Time Series Analysis

MODEL SPECIFICATION

Hossein Moradi Rekabdarkolaee

South Dakota State University

email: *hossein.moradirekabdarkolaee@sdstate.edu*.

Office: CAME Building, *Room: 240*.

Model Specification

- Up to this point, we have studied a large class of parametric models for both stationary and nonstationary time series - the ARIMA models.
- One major question is how to implement these models and make inferences about them.
- The subjects of the next three chapters, respectively, are:
  1. How to choose appropriate values for p, d, and q for a given series;
  2. How to estimate the parameters of a specific ARIMA(p, d, q) model;
  3. How to check on the appropriateness of the fitted model and improve it if needed.

Model-Building Strategy

- Finding appropriate models for time series is a nontrivial task.
- We will develop a multistep model-building strategy espoused so well by Box and Jenkins (1976).
- There are three main steps in the process, each of which may be used several times:
  1. model specification (or identification),
  2. model fitting, and
  3. model diagnostics.
- In choosing a model, the goal is to follow the principle of parsimony; that is, the model should require the smallest number of parameters that will adequately represent the time series.

Properties of the Sample Autocorrelation Function

- For the observed series $Y_1, Y_2, \ldots, Y_n$, the sample or estimated autocorrelation function is defined as

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}, \qquad k = 1, 2, \ldots$$

- Our goal is to find a patterns in r_k that mimics the characteristic of the known patterns in ρ_k for an ARMA models.

- Knowing the sampling properties of r_k would facilitate the comparison of estimated correlations with theoretical correlations.
- Suppose

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \qquad \epsilon_t \sim WN(0, \sigma^2)$$

- We assume $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $(\rightarrow)$ $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ (causality $\rightarrow$ stationary condition)

Properties of r_k

- For any fixed m, it can be shown that the joint distribution of

$$\begin{pmatrix} \sqrt{n}(r_1 - \rho_1) \\ \sqrt{n}(r_2 - \rho_2) \\ \vdots \\ \sqrt{n}(r_m - \rho_m) \end{pmatrix} \longrightarrow N_p(\mathbf{0}, \mathbf{C}) \quad \text{as} \quad n \to \infty$$

where $\mathbf{C} = (c_{ij})$ is the covariance matrix with dimension $m \times m$, and

$$c_{ij} = \sum_{k=\infty}^{\infty} \rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} + 2\rho_i\rho_j\rho_k^2$$

- For large n, r_k is approximately normally distributed:

$$r_k \sim N(\rho_k, \frac{c_{kk}}{n}), \quad \text{moreover} \quad Corr(r_k, r_j) \approx \frac{c_{kj}}{\sqrt{c_{kk}c_{jj}}} \quad \text{for} \quad k \neq j$$

Properties of r_k

- If Y_t is a white noise process, it can be shown that $Var(r_k) \approx \frac{1}{n}$, i.e.,

$$r_k \sim N(0, \frac{1}{n}), \quad \text{and} \quad Corr(r_k, r_j) \approx 0 \quad \text{for} \quad k \neq j$$

- Suppose Y_t is an AR(1) model with $\rho_k = \phi^k$ for $k > 0$.
- Then, it can be shown that

$$Var(r_k) \approx \frac{1}{n}\left[\frac{(1+\phi^2)(1-\phi^{2k})}{1-\phi^2} - 2k\phi^{2k}\right]$$

- In particular, for $k = 1$, and large k, we have

$$Var(r_1) \approx \frac{1-\phi^2}{n}, \qquad Var(r_k) \approx \frac{1}{n}\left[\frac{1+\phi^2}{1-\phi^2}\right] \quad \text{for large } k.$$

Properties of r_k - AR(1) model

- Suppose Y_t be an AR(1) model, then for general $0 < i < j$, we have

$$c_{ij} = \frac{(\phi^{j-i} - \phi^{j+i})(1+\phi^2)}{1-\phi^2} + (j-i)\phi^{j-i} - (j+i)\phi^{j+i},$$

- In particular,

$$Corr(r_1, r_2) \approx \frac{c_{12}}{\sqrt{c_{11}c_{22}}} = 2\phi\sqrt{\frac{1-\phi^2}{1+2\phi^2-3\phi^4}}.$$

Properties of r_k - MA(q) model

- For a general MA(q) process, it can be shown that

$$c_{kk} = 1 + 2\sum_{j=1}^{q} \rho_j^2, \qquad k > q,$$

therefore

$$Var(r_k) = \frac{1}{n}\left[1 + 2\sum_{j=1}^{q} \rho_j^2\right], \qquad k > q,$$

- For an observed time series, we can replace ρ's by r's, take the square root, and obtain an estimated standard deviation of r_k, that is, the standard error of r_k for large lags.
- A test of the hypothesis that the series is MA(q) could be carried out by comparing r_k to plus and minus two standard errors.
- We would reject the null hypothesis if and only if r_k lies outside these bounds.

