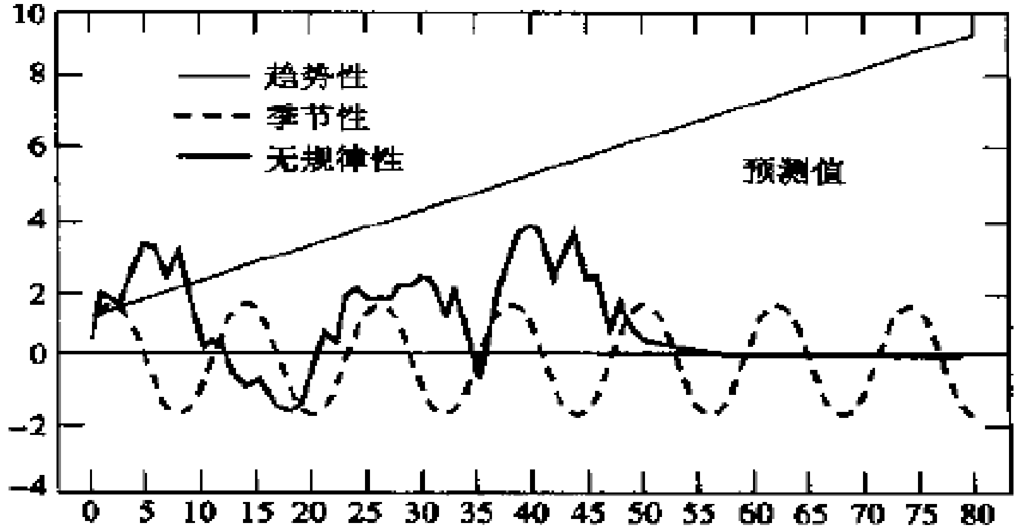
Difference Equation
Stationary Processes
Autocorrelation Function (ACF)
Partial Autocorrelation Function (PACF)
Estimation
Forecast

Difference Equation

- Express the value of a variable as a function of its own lagged values, time, and other variables.
- The trend and seasonal terms are both functions of time.
- The irregular term is a function of its owns lagged value and of the stochastic variables.
- Time-series Econometrics: estimation of difference equations containing stochastic components.

Difference Equation





Difference Equation: Examples

- Random walk: $y_t = y_{t-1} + u_t$
- Reduced-form and Structural equation:

$$\begin{cases} y_t = c_t + i_t \\ c_t = \alpha y_{t-1} + \varepsilon_{ct},\ 0 < \alpha < 1 \\ i_t = \beta(c_t - c_{t-1}) + \varepsilon_{it},\ \ \beta > 0 \end{cases}$$

$$y_t = \alpha(1+\beta)y_{t-1} - \alpha\beta y_{t-2} + (1+\beta)\varepsilon_{ct} + \varepsilon_{it} - \beta\varepsilon_{ct-1}$$

- Error-Correction:

$$s_{t+2} = s_{t+1} - \alpha(s_{t+1} - f_t) + \varepsilon_{s,t+2} \quad \alpha > 0$$

$$f_{t+1} = f_t + \beta(s_{t+1} - f_t) + \varepsilon_{f,t+2} \quad \beta > 0$$

Difference Equation: Constant Coefficients

- Linear parametric difference equation:

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + x_t$$

where n is the **order** of the difference equation; x_t is the **forcing** process that can be any function of time, current and lagged values of other variables or stochastic disturbances.

- A modified version:

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^{n} a_i y_{t-i} + x_t \qquad \gamma = a_1 - 1$$

- A solution: a function of the elements of $\{x_t\}$ and t or some given values of the $\{y_t\}$ (initial conditions).

Solve Difference Equation: AR(1) (1/4)

- Iterative method:
- Example 1: $y_t = a_0 + a_1 y_{t-1} + u_t$

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i u_{t-i}. \quad \text{If } y_0 \text{ is known}$$

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

If y_{-m-1} is known

If $|a_1| < 1$, as $m \rightarrow \infty$.

$$\text{a special solution} = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i u_{t-i}.$$

Solve Difference Equation: AR(1)   (2/4)

- An alternative method:

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

- Homogeneous equation: $y_t = a_1 y_{t-1}$

  $y_t = Aa_1^t$ is the **homogeneous solution**

- General Solution: If a_1 is not equal to 1

$$y_t = Aa_1^t + \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i u_{t-i}.$$

- So we can calculate A to obtain a special solution from the general solution by a given initial condition, e.g. at t=0, $y=y_0$.

Solve Difference Equation: AR(1)   (3/4)

- Obtain a special solution satisfying the initial condition y_0 :

