Model diagnostics and selection for ARMA models

(Updated Spring 2021)

Chaoyi Chen
Institute of MNB, Corvinus University of Budapest

Empirical Financial Econometrics

Outlines

- Model Diagnostics and Selection Test ⏩ Jump [Online Lecture (minor) + Self-study (major)]

Model Diagnostics and Selection Test

- Model selection Criteria
- Auto Covariance and Auto relation functions
- Covariance stationarity and the ACF
- Calculating population(theoretical) ACF
- White Noise ACF
- ACF FOR MA(1) Model
- ACF FOR AR(1) Model
- Estimated ACF
- PACF-Partial Autocorrelation Function
- Common Factors

Model selection Criteria

Intuition

- Intuition: Posit $y_t \sim ARMA(p, q)$
- if p and/or q too small
  - model may be wrong
  - estimates and forecasts biased
  - tests may over-rejected

Model selection Criteria

- Example: the true model is AR(2)

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

But we estimate an AR(1)

$$y_t = \hat{a}_0 + \hat{a}_1 y_{t-1} + \hat{u}_t \tag{2}$$

Model selection Criteria

- if p and/or q too big
  - model still correct
  - But needlessly estimate extra coefficients, whose true value is zero.
  - Lose degrees of freedom
  - Increase variance of parameter estimates and forecasts
  - Increase probability if type 2 error.

Model selection Criteria

- Example: the true model is AR(1)

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t \tag{3}$$

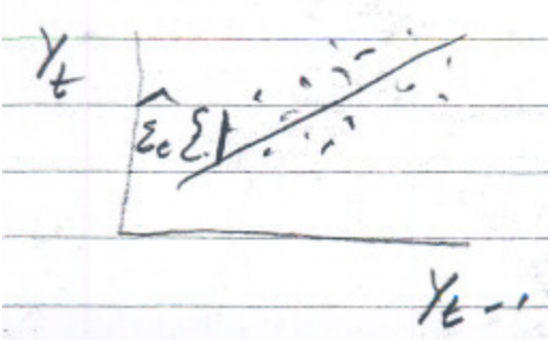
But we estimate an AR(1)

$$y_t = \hat{a}_0 + \hat{a}_1 y_{t-1} + \hat{a}_2 y_{t-2} + \hat{\varepsilon}_t \tag{4}$$

- Recall Sum of Squared Residuals:

$$SSR = \sum_{t=1}^{T} \hat{\varepsilon_t}^2$$

- a measure of how well the model fits the data.



The smaller the better.

Model selection Criteria

- **Temptation:** Use SSR or (R^2) to compare ARMA(p,q) models for different values of p and g to see which fits data best.
- **Problem:** Recall that the SSR always goes down(and R^2 goes up) as add regressors or as p increase or q increase.

  So following (C) would give us huge and terrible values of p and p.
- **Solution:** compare SSR but penalize larger model.

Model selection Criteria

- One such criteria is **Akaike Information Criteria(AIC)**

$$AIC = T \ln(\frac{SSR}{T}) + 2(\text{no parameters})$$

- Another is the **Schwartz Bayesian Criteria (SBC)**[1]

$$SBC = T \ln(\frac{SSR}{T}) + (\text{no parameters}) \ln(T)$$

- Smaller AIC or SBC suggests better model.

[1] a.k.a. **Bayesian Information Criteria (BIC)**

Model selection Criteria

- Practical Model selection using AIC or SBC.
- Choose a maximum values of p and q, say $\bar{p}$ and $\bar{q}$
- Estimate the ARMA(p,q) for each combination of p and q satisfying, $0 \leq p \leq \bar{p}$ and $0 \leq q \leq \bar{q}$. And record the AIC and SBC.
- Choose the (p,q) with the smallest AIC or SBC as your model.
- This can be done using a double loop

Model selection Criteria

Without using exact code, your program might have structure similar to:

```
AIC = zeros(p+1,q+1); % Matrix to store AIC results in
for p=1:pmax+1;% loop through AR lag orders
   for q=1:qmax+1; % loop through MA lag orders
      AIC(p,q) = ARMA_AIC(p-1,q-1);% calculate and store AIC values
   end; % end loop for p
end; % end loop for q
print AIC % print out results
```

Auto Covariance and Auto relation functions

- **Let's define:**

$$
\begin{aligned}
\gamma_0 &= var(y_t) \\
\gamma_1 &= cov(y_t, y_{t+1}) \\
&\vdots \\
\gamma_h &= cov(y_t, y_{t+h})
\end{aligned}
$$

- γ_h is the **auto-covariance function**,it is a function of h, which is the time between the two observations

