Lecture 6

ARMA Models: Inference, Diagnostics, and Model Selection

Reading: CC08: Chapter 7.2-7.5, Chapter 8.1, Chapter 6.5; BD16: Chapter 5.2, 5.3, 5.5; SS17: Chapter 3.5

MATH 8090 Time Series Analysis

Whitney Huang
Clemson University

Agenda

1. **Maximum Likelihood Estimation**
2. **Model Diagnostics and Selection**

AR(1) Log-likelihood

Let $\{\eta_1, \eta_2, \cdots, \eta_n\}$ be a realization of a zero-mean stationary AR(1) Gaussian time series. Let $\boldsymbol{\theta} = (\phi, \sigma^2)$

$$\ell_n(\boldsymbol{\theta}) = \underbrace{\log(f(\eta_1; \boldsymbol{\theta}))}_{\ell_{n,1}} + \underbrace{\sum_{t=2}^{n} \log f(\eta_t | \eta_{t-1}, \cdots, \eta_1; \boldsymbol{\theta})}_{\ell_{n,2}}.$$

AR(1) Log-likelihood

Let $\{\eta_1, \eta_2, \cdots, \eta_n\}$ be a realization of a zero-mean stationary AR(1) Gaussian time series. Let $\boldsymbol{\theta} = (\phi, \sigma^2)$

$$\ell_n(\boldsymbol{\theta}) = \underbrace{\log(f(\eta_1; \boldsymbol{\theta}))}_{\ell_{n,1}} + \underbrace{\sum_{t=2}^{n} \log f(\eta_t | \eta_{t-1}, \cdots, \eta_1; \boldsymbol{\theta})}_{\ell_{n,2}}.$$

Note that for $t \geq 2$, $f(\eta_t|\eta_{t-1}, \cdots, \eta_1) = f(\eta_t|\eta_{t-1})$, where $[\eta_t|\eta_{t-1}] \sim \mathrm{N}(\phi\eta_{t-1}, \sigma^2) \Rightarrow \ell_{n,2} =$

$$-\frac{(n-1)}{2}\log 2\pi - \frac{(n-1)}{2}\log\sigma^2 - \frac{\sum_{t=2}^{n}(\eta_t - \phi\eta_{t-1})^2}{2\sigma^2}$$

AR(1) Log-likelihood

Let $\{\eta_1, \eta_2, \cdots, \eta_n\}$ be a realization of a zero-mean stationary AR(1) Gaussian time series. Let $\boldsymbol{\theta} = (\phi, \sigma^2)$

$$\ell_n(\boldsymbol{\theta}) = \underbrace{\log(f(\eta_1; \boldsymbol{\theta}))}_{\ell_{n,1}} + \underbrace{\sum_{t=2}^{n} \log f(\eta_t | \eta_{t-1}, \cdots, \eta_1; \boldsymbol{\theta})}_{\ell_{n,2}}.$$

Note that for $t \geq 2$, $f(\eta_t | \eta_{t-1}, \cdots, \eta_1) = f(\eta_t | \eta_{t-1})$, where $[\eta_t | \eta_{t-1}] \sim \mathrm{N}(\phi \eta_{t-1}, \sigma^2) \Rightarrow \ell_{n,2} =$

$$-\frac{(n-1)}{2} \log 2\pi - \frac{(n-1)}{2} \log \sigma^2 - \frac{\sum_{t=2}^{n} (\eta_t - \phi \eta_{t-1})^2}{2\sigma^2}$$

Also, we know $[\eta_1] \sim \mathrm{N}\left(0, \frac{\sigma^2}{(1-\phi^2)}\right) \Rightarrow \ell_{1,n} =$

$$\frac{-\log 2\pi}{2} - \frac{\log \sigma^2}{2} + \frac{\log(1-\phi^2)}{2} - \frac{(1-\phi^2)\eta_1^2}{2\sigma^2}$$

AR(1) Log-likelihood

Let $\{\eta_1, \eta_2, \cdots, \eta_n\}$ be a realization of a zero-mean stationary AR(1) Gaussian time series. Let $\boldsymbol{\theta} = (\phi, \sigma^2)$

$$\ell_n(\boldsymbol{\theta}) = \underbrace{\log(f(\eta_1; \boldsymbol{\theta}))}_{\ell_{n,1}} + \underbrace{\sum_{t=2}^{n} \log f(\eta_t | \eta_{t-1}, \cdots, \eta_1; \boldsymbol{\theta})}_{\ell_{n,2}}.$$