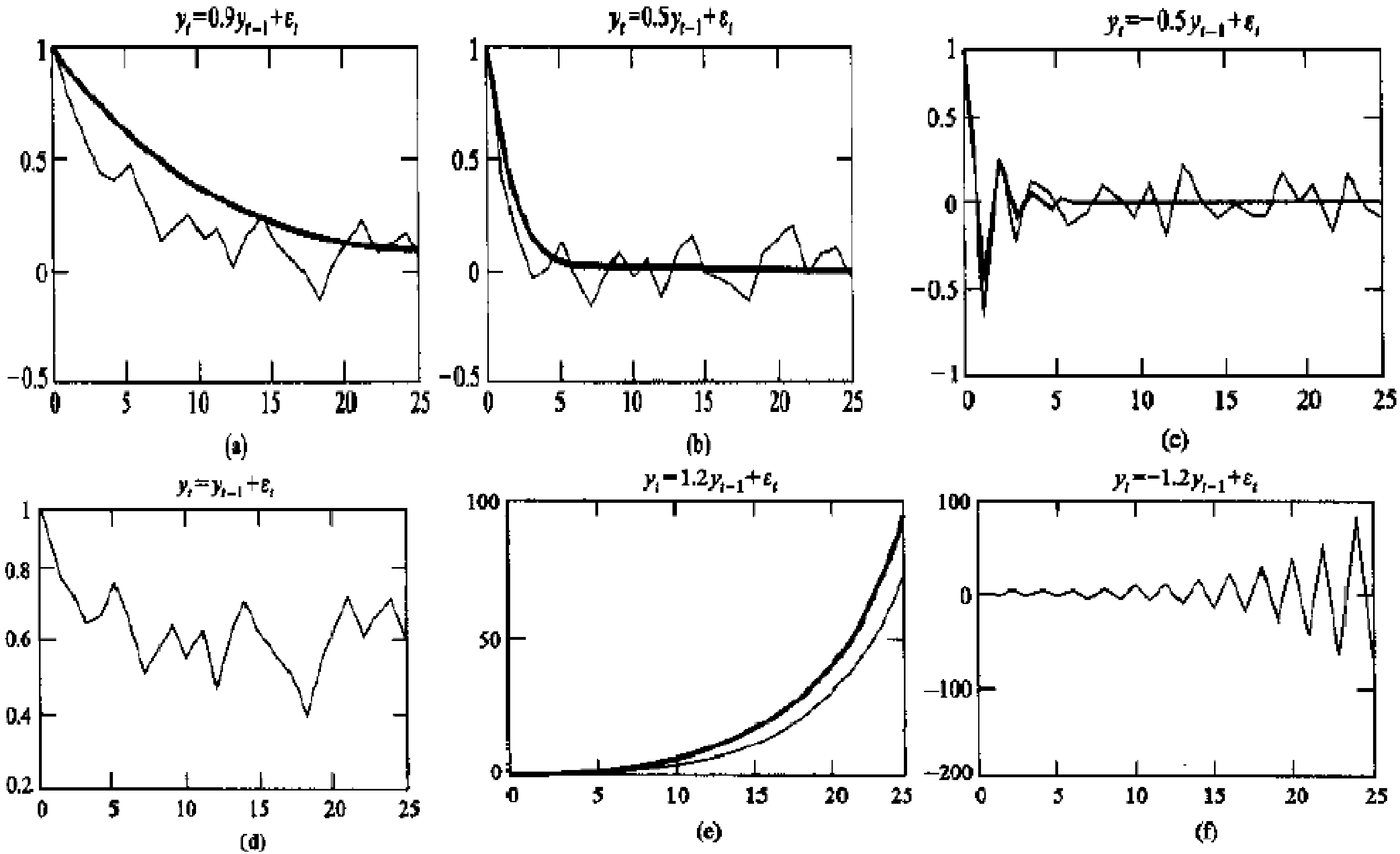
$$\text{If } |a_1| \neq 1, \quad y_t = \left(y_0 - \frac{a_0}{1-a_1} \right) a_1^{\,t} + \frac{a_0}{1-a_1} + \sum_{i=0}^{t-1} a_1^{\,i} u_{t-i}$$

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

$$\text{If } a_1 = 1, \quad y_t = a_0 t + y_0 + \sum_{i=1}^{t} u_i$$ （叠代至t=0时止）

- 在 $|a_1| \neq 1$ 时，也可由通解方程和初始条件确定 $A = y_0 - \frac{a_0}{1-a_1} - \sum_{i=0}^{\infty} a_1^{\,i} u_{-i}$，再将A代回通解方程，解出$y_t$。
- The solution shows that, when $|a_1| >= 1$, each disturbance has a permanent non-decaying effect on the value of y_t.
- Stability Condition: $\boxed{|a_1| < 1}$

Solve Difference Equation: AR(1)时间路径 (4/4)



The Cobweb Model: AR(1)

$$d_t = a - \gamma p_t \qquad \gamma > 0$$

$$s_t = b + \beta p_t^* + u_t \quad \beta > 0$$

$$s_t = d_t$$

$$\boxed{p_t^* = p_{t-1}}$$

$$\boxed{p_t = \frac{a-b}{\gamma} - \frac{\beta}{\gamma} p_{t-1} - \frac{1}{\gamma} u_t}$$

$$p_t = \frac{a-b}{\gamma+\beta} + \left(-\frac{\beta}{\gamma}\right)^t \left(p_0 - \frac{a-b}{\gamma+\beta}\right) - \frac{1}{\gamma} \sum_{i=0}^{t-1} \left(-\frac{\beta}{\gamma}\right)^i u_{t-i}$$

If the supply curve is steeper than the demand curve, i.e. $\beta / \gamma < 1$,

then the system is stable.

Solve Difference Equation: AR(2) (1/4)

- AR(2): $y_t = a_1 y_{t-1} + a_2 y_{t-2} + u_t,$
- Homogeneous Equation:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2}.$$

- Assume $y_t = A\lambda^t$ and obtain the characteristic equation:

$$\lambda^2 - a_1\lambda - a_2 = 0.$$

$$\lambda_1, \lambda_2 = \left(a_1 \pm \sqrt{a_{\mathbf{1}}^2 + 4a_2}\right)/2.$$

Solve Difference Equation: AR(2) (2/4)

- Homogeneous Solution:

according as $a_1^2 + 4a_2 > 0, = 0$ and < 0

$$y_t^h = A_1\lambda_1^t + A_2\lambda_2^t,$$

$$y_t^h = (A_1 + A_2 t)\lambda^t,\ \lambda = \lambda_1 = \lambda_2$$

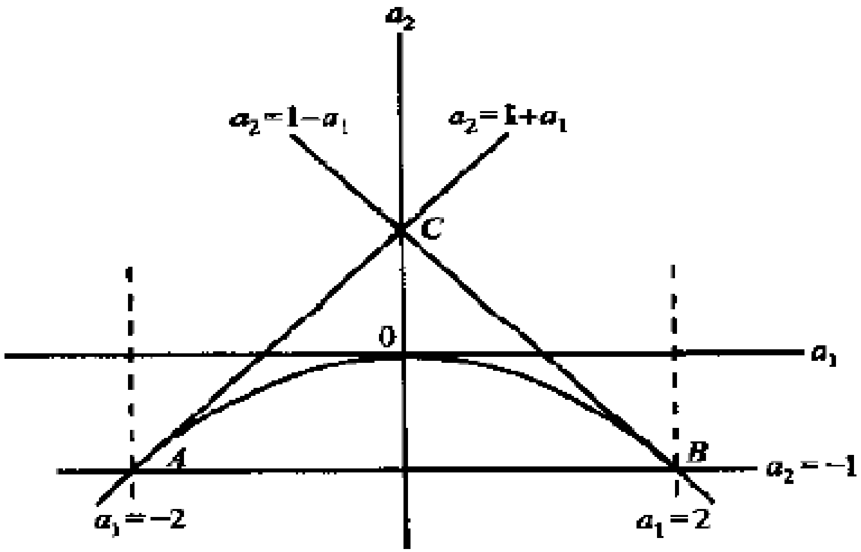
$$y_t^h = A_1 r^t \cos(\theta t + A_2),\ r = \sqrt{-a_2}, \theta = \arg tg\left(\sqrt{-a_1^2 - 4a_2}/a_1\right).$$