Partial Autocorrelation Functions (PACF)

- For MA(q) models we have $\rho_k = 0$, for $k > q$.
- For both AR(p) and ARMA(p, q) models, however, $\rho_k = 0$ tails off to 0 as k increases.
- Example: For AR(1), $Y_t = \phi Y_{t-1} + \epsilon_t$, then $\rho_k = Corr(Y_t, Y_{t-k}) = \phi^k$ (it carries over to infinite lags).
- Can we distinguish AR from ARMA, as we can do with MA from ARMA based on ACF?
- Partial ACF (PACF) is another dependence measure that can expose differences between pure AR and ARMA models.

Partial ACF

- For MA(q) models we have $\rho_k = 0$, for $k > q$, the sample autocorrelation is a good indicator to determine the order of the process.
- However, the autocorrelations of an AR(p) model do not become zero after a certain number of lags-they die off rather than cut off.
- So a different function is needed to help determine the order of autoregressive models.
- Such a function may be defined as the correlation between Y_t and Y_{t-k} after removing the effect of the intervening variables $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-k+1}$.
- This coefficient is called the partial autocorrelation at lag k and will be denoted by ϕ_{kk}.

Partial ACF

1. If Y_t is a normally distributed time series, then

$$\phi_{kk} = Corr(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-k+1})$$

2. Consider predicting Y_t based on a linear function of the intervening variables $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-k+1}$, and predicting Y_{t-k} based on the same variables $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-k+1}$ (backward in time), then

$$\begin{aligned}\phi_{kk} = Corr(&Y_t - \beta_1 Y_{t-1} - \beta_2 Y_{t-2} - \ldots - \beta_{k-1} Y_{t-k+1}, \\ &Y_{t-k} - \beta_1 Y_{t-k+1} - \beta_2 Y_{t-k+2} - \ldots - \beta_{k-1} Y_{t-1})\end{aligned}$$

- If fact, the partial autocorrelation function at lag k is defined to be the correlation between the prediction errors.

Partial ACF

- In summary:
  - For lag k, define the best linear predictors of Y_t and Y_{t-k} based on $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1}$ as follows

$$\hat{Y}_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_{k-1} Y_{t-k+1}$$
$$\hat{Y}_{t-k} = \beta_1 Y_{t-k+1} + \beta_2 Y_{t-k+2} + \ldots + \beta_{k-1} Y_{t-1}$$

    with the β's chosen to minimize the mean square error (MSE) of prediction Predictors, i.e.,

$$\min_{\beta_1,\ldots,\beta_{k-1}} E(Y_t - \hat{Y}_t)^2 \quad \& \quad \min_{\beta_1,\ldots,\beta_{k-1}} E(Y_{t-k} - \hat{Y}_{t-k})^2$$

  - To remove effect of intermediate variables $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1}$, calculate the correlation of prediction errors $(Y_t - \hat{Y}_t)$ and $(Y_{t-k} - \hat{Y}_{t-k})$.

Partial ACF - Example

- It can be shown that the best linear prediction of Y_t based on just Y_{t-1} is $\hat{Y}_t = \rho_1 Y_{t-1}$. Therefore, to find ϕ_{22} we have

$$\begin{aligned} Cov(Y_t - \hat{Y}_t, Y_{t-k} - \hat{Y}_{t-k}) &= Cov(Y_t - \rho_1 Y_{t-1}, Y_{t-2} - \rho_1 Y_{t-1}) \\ &= \gamma_0(\rho_2 - \rho_1^2) \\ Var(Y_t - \rho_1 Y_{t-1}) &= \gamma_0(1 - \rho_1^2) \\ \Rightarrow \quad \phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \end{aligned}$$

- Now consider an AR(1) model. Recall that $\rho_k = \phi^k$, so that

$$\phi_{22} = \frac{\phi^2 - \phi^2}{1 - \phi^2} = 0 \quad \& \quad \phi_{kk} = 0 \qquad k > 1.$$

Partial ACF - AR(p)

- Consider a general AR(p) case.

$$\hat{Y}_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p}$$
$$\hat{Y}_{t-k} = h(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1})$$

  are the best linear predictor of Y_t and Y_{t-k}, respectively, based on $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1}$.