Note that for $t \geq 2$, $f(\eta_t | \eta_{t-1}, \cdots, \eta_1) = f(\eta_t | \eta_{t-1})$, where $[\eta_t | \eta_{t-1}] \sim \mathrm{N}(\phi \eta_{t-1}, \sigma^2) \Rightarrow \ell_{n,2} =$

$$-\frac{(n-1)}{2} \log 2\pi - \frac{(n-1)}{2} \log \sigma^2 - \frac{\sum_{t=2}^{n} (\eta_t - \phi \eta_{t-1})^2}{2\sigma^2}$$

Also, we know $[\eta_1] \sim \mathrm{N}\left(0, \frac{\sigma^2}{(1-\phi^2)}\right) \Rightarrow \ell_{1,n} =$

$$\frac{-\log 2\pi}{2} - \frac{\log \sigma^2}{2} + \frac{\log(1-\phi^2)}{2} - \frac{(1-\phi^2)\eta_1^2}{2\sigma^2}$$

$$\Rightarrow \ell_n(\boldsymbol{\theta}) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{\sum_{t=2}^{n} (\eta_t - \phi \eta_{t-1})^2}{2\sigma^2} + \frac{\log(1-\phi^2)}{2} - \frac{(1-\phi^2)\eta_1^2}{2\sigma^2}$$

AR(1) Log-likelihood Cont'd

$$\ell_n(\boldsymbol{\theta}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 + \frac{\log(1-\phi^2)}{2} - \frac{S(\phi)}{2\sigma^2},$$

where $S(\phi) = \sum_{t=2}^{n}(\eta_t - \phi\eta_{t-1})^2 + (1-\phi^2)\eta_1^2$

- For given value of ϕ, $\ell_n(\phi, \sigma^2)$ can be maximized analytically with respect to σ^2

$$\hat{\sigma}^2 = \frac{S(\hat{\phi})}{n}$$

AR(1) Log-likelihood Cont'd

$$\ell_n(\boldsymbol{\theta}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 + \frac{\log(1-\phi^2)}{2} - \frac{S(\phi)}{2\sigma^2},$$

where $S(\phi) = \sum_{t=2}^{n}(\eta_t - \phi\eta_{t-1})^2 + (1-\phi^2)\eta_1^2$

- For given value of ϕ, $\ell_n(\phi, \sigma^2)$ can be maximized analytically with respect to σ^2

$$\hat{\sigma}^2 = \frac{S(\hat{\phi})}{n}$$

- Estimation of ϕ can be simplified by maximizing the **conditional sum-of-squares** ($\sum_{t=2}^{n}(\eta_t - \phi\eta_{t-1})^2$)

AR(1) Log-likelihood Cont'd

$$\ell_n(\boldsymbol{\theta}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 + \frac{\log(1-\phi^2)}{2} - \frac{S(\phi)}{2\sigma^2},$$

where $S(\phi) = \sum_{t=2}^{n}(\eta_t - \phi\eta_{t-1})^2 + (1-\phi^2)\eta_1^2$

- For given value of ϕ, $\ell_n(\phi, \sigma^2)$ can be maximized analytically with respect to σ^2

$$\hat{\sigma}^2 = \frac{S(\hat{\phi})}{n}$$

- Estimation of ϕ can be simplified by maximizing the **conditional sum-of-squares** $(\sum_{t=2}^{n}(\eta_t - \phi\eta_{t-1})^2)$

AR(1) Log-likelihood Cont'd

$$\ell_n(\boldsymbol{\theta}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 + \frac{\log(1-\phi^2)}{2} - \frac{S(\phi)}{2\sigma^2},$$

where $S(\phi) = \sum_{t=2}^{n}(\eta_t - \phi\eta_{t-1})^2 + (1-\phi^2)\eta_1^2$

- For given value of ϕ, $\ell_n(\phi, \sigma^2)$ can be maximized analytically with respect to σ^2

$$\hat{\sigma}^2 = \frac{S(\hat{\phi})}{n}$$

- Estimation of ϕ can be simplified by maximizing the **conditional sum-of-squares** $(\sum_{t=2}^{n}(\eta_t - \phi\eta_{t-1})^2)$

In the next slides, we are going to contrast Conditional Least Squares and Exact MLE