- Stability Condition:

$$a_2 + a_1 < 1$$

$$a_2 - a_1 < 1$$

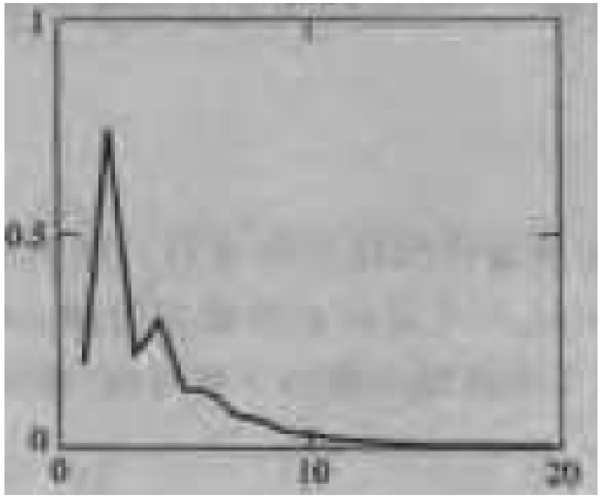
$$a_2 > -1$$



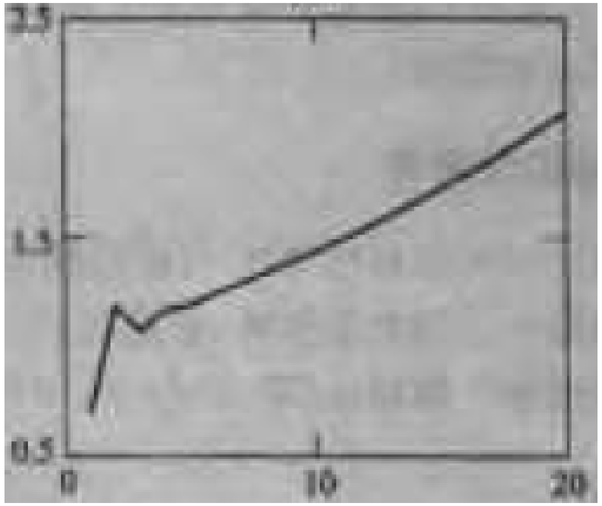
Solve Difference Equation: AR(2) (3/4)

- Example 1: $y_t=0.2y_{t-1}+0.35y_{t-2}$

  The two characteristic roots are 0.7, -0.5.



$y_t=0.7^t+(-0.5)^t$



$y_t=1.037^t+(-0.337)^t$

- Example 2: $y_t=0.7y_{t-1}+0.35y_{t-2}$

  The two characteristic roots are 1.037, -0.337.

Solve Difference Equation: AR(2)   (4/4)

- General Solution for $y_t = a_1 y_{t-1} + a_2 y_{t-2} + u_t$,

$$\begin{aligned} y_t &= y_t^p + y_t^h \\ &= y_t^d + y_t^s + y_t^h \\ &= \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i} + y_t^h \end{aligned}$$

Solve Difference Equation: AR(n) 1/2

- Difference equation: $$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + x_t$$
- First, solve homogeneous solution of $y_t = \sum_{i=1}^{n} a_i y_{t-i}$:

$$\lambda^t - \sum_{i=1}^{n} a_i \lambda^{t-i} = 0 \qquad \lambda^{n-t}\left(\lambda^t - \sum_{i=1}^{n} a_i \lambda^{t-i}\right) = 0$$

$$\lambda^n - \sum_{i=1}^{n} a_i \lambda^{n-i} = 0, \quad i.e. \quad \lambda^n - a_1\lambda^{n-1} - a_2\lambda^{n-2} - \cdots - a_n = 0$$

The linear combination: $A_1\lambda_1^t + A_2\lambda_2^t + \cdots + A_n\lambda_n^t$

- At least one characteristic root equals one if $a_1 + a_2 + \ldots + a_n = 1$.

Solve Difference Equation: AR(n) 2/2

- Particular solution for

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + x_t$$

If x_t is deterministic, e.g.

$x_t = 0$, then set $y_t^d = c$ (further if $a_1 + \cdots + a_n = 1$, set $y_t^d = ct$);

$x_t = bd^{rt}$, then set $y_t^d = c_0 + c_1 d^{rt}$;

$x_t = a_0 + bt^k$, then set $y_t^d = c_0 + c_1 t + \cdots + c_{k-1} t^{k-1} + c_k t^k$;

If $x_t = u_t$ is stochastic, set $y_t^s = \sum_{i=0}^{\infty} \alpha_i u_{t-i}$.

If x_t includes both deterministic and stochastic terms,

insert $y_t^p = y_t^d + y_t^s$ into the difference equation and solve c's and α's by the method of undetermined coefficients.

Lag operator

- Linear operator

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.