Autocovariance and Autocorrelation functions

- **Now define**

$$\begin{aligned}
\rho_0 &= cor(y_t, y_t) = 1 \\
\rho_1 &= cor(y_t, y_{t+1}) = \frac{cov(y_t, y_{t+1})}{\sqrt{var(y_t)var(y_{t+1})}} \\
\rho_2 &= cor(y_t, y_{t+2}) = \frac{cov(y_t, y_{t+2})}{\sqrt{var(y_t)var(y_{t+2})}} \\
&\vdots \\
\rho_h &= cor(y_t, y_{t+h}) = \frac{cov(y_t, y_{t+h})}{\sqrt{var(y_t)var(y_{t+h})}}
\end{aligned}$$

- ρ_n **is the auto-correlation function(ACF)**-again a function of h.

- **Covariance Stationary and the ACF**
- Suppose y_t is covariance stationary.
- Then, by definition:

  $var(y_t)$ is the same for all t.

  $\gamma_h = cov(y_t, y_{t+h})$ is the same for all t.

  (But still different for all different h).

Covariance stationarity and the ACF

- This means that for the ACF.

$$\rho_h = cor(y_t, y_{t+h}) = \frac{cov(y_t, y_{t+h})}{\sqrt{var(y_t) var(y_{t+h})}}$$

is also the same for all t. (But again different for each h)

Covariance stationarity and the ACF

- This means that for the ACF.

$$\rho_h = cor(y_t, y_{t+h}) = \frac{cov(y_t, y_{t+h})}{\sqrt{var(y_t)var(y_{t+h})}}$$

  is also the same for all t. (But again different for each h)

- The ACF also simplifies since:

$$var(y_t) = var(y_{t+h}) = \gamma_0$$

$$\rho_h = \frac{\overbrace{cov(y_t, y_{t+h})}^{\gamma_h}}{\sqrt{\underbrace{var(y_t)}_{\gamma_0}\underbrace{var(y_{t+h})}_{\gamma_0}}} = \frac{\gamma_h}{\gamma_0} \tag{5}$$

Purpose of ACF

- **Purpose of the ACF**
- Useful in model selection.
- ACF of y_t gives some idea of which model may be appropriate for y_t—The model should capture the autocorrelation in data.
- ACF of the residual in your model can be used as a diagnostic check-If you selected a good model,the residuals should not be autocorrelated.
- More on this later

Calculating population(theoretical) ACF for ARMA models

- **Purpose:** Compare theoretical ACF of given model to estimated ACF to evaluate if model is appropriate.

[2] Since the constant does not impact the covariance, we can simplify the calculation of the theoretical ACF by omitting the intercept. We would not do this in practice.

Calculating population(theoretical) ACF for ARMA models

- **Purpose:** Compare theoretical ACF of given model to estimated ACF to evaluate if model is appropriate.
- **ACF for White Noise Process**[2]

$$
\begin{aligned}
y_t &= \varepsilon_t \\
E_{t-1}\varepsilon_t &= 0 \\
var(\varepsilon_t) &= \sigma^2
\end{aligned}
$$

(Above is a white noise model)

- Let's start with covariances.

$$
\begin{aligned}
\gamma_0 &= \sigma^2 \\
\gamma_1 &= \gamma_2 = \gamma_3 = \ldots = 0
\end{aligned}
$$

[2] Since the constant does not impact the covariance, we can simplify the calculation of the theoretical ACF by omitting the intercept. We would not do this in practice.

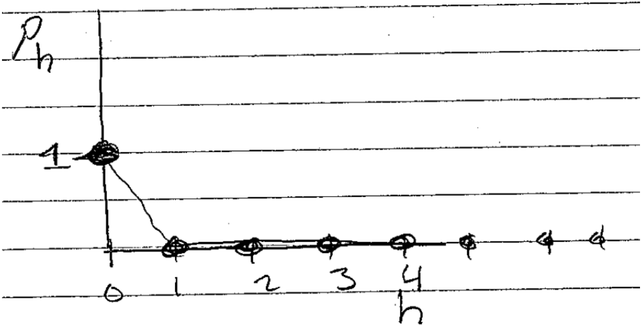
Calculating population(theoretical) ACF

- Now the ACF:

$$\begin{aligned}
\rho_0 &= \frac{\gamma_0}{\gamma_0} = 1 \text{ (always)} \\
\rho_1 &= \frac{\gamma_1}{\gamma_0} = \frac{0}{\gamma_0} = 0 \\
\rho_2 &= \rho_3 = \rho_4 = \ldots = 0
\end{aligned}$$

White Noise ACF

- Plot of White Noise ACF



White Noise ACF

- **Recognize white noise process by ACF:**

  ACF is zero (population case) or small and insignificant (sample case) for all h, except $h = 0$

  Intuition: white noise has no autocorrelation.