Conditional Least Squares vs. Exact MLE: AR(1)

- Conditional Least Squares

  Condition on η_1. Estimation minimizes $S(\phi) = \sum_{t=2}^{n}(\eta_t - \phi\eta_{t-1})^2$. Differentiate $S(\phi)$ and set to zero:

$$\frac{dS}{d\phi} = -2\sum_{t=2}^{n}\eta_{t-1}(\eta_t - \phi\eta_{t-1}) = 0 \Rightarrow \hat{\phi}_{\mathrm{CSS}} = \frac{\sum_{t=2}^{n}\eta_{t-1}\eta_t}{\sum_{t=2}^{n}\eta_{t-1}^2}.$$

  $\hat{\sigma}^2_{\mathrm{CSS}} = \frac{S(\hat{\phi}_{\mathrm{CSS}})}{n-1}$.

- MLE

  Assuming $\eta_1 \sim \mathrm{N}(0, \frac{\sigma^2}{1-\phi^2})$, the stationary distribution. Estimation minimizes $S(\phi) = \sum_{t=2}^{n}(\eta_t - \phi\eta_{t-1})^2 + (1-\phi^2)\eta_1^2 \Rightarrow \hat{\phi}_{\mathrm{ML}}$ no closed-from in general.
  $\hat{\sigma}^2_{\mathrm{ML}} = \frac{S(\hat{\phi}_{\mathrm{ML}})}{n}$

`arima` in `R` with the Lake Huron Example

arima: ARIMA Modelling of Time Series

Description

Fit an ARIMA model to a univariate time series.

Usage

```
arima(x, order = c(0L, 0L, 0L),
      seasonal = list(order = c(0L, 0L, 0L), period = NA),
      xreg = NULL, include.mean = TRUE,
      transform.pars = TRUE,
      fixed = NULL, init = NULL,
      method = c("CSS-ML", "ML", "CSS"), n.cond,
      SSinit = c("Gardner1980", "Rossignol2011"),
      optim.method = "BFGS",
      optim.control = list(), kappa = 1e6)
```

`arima` in `R` with the Lake Huron Example

arima: ARIMA Modelling of Time Series

Description

Fit an ARIMA model to a univariate time series.

Usage

```
arima(x, order = c(0L, 0L, 0L),
      seasonal = list(order = c(0L, 0L, 0L), period = NA),
      xreg = NULL, include.mean = TRUE,
      transform.pars = TRUE,
      fixed = NULL, init = NULL,
      method = c("CSS-ML", "ML", "CSS"), n.cond,
      SSinit = c("Gardner1980", "Rossignol2011"),
      optim.method = "BFGS",
      optim.control = list(), kappa = 1e6)
```

```{r}
(MLE_est1 <- arima(lm$residuals, order = c(2, 0, 0), method = "ML"))
```

```
Call:
arima(x = lm$residuals, order = c(2, 0, 0), method = "ML")

Coefficients:
         ar1      ar2  intercept
      1.0047  -0.2919     0.0197
s.e.  0.0977   0.1004     0.2350

sigma^2 estimated as 0.4571:  log likelihood = -101.25,  aic = 210.5
```

Using Innovations for Likelihood Calculation in ARMA Models

Let the best linear one-step predictor of η_t be

$$\hat{\eta}_t = \begin{cases} 0, & t = 1; \\ P_{t-1}\eta_t, & t = 2, \cdots, n \end{cases}$$

- The one-step prediction errors or innovations are defined

$$U_t = \eta_t - \hat{\eta}_t, \quad t = 1, \cdots, n,$$

and the associated mean squared error is

$$\nu_{t-1} = \mathbb{E}\left[(\eta_t - \hat{\eta}_t)^2\right] = \mathbb{E}(U_t^2), \quad t = 1, \cdots, n.$$

- For a causal ARMA process we can write $\nu_{t-1} = \sigma^2 r_{t-1}$, where r_t only depends on the AR and MA parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, but not σ^2

Working with the Innovations

- Result I: $\{U_t\}$ is an independent set of RVs with

$$U_t \sim \mathrm{N}(0, \nu_{t-1}), t = 1, \cdots, n$$

  $\Rightarrow$ the one-step prediction errors are uncorrelated with one another, and each each a normal distribution