Some Processes

- 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,

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

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

Stationary Process and Autocorrelation

- $\{y_t\}$ **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

- autocorrelation between y_t and y_{t-s}:

$$\rho_s \equiv \frac{\gamma_s}{\gamma_0} = \frac{Cov(y_t, y_{t-s})}{Cov(y_t, y_t)} = \frac{Cov(y_t, y_{t-s})}{\sigma_y^2}, \ \rho_0 = 1.$$

Conditions for AR(1) Stationary Process

- AR(1): $\boxed{y_t = a_0 + a_1 y_{t-1} + \varepsilon_t,}$ where a_1 is not equal to 1.
- For any given initial condition y_0 (suppose not random)

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

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

- **y_0 can not be fixed! It should be a random variable.**
- Restrictions: $\boxed{y_t^h \equiv 0,}$ i.e. A=0; y_0 is a random variable;

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

$$y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}. \qquad Ey_t = \frac{a_0}{1-a_1} = Ey_{t-s},$$

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

Conditions for ARMA(2,1) Stationary Process

- ARMA(2,1): $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$ where ε_t is white-noise.
- One Restriction on homogeneous solution: $\boxed{y_t^h \equiv 0,}$
- Try a particular solution: $\boxed{y_t = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i}}$

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

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

- Hence: $y_t = \varepsilon_t + (a_1 + \beta_1)\varepsilon_{t-1} + \sum_{i=2}^{\infty} \alpha_i \varepsilon_{t-i},$ where

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

- If the characteristic roots of ARMA(2,1) lie within the unit circle, $\{\alpha_i\}$ is a convergent sequence and $\{y_t\}$ is **stationary**:

$$Ey_t = 0 = Ey_{t-s}, \forall t, s, \quad Var(y_t) = Var(y_{t-s}) = \sigma^2 \sum_{i=0}^{\infty} \alpha_i^2, \forall t, s, \quad Cov(y_t, y_{t-s}) = \sigma^2 \sum_{j=0}^{\infty} \alpha_{s+j} \alpha_j.$$

A Theorem about Stationarity

- **Theorem** Let the roots of the polynomial equation

$$\lambda^p + a_1\lambda^{p-1} + a_2\lambda^{p-2} + \cdots + a_p = 0$$

be less than one in absolute value, where $a_p \neq 0$, and let the weights $\{\alpha_i\}_{i=0}^{\infty}$ be defined by the solution of the homogeneous difference equation

$$\alpha_i + a_1\alpha_{i-1} + a_2\alpha_{i-2} + \cdots + a_p\alpha_{i-p} = 0, j = p, p+1, \cdots$$

subject to the initial condition $\alpha_0, \alpha_1, \cdots, \alpha_{p-1}$. Let $\{e_t\}$ be a sequence of uncorrelated $(0, \sigma^2)$ random variables. Then the sequence $\{y_t\}$, where $y_t = \sum_{i=0}^{\infty} \alpha_i e_{t-i}$, is a stationary moving average process. Moreover, $\{y_t\}$ satisfies the stochastic difference equation

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} = e_t$$

for almost every realization of $\{e_t\}$

Note: The proof requires the following theorems:

1. If the characteristic roots of the above difference equation of α_i are less than one in absolute value, the sequence $\{\alpha_i\}_{i=0}^{\infty}$ is absolutely summable, i.e. $\sum_{i=0}^{\infty} |\alpha_i| < \infty$.

2. If $\{\alpha_i\}_{i=0}^{\infty}$ is absolutely summable, then $\sum_{i=0}^{\infty} {\alpha_i}^2 < \infty$ and $\sum_{i=0}^{\infty} \alpha_i\alpha_{i+s} < \infty$.

Conditions for $MA(\infty)$ Stationary Process

$MA(\infty): x_t = \sum_{i=0}^{\infty} \beta_i \varepsilon_{t-i},\ \beta_0 = 1.$

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

If $\sum_{i=0}^{\infty} \beta_i \beta_{i+s} < \infty\ \forall s,\ MA(\infty)$ will be stationary.

$\boxed{MA(q) \text{ is always stationary for any finite } q.}$

Conditions for AR(p) Stationary Process

- 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

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

since the series $\{\alpha_i\}$ solve the difference equation $\alpha_i - \sum_{j=1}^{p} a_j \alpha_{i-j} = 0$.

$$Ey_t = Ey_{t-s} = \frac{a_0}{1 - \sum_{i=1}^{p} a_i}, \qquad Var(y_t) = \sum_{i,j=0}^{\infty} \alpha_i \alpha_j \sigma^2 \delta_{i,j} = \sigma^2 \sum_{i=0}^{\infty} \alpha_i^2 < \infty,$$

$$Con(y_t, y_{t-s}) = \sigma^2 \sum_{j=0}^{\infty} \alpha_{j+s} \alpha_j < \infty$$

Conditions for ARMA(p,q) Stationary Process

- 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 1/4

- The autocorrelation function $\rho_s = \frac{\gamma_s}{\gamma_0} = \frac{Cov(y_t, y_{t-s})}{Var(y_t)}$
- **ACF for $AR(1)$ process:** $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$.

$$\gamma_0 = Var(y_t) = \sigma^2/(1 - a_1^2),$$

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

$$\boxed{\rho_0 = 1, \ \rho_s = a_1^s \text{ for } s \geq 1.}$$

$\boxed{\text{converge to 0 geometrically}}$ as $s \to \infty$, provided $|a_1| < 1$.

Autocorrelation function ACF 2/4

- **ACF for $AR(2)$ process:** $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$.

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

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

ρ_s satisfy the difference equation $\rho_s = a_1\rho_{s-1} + a_2\rho_{s-2},\ s > 0$

is stationary since the characteristic roots lie inside the unit circle.

$\boxed{\text{The ACF converge to 0 (directly or oscillatorily) as } s \to \infty.}$

Autocorrelation function ACF 3/4

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

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

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

$\boxed{\text{The ACF } \rho_s \ (s = 1, 2, \ldots) \text{ has one spike } (\rho_1 \neq 0) \text{ and then cuts to } 0.}$

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

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

$\boxed{\text{The ACF has two spikes and then cuts to } 0 : \rho_s = 0, \ s \geq 3.}$

- **ACF for $MA(q)$** : $\boxed{\text{has } q \text{ spikes and then cuts to}}\boxed{0 : \rho_s = 0, \ s \geq q+1.}$

Autocorrelation function ACF 4/4

- **ACF for** $ARMA(1,1)$ **process:** $y_t = a_1 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$.