White Noise ACF

- **Recognize white noise process by ACF:**

  ACF is zero (population case) or small and insignificant (sample case) for all h, except $h = 0$

  Intuition: white noise has no autocorrelation.

- **Use as a residual diagnostic:**
  - Model residuals should be white noise if you picked a good model.
  - So their ACF should resemble the one above.

White Noise ACF

- **Recognize white noise process by ACF:**

  ACF is zero (population case) or small and insignificant (sample case) for all h, except $h = 0$

  Intuition: white noise has no autocorrelation.

- **Use as a residual diagnostic:**
  - Model residuals should be white noise if you picked a good model.
  - So their ACF should resemble the one above.

- **Application to finance**

  If returns not predictable, then their ACF should look like White Noise ACF.

ACF FOR MA(1) Model

- **ACF for MA(1) Model**

$$\begin{aligned} y_t &= \varepsilon_t + b_1 \varepsilon_{t-1} \\ E_{t-1}\varepsilon_t &= 0 \\ var(\varepsilon_t) &= \sigma^2 \end{aligned}$$

(Above is a MA(1) model)

$$\begin{aligned} \gamma_0 = var(y_t) &= var(\underbrace{\varepsilon_t + b_1 \varepsilon_{t-1}}_{\varepsilon_{t-1},\ \varepsilon_t \text{ orthogonal}}) \\ &= \underbrace{var(\varepsilon_t)}_{\sigma^2} + b_1^2 \underbrace{var(\varepsilon_{t-1})}_{\sigma^2} \\ &= (1 + b_1^2)\sigma^2 \qquad (6) \end{aligned}$$

ACF FOR MA(1) Model

$$\begin{aligned}
\gamma_1 &= cov(y_t, y_{t+1}) && (7)\\
&= cov[\overbrace{(\varepsilon_t + b_1\varepsilon_{t-1})}^{y_t}, \overbrace{(\varepsilon_{t+1} + b_1\varepsilon_t)}^{y_{t+1}}]\\
&\quad \text{(only } \varepsilon_t \text{ and } b_1\varepsilon_t \text{ non-orthogonal)}\\
&= b_1 E[\varepsilon_t^2]\\
&= b_1\sigma^2 && (8)
\end{aligned}$$

$$\begin{aligned}
\gamma_2 &= cov(y_t, y_{t+2})\\
&= cov\left(\varepsilon_t + b_1\varepsilon_{t-1}, \varepsilon_{t+2} + b_1\varepsilon_{t+1}\right)\\
&= 0
\end{aligned}$$

- Similarly $\gamma_3 = \gamma_4 = \gamma_5 = ... = 0$
- **Intuition:** MA(1) only overlaps 1 period.

- **Now, the ACF**

$$\rho_0 = 1 \qquad \text{always} \tag{9}$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} \tag{10}$$

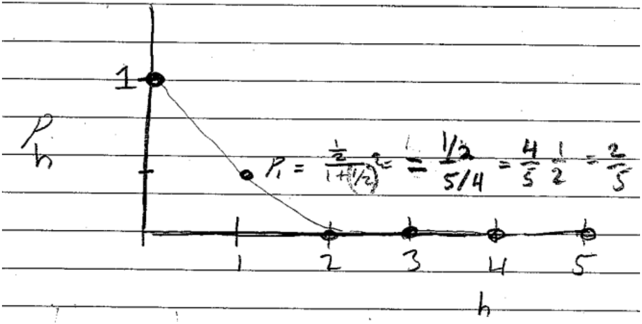
$$= \frac{b_1\sigma^2}{(1+b_1^2)\sigma^2} \tag{11}$$

$$= \frac{b_1}{1+b_1^2} \tag{12}$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = 0,\ \rho_3 = \frac{\gamma_3}{\gamma_0} = 0,\ \rho_4 = 0, \dots \tag{13}$$

ACF FOR MA(1) Model

- **Plot ACF for** $b_1 = \frac{1}{2}$

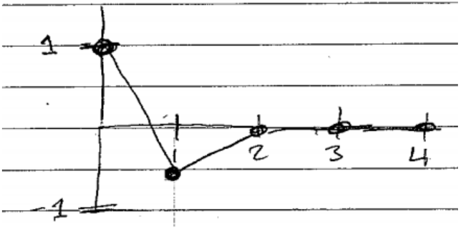


- Recognize MA(1) by its ACF:
  – ACF noticeable and significant at lag h=1
  – ACF small and insignificant for $h > 1$

ACF FOR MA(1) Model

- **Plot ACF for $b_1 = -1$ (an over-differenced series)**

$$\rho_1 = \frac{-1}{1+(-1)^2} = -\frac{1}{2}$$



- Recognize over-differenced seroes by its ACF:
  (i) ACF negative (close to $-\frac{1}{2}$) at lag 1
  (ii) Generally close to zero or insignificant at lag $h > 1$

ACF FOR MA(1) Model

Why "over differenced"?