Working with the Innovations

- Result I: $\{U_t\}$ is an independent set of RVs with

  $$U_t \sim \mathrm{N}(0, \nu_{t-1}), t = 1, \cdots, n$$

  $\Rightarrow$ the one-step prediction errors are uncorrelated with one another, and each each a normal distribution

- Result II: The likelihoods are the same if we use a model based on realizations of $\{\eta_t\}$ or a model based on realizations of $\{U_t\}$

Working with the Innovations

- Result I: $\{U_t\}$ is an independent set of RVs with

$$U_t \sim \mathrm{N}(0, \nu_{t-1}), t = 1, \cdots, n$$

$\Rightarrow$ the one-step prediction errors are uncorrelated with one another, and each each a normal distribution

- Result II: The likelihoods are the same if we use a model based on realizations of $\{\eta_t\}$ or a model based on realizations of $\{U_t\}$

- Therefore

$$\ell_n(\boldsymbol{\omega}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\log(\nu_{t-1}) - \frac{1}{2}\sum_{t=1}^{n}\left(\frac{u_t^2}{\nu_{t-1}}\right).$$

For a causal ARMA process this becomes

$$\ell_n(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\sum_{t=1}^{n}\log(r_{t-1}) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}\left(\frac{u_t^2}{r_{t-1}}\right)$$

The MLEs of σ^2, ϕ, and θ

- Now take the derivative of ℓ_n with respect to σ^2, setting the derivative equal to zero and solving for $\sigma^2 \Rightarrow$

$$\hat{\sigma}^2 = \frac{S(\boldsymbol{\phi}, \boldsymbol{\theta})}{n},$$

where

$$S(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{t=1}^{n} \left(\frac{u_t^2}{r_{t-1}} \right).$$

- Substituting $\hat{\sigma}^2$ into ℓ_n, the MLE estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$, respectively, are those values which **maximize**

$$\tilde{\ell}_n(\boldsymbol{\phi}, \boldsymbol{\theta}, \hat{\sigma}^2) = -\frac{n}{2}\log(\hat{\sigma}^2) - \frac{1}{2}\sum_{t=1}^{n}\log(r_{t-1}) - \frac{1}{2\hat{\sigma}^2}\sum_{t=1}^{n}\left(\frac{u_t^2}{r_{t-1}}\right)$$

Inference for the ARMA Parameters

Motivating example: What is an approximate 95% CI for ϕ_1 in an AR(1) model?

- **Standard errors** can be obtained by computing the inverse of the **Hessian matrix**: $\mathrm{Var}(\hat{\boldsymbol{\omega}}) = H(\hat{\boldsymbol{\omega}})^{-1}$, where $H(\boldsymbol{\theta}) = \frac{\partial^2 \ell_n(\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T}$

Inference for the ARMA Parameters

Motivating example: What is an approximate 95% CI for ϕ_1 in an AR(1) model?

- **Standard errors** can be obtained by computing the inverse of the **Hessian matrix**: $\mathrm{Var}(\hat{\boldsymbol{\omega}}) = H(\hat{\boldsymbol{\omega}})^{-1}$, where $H(\boldsymbol{\theta}) = \frac{\partial^2 \ell_n(\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T}$
- Let $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)$ denote the ARMA parameters (excluding σ^2), and let $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$ be the ML estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$. Then for "large" n, $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ have approximately a joint normal distribution:

$$\begin{bmatrix} \hat{\boldsymbol{\phi}} \\ \hat{\boldsymbol{\theta}} \end{bmatrix} \dot{\sim} \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\phi} \\ \boldsymbol{\theta} \end{bmatrix}, \frac{V(\boldsymbol{\phi}, \boldsymbol{\theta})}{n} \right)$$

Inference for the ARMA Parameters

Motivating example: What is an approximate 95% CI for ϕ_1 in an AR(1) model?