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

$$\gamma_0 = \frac{1+\beta_1^2+2a_1\beta_1}{1-a_1^2}\sigma^2 \text{ and } \gamma_1 = \frac{(1+a_1\beta_1)(a_1+\beta_1)}{1-a_1^2}\sigma^2$$

$$\rho_1 = \frac{(1 + a_1\beta_1)(a_1 + \beta_1)}{1 + \beta_1^2 + 2a_1\beta_1} \qquad \rho_s = a_1 \rho_{s-1}, \; s \geq 2,$$

ρ_s converge to 0 geometrically as $s \to \infty$ provided $|a_1| < 1$.

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

ρ_s $(s > q)$ satisfy (Note that $\rho_s = \rho_{-s}$): $\boxed{\rho_s = a_1 \rho_{s-1} + \cdots + a_p \rho_{s-p}, \; s \geq q+1.}$

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 1/2

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

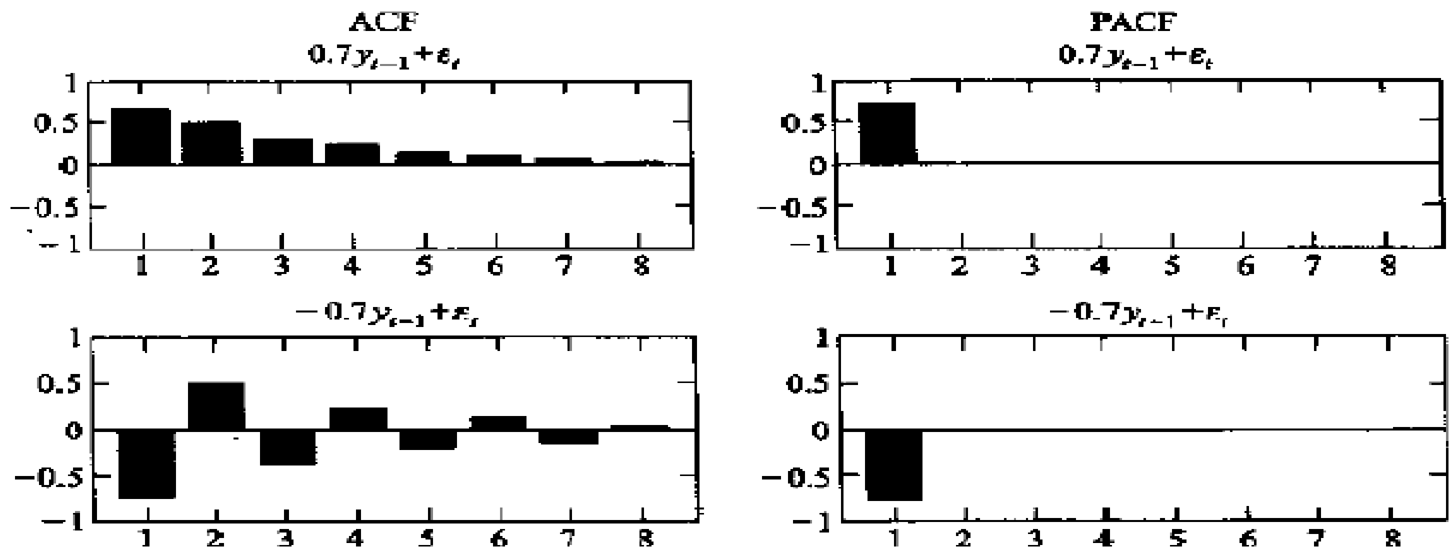
  $\boxed{\text{then } \phi_{11} = PACF \text{ between } y_t \text{ and } y_{t-1}.}$

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

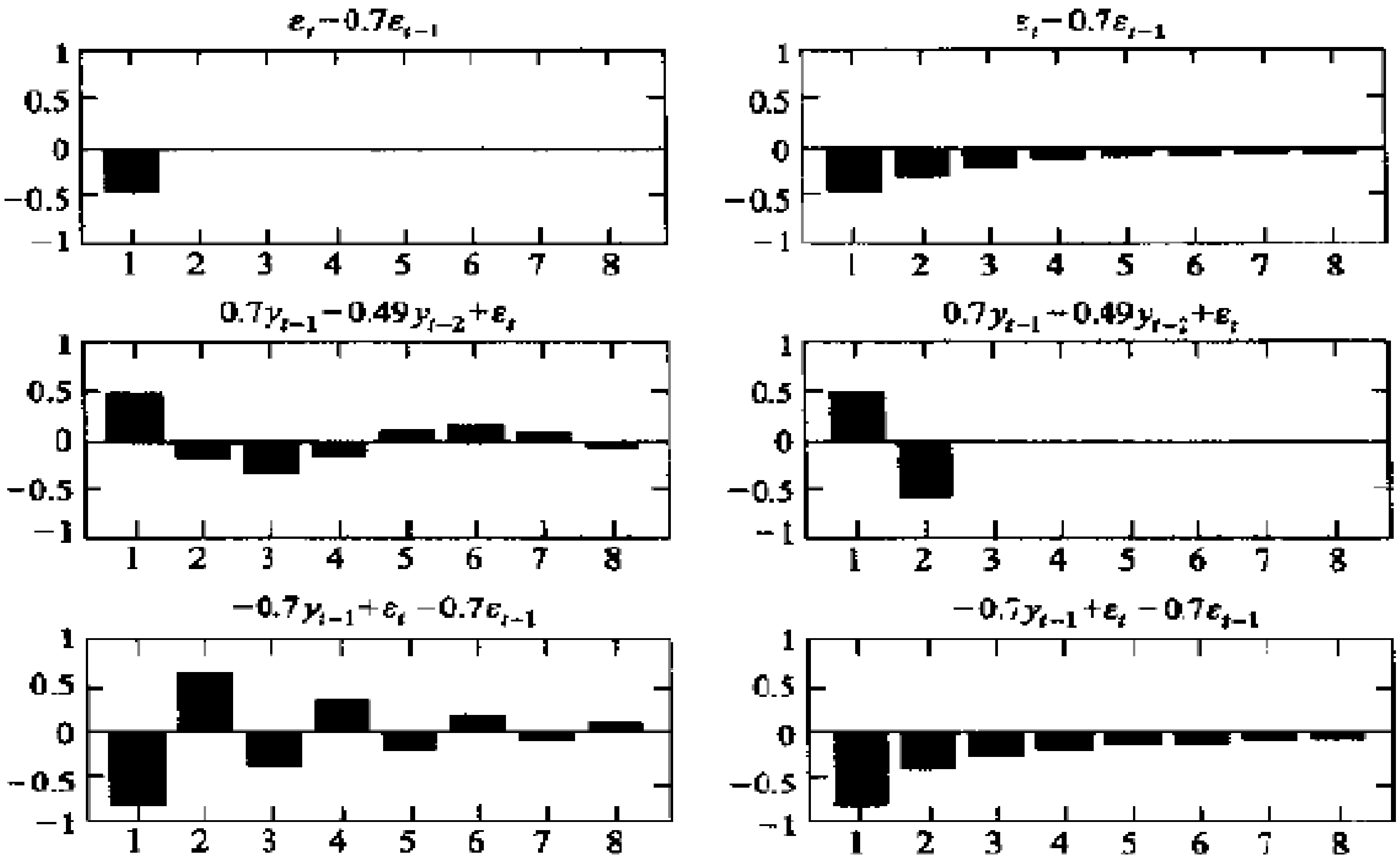
  $\boxed{\text{then } \phi_{22} = PACF \text{ between } y_t \text{ and } y_{t-2}.}$
- AR(P): no direct correlation between y_t and y_{t-s} for $s > p$, i.e. $\phi_{ss} = 0$ for $s \geq p+1$.
- MA(q) results in infinite-order AR representation. The PACF exhibit a decay.
- ARMA(p,q): ACF begin to decay 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).