- For $b_1 = -1$, MA(1) becomes $\varepsilon_t - \varepsilon_{t-1} = \Delta\varepsilon_t$ a difference.
- Sometimes differencing is employed when variables is non-stationary.
- But ε_t always stationary — it didn't need to be differenced.

ACF FOR AR(1) Model

$$y_t = a_1 y_{t-1} + \varepsilon_t \qquad AR(1)^3$$

- Solve for the covariance one time period apart:

$$\begin{aligned}
\gamma_1 &= cov(y_t, y_{t+1}) \qquad \text{(Plug in } y_{t+1} = a_1 y_t + \varepsilon_{t+1}) && (14)\\
&= cov(y_t, (a_1 y_t + \varepsilon_{t+1})) && (15)\\
&= a_1 cov(y_t, y_t) + cov(y_t, \varepsilon_{t+1}) && (16)\\
&= a_1 var(y_t) && (17)\\
&= a_1 \gamma_0 && (18)
\end{aligned}$$

[3] As usual also assume $E_{t-1}\varepsilon_t$, $var(\varepsilon_t)\sigma^2$ and $|a_1| < 1$

ACF FOR Stationary AR(1) Model

- Solve for the covariance two time periods apart:

$$\gamma_2 = cov(y_t, y_{t+2}) \qquad (\text{Plug in } y_{t+2} = a_1 y_{t+1} + \varepsilon_{t+2}) \tag{19}$$
$$= cov(y_t, a_1 y_{t+1} + \varepsilon_{t+2}) \tag{20}$$
$$= a_1 \underbrace{cov(y_t, y_{t+1})}_{\gamma_1} + \underbrace{cov(y_t, \varepsilon_{t+2})}_{0} \tag{21}$$
$$= a_1 \underbrace{cov(y_t, y_{t+1})}_{\gamma_1} \tag{22}$$
$$\tag{23}$$

- Now Recognize the pattern:

$$\gamma_1 = a_1 \gamma_0 \tag{24}$$
$$\gamma_2 = a_1 \gamma_1 = a_1^2 \gamma_0 \tag{25}$$
$$\vdots \qquad \vdots$$
$$\gamma_h = a_1^h \gamma_0 \qquad h = 0, 1, 2, 3, \ldots \tag{26}$$

ACF FOR AR(1) Model

- **Now the ACF**

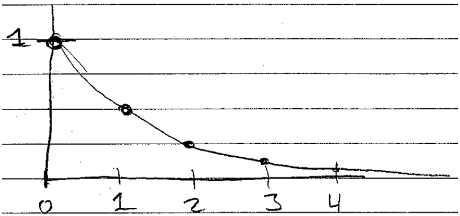
$$
\begin{aligned}
\rho_0 &= 1 && (27)\\
\rho_1 &= \frac{\gamma_1}{\gamma_0} = \frac{a_1\gamma_0}{\gamma_0} = a_1 && (28)\\
\rho_2 &= \frac{\gamma_2}{\gamma_0} = \frac{a_1^2\gamma_0}{\gamma_0} = a_1^2 && (29)\\
\vdots & \quad \vdots \\
\rho_h &= \frac{\gamma_h}{\gamma_0} = \frac{a_1^h\gamma_0}{\gamma_0} = a_1^h && (30)
\end{aligned}
$$

ACF FOR AR(1) Model

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

- **Plot ACF of AR(1) with** $a_1 = 1/2$

$$\rho_0 = 1,\ \rho_1 = 1/2,\ \rho_2 = (\frac{1}{2})^2 = \frac{1}{4},\ \rho_3 = (\frac{1}{2})^3 = \frac{1}{8},\ \rho_4 = \frac{1}{16}, \dots$$



ACF FOR AR(1) Model

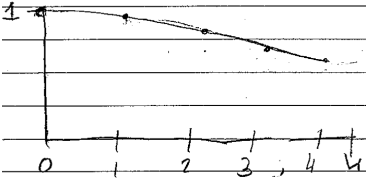
- **Recognize stationary AR(1) with $a_1 > 0$**
- Geometric decline of ACF
- ACF starts large/significant and positive.
- Becomes smaller and eventually insignificant

ACF for nearly nonstitonary AR(1)

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

- **Plot of nearly nonstationary AR(1):** $a_1 = 0.9$

$$\rho_0 = 1, \rho_1 = 0.9, \quad \rho_2 = (0.9)^2 = 0.81,$$
$$\rho_3 = (0.9)^3 = 0.729, \quad \rho_4 = (0.9)^4 = 0.6561$$



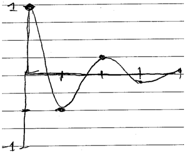
- **Recognize Possible (or almost) nonstationary AR(1)**
- ACF is slow to decline– stays large, positive, significant for many lags.