- So,

$$Cov(Y_t - \hat{Y}_t, Y_{t-k} - \hat{Y}_{t-k}) = Cov(e_t, h(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1})) = 0$$

- Thus for AR(p) model, we have

$$\phi_{kk} = 0 \quad \forall \ k > p.$$

Partial ACF

- A general method for finding the partial autocorrelation function for any stationary process is as follows (see Anderson 1971).
  - Suppose that Y_t is zero mean stationary process.
  - Consider a regression model where Y_{t+k} is regressed on k lagged variables

$$Y_{t+k} = \phi_{k1} Y_{t+k-1} + \phi_{k2} Y_{t+k-2} + \ldots + \phi_{kk} Y_t + \epsilon_{t+k}$$

  - ϕ_{ki} denotes the i^{th} regression parameter, ϵ_{t+k} is a normal error term uncorrelated with Y_{t+k-j} for $j \geq 1$
  - Multiplying Y_{t+k-j} on both sides of the above regression equation and taking the expectation, we get

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \ldots + \phi_{kk}\gamma_{j-k}$$

Partial ACF

- If we divide both sides by γ_0 we get

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \ldots + \phi_{kk}\rho_{j-k} \tag{1}$$

- For $j = 1, 2, \ldots, k$, we have the following system of equations:

$$\begin{cases} \rho_1 = & \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \ldots + \phi_{kk}\rho_{k-1} \\ \rho_2 = & \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \ldots + \phi_{kk}\rho_{k-2} \\ \vdots & \\ \rho_k = & \phi_{k1}\rho_{k-11} + \phi_{k2}\rho_{k-2} + \ldots + \phi_{kk}\rho_0 \end{cases}$$

- Using Cramer's rule successively for $k = 1, 2, \ldots$, we have

$$\phi_{11} = \rho_1$$

Partial ACF

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}, \qquad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}} \tag{2}$$

Partial ACF

- Note that if the process is an AR(p), then since for $k = p$ Equations (1) are just the Yule-Walker equations, therefore

$$\begin{aligned} \phi_{kj} &= \phi_j, \qquad j = 1, 2 \ldots, p \\ \phi_{pp} &= \phi_p \\ \phi_{kj} &= 0 \qquad j = p+1, \ldots, k \end{aligned}$$

- Levinson (1947) and Durbin (1960) showed that ϕ_{kk} (the solution of the determinant in Equation (2)) can be found as follows

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}} \tag{3}$$

where

$$\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j} \qquad \text{for} \quad j = 1, 2, \ldots, k-1$$

Partial ACF

- Therefore,

$$\phi_{11} = \rho_1,$$

$$\phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1 - \phi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2},$$

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11},$$

$$\phi_{33} = \frac{\rho_3 - \phi_{21}\rho_2 - \phi_{22}\rho_1}{1 - \phi_{21}\rho_1 - \phi_{22}\rho_2},$$

$$\vdots$$

The Sample PACF

- For an observed time series, we need to be able to estimate the partial autocorrelation function at a variety of lags.
- The recursive equations in (3) give us the theoretical partial ACF, but by replacing ρ's with r's, we obtain the estimated or sample partial ACF, i.e., $\hat{\phi}_{kk}$.
- Under the hypothesis that *an AR(p) model is correct*, then

$$\hat{\phi}_{kk} \sim N(0, \frac{1}{n}) \qquad \text{for} \quad k > p$$

- Thus, for $k > p$, $\pm 2/\sqrt{n}$ can be used as critical limits on $\hat{\phi}_{kk}$ to test the null hypothesis that an AR(p) model is correct.

ARMA model

- When the underlying random process that governs the behavior of the time series is an AR(p) or MA(q) the sample ACF and PACF can be enough to determine the model order.

- However, for a mixed ARMA model, if $p, q \geq 1$, it is not as easy, and it is difficult to identify p and q by the sample ACF and PACF.

Function	Model	Feature
ACF	MA(q)	Cutting-off at lag q
PACF	AR(p)	Cutting-off at lag p
EACF	ARMA(p, q)	A triangle with vertex (p, q)

Extended ACF

- The Extended ACF (EACF) is one (of several) tools to improve the choice of orders of ARMA(p, q) processes.
- For an ARMA(1,1) i.e.