ACF and PACF 1/2

- 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.



ACF and PACF 2/2



Sample ACF and PACF 1/2

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

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

The sample PACF $\hat{\phi}_{ss}$ is the estimator of ϕ_{ss} in

$$y_t^* = \phi_{s1} y_{t-1}^* + \cdots + \phi_{ss} y_{t-s}^* + e_t.$$

where y* is the subtracted mean of the series from each observation

Sample ACF and PACF 2/2

recursively

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

where $\hat{\phi}_{s,j} = \hat{\phi}_{s-1,,j} - \hat{\phi}_{ss}\hat{\phi}_{s-1,s-j}$, $j = 1, 2, \cdots, s-1$.

If the sample value of r_s is zero (corresponding to $\rho_s = 0$), the process is expected to be $MA(s-1)$; if the estimated value of $\ddot{\phi}_{ss}$ is zero (corresponding to $\phi_{ss} = 0$), the process is expected to be $AR(s-1)$.

Hence a test for AR or MA.

Testing for AR and MA 1/2

- 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.

Testing for AR and MA 2/2

(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.

Model Selection of Lags

- Use AIC or 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 are, the better is the selected model**

Since $\ln T > 2$, the SBC will always select a more parsimonious model than the AIC.

Model Estimation: AR(1) 1/4

$$Y_t = a_0 + a_1 Y_{t-1} + \varepsilon_t, \text{where } \varepsilon_t \sim i.i.d.N(0,\sigma^2),\ \Theta = (a_0, a_1, \sigma^2)',\ |a_1| < 1.$$

$$Y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \qquad EY_t = \frac{a_0}{1-a_1} \equiv \mu \qquad Var(Y_t) = \frac{\sigma^2}{1-a_1^2}$$

1) Exact MLE $Y_1 \sim N\left(\mu, \frac{\sigma^2}{1-a_1^2}\right)$ $(Y_2|Y_1) \sim N\left(a_0 + a_1 Y_1, \sigma^2\right)$

$$f_{Y_2,Y_1}(y_2, y_1; \Theta) = f_{Y_2|Y_1}(y_2|y_1;\Theta) f_{Y_1}(y_1;\Theta).$$

$$\begin{aligned} f_{Y_3,Y_2,Y_1}(y_3,y_2,y_1;\Theta) &= f_{Y_3|Y_2,Y_1}(y_3|y_2,y_1;\Theta) f_{Y_2,Y_1}(y_2,y_1;\Theta) \\ &= f_{Y_3|Y_2,Y_1}(y_3|y_2,y_1;\Theta) f_{Y_2|Y_1}(y_2|y_1;\Theta) f_{Y_1}(y_1;\Theta). \end{aligned}$$

The joint density of the sample:

$$f_{Y_T,Y_{T-1},\cdots,Y_1}(y_T, y_{T-1}, \cdots, y_1;\Theta) = f_{Y_1}(y_1;\Theta) \prod_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\Theta)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2/(1-a_1^2)}} \exp\left(-\frac{(y_1-\mu)^2}{2\sigma^2/(1-a_1^2)}\right) \prod_{t=2}^{T}\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - a_0 - a_1 y_{t-1})^2}{2\sigma^2}\right)\right]$$

$$= \sqrt{1-a_1^2}\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^T \exp\left(-\frac{(y_1-\mu)^2}{2\sigma^2/(1-a_1^2)} - \sum_{t=2}^{T} \frac{(y_t - a_0 - a_1 y_{t-1})^2}{2\sigma^2}\right).$$

Model Estimation: AR(1) 2/4

- The log likelihood of the sample:

$$\ln L = \frac{1}{2}\log(1-a_1^2) - \frac{T}{2}\log(2\pi) - \frac{(y_1 - a_0/(1-a_1))^2}{2\sigma^2/(1-a_1^2)} - \sum_{t=2}^{T}\frac{(y_t - a_0 - a_1 y_{t-1})^2}{2\sigma^2}.$$