ACF for negative AR(1) slope coefficient

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

- **Plot ACF with $a_1 = -1/2$,(Negative Root)**

$$\rho_0 = 1, \rho_1 = -1/2, \rho_2 = (-\frac{1}{2})^2 = \frac{1}{4}, \rho_3 = (-\frac{1}{2})^3 = -\frac{1}{8}, \rho_4 = \frac{1}{16}$$



- **Recognize AR(1) with negative root:**
- Oscillates between positive and negative
- Magnitude decays geometrically

Estimated ACF

- So far— looked at ACF implied by various models
- Now— estimate ACF directly form data without a model
- Use— If the estimated ACF similar to one of the model based ACF_S this may suggest a tentative model
- Note— The estimated ACF is always noisy and never looks exactly like model based ACF - look for (broadly) similar features.

Estimated ACF

The estimated auto-covariance is:

$$\hat{\gamma}_h = \frac{1}{T} \sum_{t=h+1}^{T} (y_t - \bar{y})(y_{t-h} - \bar{y}) \qquad \text{(estimated covariance function)}$$

The estimated autocorrelation is:

$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0} = \frac{\frac{1}{T} \sum_{t=h+1}^{T} (y_t - \bar{y})(y_{t-h} - \bar{y})}{\frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^2} \qquad \text{(estimated ACF)}$$

Estimated ACF

- **Standard errors for ACF**
- It has been shown that:

$$var(\hat{\rho}_n) = \begin{cases} \frac{1}{T}, & h = 1 \\ \frac{1}{T}(1 + 2\sum_{j=1}^{h-1} \rho_j^2) & h > 1 \end{cases} \quad (31)$$

- So practical formula for standard error is:

$$se(\hat{\rho}_h) = \begin{cases} \frac{1}{\sqrt{T}} & h = 1 \\ \frac{1}{\sqrt{T}}\sqrt{1 + 2\sum_{j=1}^{s-1} \hat{\rho}_i^2} & h > 1 \end{cases} \quad (32)$$

- Often we want to compare our ACF to white noise ACF. Then we may calculate standard errors under null hypothesis of white noise ACF, for which $\rho_h = 0$ for all h.

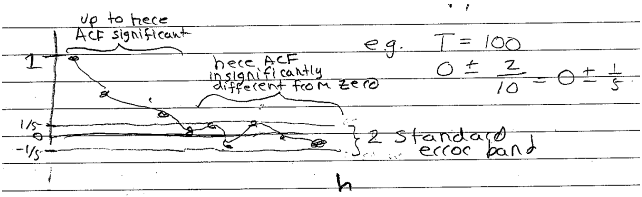
Estimated ACF

- In this case $se(\hat{\rho}_h)$ simplifies to:

$$se(\hat{\rho}_h) = \frac{1}{\sqrt{T}} \qquad h \geq 1$$

- Common to include two-standard-error bands away from zero in ACF using:

$$0 \pm 2se(\hat{\rho_n}) = 0 \pm \frac{2}{\sqrt{T}}$$



- **Box-Pierce and Ljung Box Statistics**
- Problem with 2 standard error bands is that each one can be wrong about 5% of time.
- So if we calculate ACF for $h = 1, 2, ..., 20$, there is a good chance that the ACF will lie outside the error bands for at least one or two values of h, even if the true model is White Noise.
- Box Pierce and Ljung Box statistics provide a joint test that the ACF is zero for the first H lags (e.g:H=20)

$$H_0 : \quad \rho_1 = \rho_2 = \rho_3 = ... = \rho_H = 0 \tag{33}$$

$$H_A : \quad \text{Not } H_0 \tag{34}$$

Box-Pierce and Ljung-Box Test Statistics

- Box-Pierce Test Statistic (Q):

$$Q = T \sum_{h=1}^{H} (\hat{\rho}_h)^2 \underset{H_0}{\sim} \chi^2(H) \qquad \text{(for T large)}$$