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

here is the procedure:

  - Obtain the residuals of the simple regression of Y_t on Y_{t-1}.
  - Get the lag 1 of the residuals.
  - Run a multiple linear regression of Y_t on Y_{t-1} and $\nabla \epsilon_t$.
  - The coefficient of Y_{t-1} is a consistent estimate for the ϕ.

Extended ACF

- For ARMA(1,2), we need to run a third regression of Y_t on Y_{t-1}, the lag 1 of the residuals from the second regression, and the lag 2 of the residual of the first regression to obtain a consistent estimator of ϕ.
- This process can be extended for the ARMA(p,q).
- For AR of order k and MA order j, let

$$W_{t,k,j} = Y_t - \tilde{\phi}_1 Y_{t-1} - \tilde{\phi}_2 Y_{t-2} - \ldots - \tilde{\phi}_k Y_{t-k}$$

- to be the autoregressive residuals defined with the AR coefficients estimated iteratively.
- The sample ACF of $W_{t,k,j}$ called extended sample autocorrelation.

Extended ACF

- For $k = p$ and $j \geq q$, $\{W_{t,k,j}\}$ is approximately an MA(q) model.
- Therefore, its theoretical autocorrelations of lag $q + 1$ or higher are equal to zero.
- For $k > p$, an overfitting problem occurs, and this increases the MA order for the W process by the minimum of $k - p$ and $j - q$.
- Since the sample autocorrelation of $\{W_{t,k,j}\}$ process is asymptotically $N(0, 1/(n - k - j))$, we can test if it is 0 or not.

Non-stationary - AR(1)

- For nonstationary series, the sample ACF typically fails to die out rapidly as the lags increase.
- This is due to the tendency for nonstationary series to drift slowly, either up or down, with apparent “trends.”
- Let us look at the “Monthly Price of Oil: January 1986–January 2006” data.

Over-differencing

- We know that the difference of any stationary time series is also stationary.
- Over-differencing can introduce artificial dependencies.
- Consider

$$Y_t = Y_{t-1} + \epsilon_t$$

then

$$\nabla Y_t = \epsilon_t$$

but

$$\nabla^2 Y_t = \epsilon_t - \epsilon_{t-1} \sim \text{MA}(1)$$

- The variance of the over-differenced process will be larger than that of the original process.
- Over-differencing also creates a non-invertible model.

Unit Roots

- Consider the random walk model

$$Y_t = Y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim WN(0, \sigma^2)$$

- Because the autoregressive lag polynomial, has one root equal to one, we say it has a unit root.
- Because time series with unit roots are not covariance stationary, so unit roots require some special treatment.
- To identify the correct underlying time series model, we must test whether a unit root exists or not.

The Dickey-Fuller Unit-Root Test

- Consider the model

$$Y_t = \alpha Y_{t-1} + X_t, \qquad \text{where} \quad X_t \ \text{stationary process}$$

- The process Y_t is nonstationary if the coefficient $\alpha = 1$, but it is stationary if $|\alpha| < 1$.

$$H_0 : \alpha = 1$$
$$H_A : |\alpha| < 1$$

- Suppose X_t is an AR(k) process. Then under the null hypothesis we have, $X_t = Y_t - Y_{t-1}$

$$\begin{aligned} Y_t - Y_{t-1} &= (\alpha - 1) Y_{t-1} + X_t \\ &= aY_{t-1} + \phi_1 X_{t-1} + \ldots + \phi_k X_{t-k} + \epsilon_t \\ &= aY_{t-1} + \phi_1 (Y_{t-1} - Y_{t-2}) + \ldots + \phi_k (Y_{t-k} - Y_{t-k1}) + \epsilon_t \end{aligned}$$

The Dickey-Fuller Unit-Root Test

- Then under the null hypothesis ($\alpha = 1$), we have

$$\nabla Y_t = \phi_1 \nabla Y_{t-1} + \phi_k \nabla Y_{t-2} + \ldots + \phi_k \nabla Y_{t-k} + \epsilon_t$$

- Under the alternative hypothesis, ($|\alpha| < 1$) , Y_t is a stationary AR($k+1$) model with different coefficients, and the characteristic equation

$$\Phi(x)(1 - \alpha x) = 0, \qquad \Phi(x) = 1 - \phi_1 x - \ldots - \phi_k x^k$$

Statistical Issues with Unit Root Tests

- Unit root tests generally have nonstandard and non-normal asymptotic distributions.
- The distributions are affected by the inclusion of deterministic terms, e.g. constant, time trend, dummy variables, and so different sets of critical values must be used for test regressions with different deterministic terms.
- It is not always easy to tell if a unit root exists because these tests have low power against near-unit-root alternatives (e.g. $\phi = 0.95$)
- However, the truth is that the ADF test is a critical tool we use to identify the underlying time series model.
- That is, do we have: “ARMA”, or “trend + ARMA”, or “ ARIMA”?