Denote $Y = (Y_1, Y_2, \cdots, Y_T)'$. Then $EY = \mu\mathbf{1}$, where $\mathbf{1} = (1, 1, \cdots, 1)'$ is a $T \times 1$ matrix, and

$$Y = (Y_1, Y_2, \cdots, Y_T)' \sim N(\mu\mathbf{1}, \Omega).$$

$$\Omega \equiv E\left[(Y-\mu\mathbf{1})(Y-\mu\mathbf{1})'\right]$$

$$= \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{T-2} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \gamma_0 \end{pmatrix} = \frac{\sigma^2}{1-a_1^2}\begin{pmatrix} 1 & a_1 & \cdots & a_1^{T-1} \\ a_1 & 1 & \cdots & a_1^{T-2} \\ \vdots & \vdots & \cdots & \vdots \\ a_1^{T-1} & a_1^{T-2} & \cdots & 1 \end{pmatrix}$$

$$f_Y(y;\Theta) = (2\pi)^{-T/2}|\Omega^{-1}|^{1/2}\exp\left(-\frac{1}{2}(Y-\mu\mathbf{1})'\Omega^{-1}(Y-\mu\mathbf{1})\right)$$

$$\ln L = -\frac{T}{2}\log(2\pi) + \frac{1}{2}\log|\Omega^{-1}| - \frac{1}{2}(Y - a_0/(1-a_1)1)'\Omega^{-1}(Y - a_0/(1-a_1)1)$$

Model Estimation: AR(1) 3/4

2) Conditional MLE Take y_1 as deterministic

$$f_{Y_T,Y_{T-1},\cdots,Y_2|Y_1}(y_T,y_{T-1},\cdots,y_2|y_1;\Theta)=\prod_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\Theta)$$

$$=\prod_{t=2}^{T}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_t-a_0-a_1y_{t-1})^2}{2\sigma^2}\right)\right]$$

$$=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{T-1}\exp\left(-\sum_{t=2}^{T}\frac{(y_t-a_0-a_1y_{t-1})^2}{2\sigma^2}\right).$$

$$\boxed{\ln L=-\frac{T-1}{2}\log(2\pi)-\frac{T-1}{2}\log(\sigma^2)-\sum_{t=2}^{T}\frac{(y_t-a_0-a_1y_{t-1})^2}{2\sigma^2}.}$$

equivalent to minimization of $\sum_{t=2}^{T}(y_t-a_0-a_1y_{t-1})^2,$

$$\hat{\sigma}^2=\frac{1}{T-1}\sum_{t=2}^{T}(y_t-\hat{a}_0-\hat{a}_1y_{t-1})^2.$$

Model Estimation: AR (1) 4/4

Compare exact MLE and conditional MLE:

$$f_{Y_1}(y_1;\Theta)\prod_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\Theta) \approx \prod_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\Theta),$$

$$f_{Y_T,Y_{T-1},\cdots,Y_1}(y_T,y_{T-1},\cdots,y_1;\Theta) \approx f_{Y_T,Y_{T-1},\cdots,Y_2|Y_1}(y_T,y_{T-1},\cdots,y_2|y_1;\Theta).$$

Both exact MLE and conditional MLE have the same large-sample distribution.

When $|a_1| > 1$, conditional MLE continues to provide consistent estimates, whereas exact MLE does not, since

$$\frac{1}{\sqrt{2\pi\sigma^2/(1-a_1^2)}}\exp\left(-\frac{(y_1-\mu)^2}{2\sigma^2/(1-a_1^2)}\right)$$

does not accurately describe the density $f_{Y_1}(y_1;\Theta)$ of Y_1. Hence, in most applications the parameters of autoregression are estimated by conditioanl MLE (OLS) rather than exact MLE.

Model Estimation: AR (p)

$$Y_t = a_0 + a_1 Y_{t-1} + \cdots + a_p Y_{t-p} + \varepsilon_t$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, $\Theta = (a_0, a_1, \cdots, a_p, \sigma^2)'$.

All the characteristic roots lie within the unit circle.

$\boxed{\text{Conditional likelihood for the sample is}}$

$$f_{Y_T, Y_{T-1}, \cdots, Y_{p+1} | Y_p, \cdots, Y_1}(y_T, y_{T-1}, \cdots, y_{p+1} | y_p, y_{p-1}, \cdots, y_1; \Theta)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{t=p+1}^{T} f_{Y_t | Y_{t-1}, \cdots, Y_{t-p}}(y_t | y_{t-1}, \cdots, y_{t-p}; \Theta)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \sum_{t=p+1}^{T} \frac{(y_t - a_0 - a_1 Y_{t-1} - \cdots - a_p Y_{t-p})^2}{2\sigma^2} \right).$$

Model Estimation: MA (1) 1/3

$Y_t = a_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1}$, where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, $\Theta = (a_0, \theta_1, \sigma^2)'$.

$$EY_t = a_0, \quad Var(Y_t) = \sigma^2(1 + \theta_1^2).$$

1) Exact MLE

Denote $Y = (Y_1, Y_2, \cdots, Y_T)'$.

$$\sim N(a_0 \mathbf{1}, \Omega).$$

$$\boxed{f_Y(y;\Theta) = (2\pi)^{-T/2}|\Omega^{-1}|^{1/2} \exp\left(-\frac{1}{2}(Y - a_0\mathbf{1})'\Omega^{-1}(Y - a_0\mathbf{1})\right)}$$

$$\Omega \equiv E\left[(Y - a_0\mathbf{1})(Y - a_0\mathbf{1})'\right]$$

$$= \sigma^2 \begin{pmatrix} 1+\theta_1^2 & \theta_1 & 0 & \cdots & 0 & 0 \\ \theta_1 & 1+\theta_1^2 & \theta_1 & \cdots & 0 & 0 \\ 0 & \theta_1 & 1+\theta_1^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\theta_1^2 & \theta_1 \\ 0 & 0 & 0 & \cdots & \theta_1 & 1+\theta_1^2 \end{pmatrix}$$

Model Estimation: MA (1) 2/3

2) Conditional MLE Assume $\varepsilon_0 = 0$.