- Ljung-Box Test Statistic ($\tilde{Q}$):

$$\tilde{Q} = T(T+2) \sum_{h=1}^{H} \frac{\hat{\rho}_h^2}{T-h} \underset{H_0}{\sim} \chi^2(H) \qquad \text{(for T large)}$$

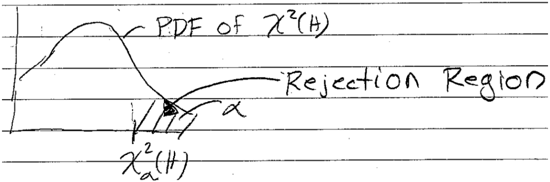
1. The distributions are asymptotic distribution that works as an approximation when T is large.
2. $\chi^2(H)$: chi-squared distribution with degree of freedom H.
3. $\underset{H_0}{\sim}$: distribution as under the null hypothesis(i.e: this is the distribution if null hypothesis holds true)

Box-Pierce and Ljung Box Statistics

- Rejection rules:

$$\text{Reject} H_0 \qquad \text{if } Q > \chi^2_\alpha(H) \text{ using Box Pierce} \tag{35}$$

$$\text{Reject} H_0 \qquad \text{if } \tilde{Q} > \chi^2_\alpha(H) \text{ using Ljung-Box} \tag{36}$$



- In practice: Ljung-Box is preferred due to better finite sample performance.

Use of Ljung-Box Statistics Test as a Diagnostic Test

- If your model correctly specified the residuals are White Noise.
- White Noise residuals implies ACF of residuals zero at all $lags \geq 1$
- Use Ljung-Box to test if ACF of residual zero at first H lags.
- If Ljung-Box test rejects then residual not white noise and model misspecified.
- Degrees of freedom adjusted for number of parameters estimated (say K) — Now $H - k$ degrees of freedom.

Ljung-Box as Diagnostic Text: Example

- Estimate tentative model is AR(1):

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

- Obtain the fitted residual:

$$\hat{\varepsilon}_t = y_t - \hat{a}_0 - \hat{a}_1 y_{t-1}$$

- Define the population and estimated ACF of ε_t :

$$\begin{aligned} \rho_{\varepsilon,h} &= cor\left(\varepsilon_t, \varepsilon_{t+h}\right) \\ \hat{\rho}_{\varepsilon,h} &= \widehat{cor}\left(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+h}\right) \end{aligned}$$

Ljung-Box as Diagnostic Text: Example

- Decide the number of lags to test: $H = 5$
- Specify the Null Hypothesis:

$$H_0 : \rho_{\varepsilon,h} = 0 \text{ for } h = 1, 2, 3, 4, 5$$

- **Interpret the null hypothesis:** Under the null hypothesis AR(1) is model correct, ε_t is white noise and therefore first five autocorrelations are zero
- Form the Ljung-Box test statistics:

$$\tilde{Q} \; = \; T(T+2) \sum_{h=1}^{5} \frac{\hat{\rho}^2_{\hat{\varepsilon},h}}{T-h}$$

Ljung-Box as Diagnostic Text: Example

- Decide the significance level: $\alpha = 0.05$
- Under the null hypothesis our test statistic has a Chi-Squared (χ^2) distribution with $H - k = 5 - 2 = 3$ degrees of freedom:[4]

$$\tilde{Q} \underset{H_0}{\sim} \chi^2(3)$$

- So, from the statistical tables, our critical value is given by

$$\chi^2_{0.05}(3) = 7.81473$$

- Reject if the test statistic exceeds the critical value, i.e. if:

$$\tilde{Q} > 7.81473$$

Or, equivalently if

$$\text{p-value} < \alpha = 0.05$$

[4] H=5 lags,k=2 estimated 2 parameters, $H - k = 3$ degrees of freedom

Ljung-Box as Diagnostic Text: Example

If we reject the null hypothesis:

1. Our residual is <u>not</u> white noise/uncorrelated
2. Our model does not capture all the correlation in y_t
3. We may need more AR terms, MA terms or a seasonal lag to get the remaining correlation
4. Maybe try an AR(2) or ARMA(1,1)

Ljung-Box as Diagnostic Text: Example

If we fail to reject the null hypothesis:

1. Our residual is <u>not</u> white noise/uncorrelated
2. Our model does capture all the correlation in y_t, which is good
3. Is it the smallest model that can do that?
4. If so, then we are happy with our model.
5. If not, maybe we can find an even smaller model that can capture the correlation in y_t

The Partial Autocorrelation Function (PACF)

- A second widely used visual diagnostic.
- Correlation between y_t and y_{t-h}, after controlling for the first h-1 lags $(y_{t-1}, y_{t-2}, ..., y_{t-h-1})$
- PACF at $lag1 = cor(y_t, y_{t-1}) = \rho_1$, because there are no lags in between to control for.
- PACF at lag2 is the correlation between y_t and y_{t-2} after controlling for y_{t-1}

PACF-Partial Autocorrelation Function

- Easiest to estimate (and understand) by a sequence of regressions.
  Step 1: Regress

$$y_t = \phi_{1,0} + \phi_{1,1} y_{t-1} + error$$

($\phi_{1,1}$ is PACF at lag 1.)