Trend Case in Unit root test

- When testing for unit roots, it is crucial to specify the null and alternative hypotheses appropriately to characterize the trend properties of the data at hand.
  - If the observed data does not exhibit an increasing or decreasing trend, then the appropriate null and alternative hypotheses should reflect this.
  - The trend properties of the data under the alternative hypothesis will determine the form of the test regression used.
  - The type of deterministic terms in the test regression will influence the asymptotic distributions of the unit root test statistics.

Case I: Constant Only

- The test regression is

$$Y_t = a + \phi Y_{t-1} + \epsilon_t$$

and includes a constant to capture the nonzero mean under the alternative.

- The hypotheses to be tested are

$H_0 : \phi = 1, \quad a = 0$ nonstationary without drift

$H_A : |\phi| < 1$ stationary with zero mean

- This formulation is appropriate for non-trending economic and financial series like interest rates, exchange rates, and spreads.

Case II: Constant and Time Trend

- The test regression is

$$Y_t = a + bt + \phi Y_{t-1} + \epsilon_t$$

and includes a constant and deterministic time trend to capture the deterministic trend under the alternative

- The hypotheses to be tested are

$H_0 : \phi = 1, \quad b = 0$ nonstationary with drift

$H_A : |\phi| < 1$ stationary with deterministic time trend

- This formulation is appropriate for trending time series like asset prices or the levels of macroeconomic aggregates like real GDP.

Other Specification Methods

- We now consider the problem of model selection via information criteria.
- There are several information criteria available in the literature.
  - Akaike's (1973) Information Criterion (AIC).
  - corrected Akaike information criterion (AIC_c)
  - Schwarz Bayesian Information Criterion (Schwarz, 1978, Ann. Statist.) (BIC).
  - Hannan-Quinn Criteria (HQIC) (Hannan & Quinn, 1979).

AIC criterion

- The AIC criterion says to select the model that *minimizes*

$$AIC = -2\log(\text{maximum likelihood}) + 2k$$

where $k = p + q + 1$ if the model contains an intercept or constant term and $k = p + q$ otherwise.

- The AIC is an estimator of the average Kullback-Leibler divergence of the estimated model from the true model.
- Let $p(y_1, \ldots, y_n)$ be the true pdf of $Y_1, \ldots, Y_n$, and $q_\theta(y_1, \ldots, y_n)$ be the corresponding pdf under the model with parameter θ.
- The AIC estimates the following expectations, where $\hat{\theta}$ is the MLE

$$E\left[\log\left(\frac{p(y_1, \ldots, y_n)}{q_{\hat{\theta}}(y_1, \ldots, y_n)}\right)\right]$$

AIC_c

- AIC is a biased estimator, and the bias can be appreciable for large parameter per data ratios
- Hurvich and Tsai (1989) showed that the bias can be approximately eliminated by adding another nonstochastic penalty term to the AIC, resulting in the corrected AIC, denoted by AIC_c.

$$AIC_c = AIC + \frac{2(k+1)(k+2)}{n-k-2}$$

- k is the total number of parameters as above excluding the noise variance.

BIC

- Another approach to determining the ARMA orders is to select a model that *minimizes* the Schwarz Bayesian Information Criterion (BIC) defined as

$$BIC = -2\log(\text{maximum likelihood}) + k\log(n)$$

- In BIC, the penalty for additional parameters is stronger than that of the AIC.
- BIC has the most superior large sample properties.
- BIC is consistent, unbiased and sufficient.

AIC vs BIC

- Fit models with different (p, q) orders, and select one that has the minimum AIC / BIC
- In general, BIC tends to give simpler (i.e., smaller) models than AIC
- BIC puts higher penalty on the number of model parameters
- In practice, use AIC if model is meant for making predictions, and BIC if model is meant for describing TS mechanics