$(Y_1 | \varepsilon_0 = 0) \sim N(a_0, \sigma^2)$ $(Y_2 | Y_1, \varepsilon_0 = 0) \sim N(a_0 + \theta\varepsilon_1, \sigma^2)$, where $\varepsilon_1 = y_1 - a_0$

$$f_{Y_2,Y_1|\varepsilon_0=0}(y_2, y_1 | \varepsilon_0 = 0; \Theta) = f_{Y_2|Y_1,\varepsilon_0=0}(y_2 | y_1, \varepsilon_0 = 0; \Theta) \cdot f_{Y_1|\varepsilon_0=0}(y_1 | \varepsilon_0 = 0; \Theta)$$

$$f_{Y_T,Y_{T-1},\cdots,Y_1|\varepsilon_0=0}(y_T, y_{T-1}, \cdots, y_1 | \varepsilon_0 = 0; \Theta)$$
$$= f_{Y_T|Y_{T-1},\cdots,Y_1,\varepsilon_0=0}(y_T | y_{T-1}, \cdots, y_1, \varepsilon_0 = 0; \Theta) \cdot f_{Y_{T-1},\cdots,Y_1|\varepsilon_0=0}(y_{T-1}, \cdots, y_1 | \varepsilon_0 = 0;$$
$$= f_{Y_1|\varepsilon_0=0}(y_1 | \varepsilon_0 = 0; \Theta) \prod_{t=2}^{T} f_{Y_t|Y_{t-1},\cdots,Y_1,\varepsilon_0=0}(y_t | y_{t-1}, \cdots, y_1, \varepsilon_0 = 0; \Theta)$$
$$= \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^T \exp\left(-\sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2}\right),$$

where
$$\begin{aligned} \varepsilon_1 &= y_1 - a_0, \\ \varepsilon_t &= y_t - (a_0 + \theta_1 \varepsilon_{t-1}), \quad t \geq 2. \end{aligned}$$

Model Estimation: MA (1)   3/3

For any $\varepsilon_0 \neq 0$, when $|\theta_1| < 1$,

$$\varepsilon_0 = \sum_{i=0}^{\infty}(-\theta_1)^i(y_{-i} - a_0),$$

$$\varepsilon_t = \sum_{i=0}^{t-1}(-\theta_1)^i(y_{t-i} - a_0) + (-\theta_1)^t\varepsilon_0$$

$$\varepsilon_1 = y_1 - a_0 + (-\theta_1)\varepsilon_0$$

$$\varepsilon_t = y_t - (a_0 + \theta_1\varepsilon_{t-1}) + (-\theta_1)^t\varepsilon_0.$$

When $\varepsilon_0 = 0$, we obtain the previous case.

If $|\theta_1|$ is substantially less than unity,

the effects of imposing $\varepsilon_0 = 0$ will quickly die out

If $|\theta_1| > 1$, consquences of imposing $\varepsilon_0 = 0$ accumulate over time

the conditional MLE is not reasonable.

Model Estimation: MA (q)

$$Y_t = a_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \text{where } \varepsilon_t \sim i.i.d.N(0, \sigma^2)$$

$$EY_t = a_0, \quad Var(Y_t) = \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2).$$

Exact likelihood is

$$f_Y(y; \Theta) = (2\pi)^{-T/2} |\Omega^{-1}|^{1/2} \exp\left(-\frac{1}{2}(Y - a_0\mathbf{1})'\Omega^{-1}(Y - a_0\mathbf{1}) \right)$$

Conditional likelihood

$$f_{Y_T, Y_{T-1}, \cdots, Y_1 | \varepsilon_0, \varepsilon_{-1}, \cdots, \varepsilon_{-q+1} = 0}(y_T, y_{T-1}, \cdots, y_1 | \varepsilon_0, \varepsilon_{-1}, \cdots, \varepsilon_{-q+1} = 0; \Theta)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^T \exp\left(\sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2} \right) \qquad \text{where } \varepsilon_t = y_t - a_0 - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

Conditional MLE is reasonable only if all the inverse characteristic roots for the MA lie outside of the unit circle. $1 + \theta_1 z + \cdots + \theta_q z^q = 0$

Central Limit Theorem for Time Series

- Central Limit Theorem for Stationary Time Series.

- **Central Limit Theorem for a Martingale Difference Sequence:** Let $\{Y_t\}$ be a scalar martingale difference sequence with $\bar{Y}_T = (1/T)\sum_{t=1}^{T} Y_t$. Suppose that (a) $E(Y_t^2) = \sigma_t^2 > 0$ with $(1/T)\sum_{t=1}^{T} \sigma_t^2 \to \sigma^2 > 0$; (b) $E|Y_t|^r < \infty$ for some r>2 and all t; (c) $(1/T)\sum_{t=1}^{T} Y_t^2 \to \sigma^2$ in probability. Then

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_t \to N(0, \sigma^2).$$

Three-Stage Model Selection: Box-Jenkins 1/2

- Identification: Examine the time plot of the series (data), the ACF, and the PACF visually.
- Estimation: Fit the suggested models under stationarity and examine the estimates of parameters in the ARMA models. Select the model with parsimony, where Q-statistic, AIC, and SBC are used for model selection.
- 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.

Three-Stage Model Selection: Box-Jenkins 2/2

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

forward iteration

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

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

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

the j-step-ahead forecast error

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

$$E_t e_t(j) = 0 \quad Var[e_t(j)] = \sigma^2[1 + a_1^2 + \cdots + a_1^{2(j-1)}] \longrightarrow \sigma^2/(1 - a_1^2) \text{ as } j \to \infty.$$

The forecasts are unbiased, but the quality declines as j increases.

Forecast: ARMA(2,1)

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1}.$$

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

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

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

$\boxed{\text{out-of-sample forecasts}}$

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

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

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

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

Forecast: ARMA(p,q)

$$y_t = a_0 + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q}$$

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

Forecast Evaluation Methods 1/4

Fit the best $\neq$ Forecast the best.

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 $Var(v_t)$.

Forecast Evaluation Methods 2/4

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

Forecast Evaluation Methods 3/4

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

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

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.

Forecast Evaluation Methods 4/4

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

$$\bar{d} = \frac{1}{H}\sum_{i=1}^{H} d_i = \frac{1}{H}\sum_{i=1}^{H}(g(e_{1i}) - g(e_{2i})), \qquad \boxed{\bar{d}/\sqrt{var(\bar{d})} \sim N(0,1)}$$

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

If $\{d_i\}$ is serially correlated and $(\gamma_1, \cdots, \gamma_q) \neq 0$,

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

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