PACF-Partial Autocorrelation Function

- Easiest to estimate (and understand) by a sequence of regressions.
  Step 1: Regress
  $$y_t = \phi_{1,0} + \phi_{1,1} y_{t-1} + error$$
  ($\phi_{1,1}$ is PACF at lag 1.)
  Step 2: Regress
  $$y_t = \phi_{2,0} + \phi_{2,1} y_{t-1} + \phi_{2,2} y_{t-2} + error$$
  ($\phi_{2,2}$ is PACF at lag 2.)

PACF-Partial Autocorrelation Function

- Easiest to estimate (and understand) by a sequence of regressions.
  Step 1: Regress

$$y_t = \phi_{1,0} + \phi_{1,1} y_{t-1} + error$$

($\phi_{1,1}$ is PACF at lag 1.)

Step 2: Regress

$$y_t = \phi_{2,0} + \phi_{2,1} y_{t-1} + \phi_{2,2} y_{t-2} + error$$

($\phi_{2,2}$ is PACF at lag 2.)

Step 3: Regress

$$y_t = \phi_{3,0} + \phi_{3,1} y_{t-1} + \phi_{3,2} y_{t-2} + \phi_{3,3} y_{t-3} + error$$

($\phi_{3,3}$ is PACF at lag 3)

ect

PACF-Partial Autocorrelation Function

- **PACF of white noise**
- Population PACF of white noise is zero for all $h \geq 1$
- Sample PACF of white noise generally small and insignificant.

PACF-Partial Autocorrelation Function

- **PACF of AR(1)**

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t \qquad \text{AR(1) true model}$$

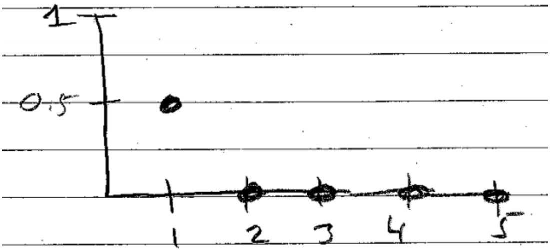
- Note that the regressions in step 1,2,3.. match the true AR(1) model:

$$\phi_{1,1} = a_1 \qquad \text{PACF at lag 1}$$

$$\phi_{2,2} = \phi_{3,3} = \phi_{4,4} = ... = \phi_{h,h} = 0 \qquad \text{PACF at lags h}$$

PACF-Partial Autocorrelation Function

- **Plot PACF for AR(1) with** $a_1 = 0.5$



- **Recognize AR(1) by its PACF**
- PACF large and significant at lag 1
- PACF small and insignificant thereafter

PACF for AR(2) model

- **Recognize AR(2) by its PACF**

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

- PACF large and significant at first **two** lags.
- PACF small and insignificant thereafter.

PACF for MA(1) model

$$y_t = \varepsilon_t + b_1 \varepsilon_{t-1} \qquad \textbf{MA(1)}$$

- In this form PACF not obvious
- Invert MA(1) model to get $AR(\infty)$:[5]

$$y_t = b_1 y_{t-1} - b_1^2 y_{t-2} + b_1^3 y_{t-1} + \ldots + \varepsilon_t \qquad AR(\infty)$$

item PACF of MA(1) declines **gradually**

- PACF of MA(1) oscillates in sign when $b_1 > 0$

[5] See next slide for derivation

Derivation: Converting MA(1) to AR(∞)

$$
\begin{aligned}
y_t &= \varepsilon_t + b_1\varepsilon_{t-1} \\
y_t &= (1 + b_1 L)\varepsilon_t \qquad \text{Convert to lag notation} \\
\frac{1}{1-(-b_1 L)} y_t &= \varepsilon_t
\end{aligned}
$$

Recall

$$
\sum_{j=0}^{\infty} x^j = \frac{1}{1-x} \Rightarrow \sum_{j=0}^{\infty} (-b_1 L)^j = \frac{1}{1-(-b_1 L)} = \frac{1}{1+b_1 L}
$$

$$
\begin{aligned}
\sum_{j=0}^{\infty} (-b_1 L)^j y_t &= \varepsilon_t \\
(1 - b_1 L + b_1 L^2 - b_1^3 L^3 + \ldots) y_t &= \varepsilon_t \\
y_t - b_1 y_{t-1} + b_1^2 y_{t-2} - b_1^3 y_{t-3} &= \varepsilon_t
\end{aligned}
$$

$$
y_t = b_1 y_{t-1} - b_1^2 y_{t-2} + b_1^3 y_{t-1} + \ldots + \varepsilon_t \qquad AR(\infty)
$$

Elimination of Common Factors

- **Intuition and purpose**
- generally there are many equivalent ARMA representations of the same time series process
- Examples already covered are the $MA(\infty)$ representation of an AR(1) and the $AR(\infty)$ representation of an MA(1).
- Other examples involve ARMA models with common factors (example below)
- Because estimation of additional parameters is costly, we want to select the representation that is most parsimonious (i.e. fewest parameters)
- In order to do this, we want to be sure to eliminate any common factors.

Common Factors-Example 1

- **Common Factors- Example 1** Start with ARMA(0,0) or white noise $\varepsilon_t \sim WN$, $y_t = \varepsilon_t$ This is an ARMA(0,0)
- Introduce a common factor by multiplying both sides by $(1 - a_1 L)$:

$$\begin{aligned} y_t &= \varepsilon_t && (37) \\ (1 - a_1 L) y_t &= (1 - a_1 L)\varepsilon_t \\ y_t - a_1 L y_t &= \varepsilon_t - a_1 L \varepsilon_t \\ y_t - a_1 y_{t-1} &= \varepsilon_t - a_1 \varepsilon_{t-1} \\ y_t &= a_1 y_{t-1} + \varepsilon_t - a_1 \varepsilon_{t-1} && (38) \end{aligned}$$

(38) is ARMA(1,1), representation of ARMA(0,0)

Common Factors-Example

- Another example:

$$
\begin{aligned}
y_t &= a_1 y_{t-1} + \varepsilon_t \text{ (Parsimonious AR(1))} \\
(1 - a_1 L) y_t &= \varepsilon_t \text{ (Rewrite using lag notation)} \\
(1 + \beta_1 L)(1 - a_1 L) y_t &= (1 + \beta_1 L)\varepsilon_t \\
\text{(Introduce a common factor} &\Rightarrow (1 + \beta_1 L)) \\
(1 - (a_1 - \beta_1)L - a_1\beta_1 L^2) y_t &= (1 + \beta_1 L)\varepsilon_t \\
y_t - (a_1 - \beta_1) y_{t-1} - a_1\beta_1 y_{t-2} &= \varepsilon_t + \beta_1 \varepsilon_{t-1}
\end{aligned}
$$

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

$$y_t = (a_1 - \beta_1) y_{t-1} + a_1\beta_1 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1} \qquad \text{ARMA(2,1)}$$

- **e.g** $a_1 = 0$, y_t is WN,using $\beta_1 = -1$ rewrite as $y_t = y_{t-1} + \varepsilon_t - \varepsilon_{t-1}$

Common Factors-Example

- **Interpretation**
- By adding common factor$(1 + \beta_1 L)$, we have re-expressed AR(1) as ARMA(2,1)
- These are 2 equivalent representations of the same model.
- But the AR(1) representation is preferred because it is more parsimonious.
- **Practical Use**
- We would never convert our AR(1) to an ARMA(2,1) as above- this was to illustrate the problem.
- Suppose we estimated an ARMA(2,1)
- We check to see if there is a common factor to both the lag polynomials for the MA and AR components that approximately cancel

Common Factors-Example

- If we can find a common factor, we can simply our model by eliminating it.
- In the example above, if we estimated as an ARMA(2,1) and noticed that $(1 + \beta_1 L)$ was a common factor that could be eliminated from both sides of

$$(1 + \beta_1 L)(1 - a_1 L)y_t = (1 + \beta_1 L)\varepsilon_t$$

We would realize that our ARMA(2,1) had an AR(1) representation. Then we would drop the ARMA(2,1) and estimate a more parsimonious AR(1) instead.

- In practice, it is unlikely that our estimated ARMA(2,1) would exactly have common factors, but if it had two factors that were vary similar, we might still want to consider a more parsimonious model, e.g

$$(1+0.42L)(1+a_1L)y_t = (1+0.47L)\varepsilon_t$$

(The two coefficients at the front don't cancel exactly, but are so close that we would probably drop this ARMA(2,1) in favor of $(1-aL)y_t = \varepsilon_t$ an AR(1))