- **Standard errors** can be obtained by computing the inverse of the **Hessian matrix**: $\mathrm{Var}(\hat{\boldsymbol{\omega}}) = H(\hat{\boldsymbol{\omega}})^{-1}$, where $H(\boldsymbol{\theta}) = \frac{\partial^2 \ell_n(\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T}$
- Let $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)$ denote the ARMA parameters (excluding σ^2), and let $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$ be the ML estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$. Then for "large" n, $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ have approximately a joint normal distribution:

$$\begin{bmatrix} \hat{\boldsymbol{\phi}} \\ \hat{\boldsymbol{\theta}} \end{bmatrix} \dot{\sim} \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\phi} \\ \boldsymbol{\theta} \end{bmatrix}, \frac{V(\boldsymbol{\phi}, \boldsymbol{\theta})}{n} \right)$$

- $V(\boldsymbol{\phi}, \boldsymbol{\theta})$ is a known $(p+q) \times (p+q)$ matrix depending on the ARMA parameters

$V(\phi, \theta)$ for AR Processes

- For an AR(p) process

$$V(\boldsymbol{\phi}) = \sigma^2 \Gamma^{-1},$$

where Γ is the $p \times p$ covariance matrix of the series $(\eta_1, \cdots, \eta_p)$

- AR(1) process:

$$V(\phi_1) = 1 - \phi_1^2$$

- AR(2) process:

$$V(\phi_1, \phi_2) = \begin{bmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{bmatrix}$$

Other Examples of $V(\phi, \theta)$

- MA(1) process:

$$V(\theta_1) = 1 - \theta_1^2$$

- MA(2) process:

$$V(\theta_1, \theta_2) = \begin{bmatrix} 1 - \theta_2^2 & \theta_1(1 - \theta_2) \\ \theta_1(1 - \theta_2) & 1 - \theta_2^2 \end{bmatrix}$$

- Casual and invertible ARMA(1,1) process

$$V(\phi, \theta) = \frac{1 + \phi\theta}{(\phi + \theta)^2} \begin{bmatrix} (1 - \phi^2)(1 + \phi\theta) & -(1 - \phi^2)(1 - \theta^2) \\ -(1 - \phi^2)(1 - \theta^2) & 1 - \theta_2^2 \end{bmatrix}$$

- More generally, for "small" n, the covariance matrix of $(\hat{\phi}, \hat{\boldsymbol{\theta}})$ can be approximated using the second derivatives of the log-likelihood function, known as the Hessian matrix

MLE for Trend and Temporal Correlation in One Step

```{r}
(MLE_est4 <- arima(LakeHuron, order = c(2, 0, 0), xreg = yr))
```

```
Call:
arima(x = LakeHuron, order = c(2, 0, 0), xreg = yr)

Coefficients:
         ar1      ar2  intercept     xreg
      1.0048  -0.2913   620.5115  -0.0216
s.e.  0.0976   0.1004    15.5771   0.0081

sigma^2 estimated as 0.4566:  log likelihood = -101.2,  aic = 210.4
```

Fitted model:

$$Y_t = 620.51 - 0.022\text{Year} + \eta_t,$$

where

$$\eta_t = 1.00\eta_{t-1} - 0.29\eta_{t-2} + Z_t, \quad Z_t \sim \text{N}(0, \sigma^2 = 0.46^2).$$

What About Non-Gaussian Processes?

It is more challenging to express the joint distribution of η_t for non-Gaussian processes. Instead, we often rely on the Gaussian likelihood as an approximate likelihood

- In practice:
  - Transform the data to make the series as close to Gaussian as possible (e.g., using a log, square-root, or Box-Cox transformation)
  - Then use the Gaussian likelihood to estimate parameters, assuming the transformed series follows a near-Gaussian structure
  - For many real-world applications, this approximation works well and simplifies estimation. However, residual diagnostics are needed to ensure the model fits the data adequately

Assessing Fit / Comparing Different Time Series Models

- We can use diagnostic plots for the "residuals" of the fitted time series, along with Box tests to assess whether an i.i.d. process is reasonable

```
> Box.test(resids, lag = 10, type = "Ljung-Box", fitdf = 2)

        Box-Ljung test

data:  resids
X-squared = 3.7882, df = 8, p-value =
0.8757
```

Assessing Fit / Comparing Different Time Series Models

- We can use diagnostic plots for the "residuals" of the fitted time series, along with Box tests to assess whether an i.i.d. process is reasonable

```
> Box.test(resids, lag = 10, type = "Ljung-Box", fitdf = 2)

        Box-Ljung test

data:  resids
X-squared = 3.7882, df = 8, p-value =
0.8757
```

- Use confidence intervals for the parameters. Intervals that contain zero may indicate that we can simplify the model

Assessing Fit / Comparing Different Time Series Models

- We can use diagnostic plots for the "residuals" of the fitted time series, along with Box tests to assess whether an i.i.d. process is reasonable

```
> Box.test(resids, lag = 10, type = "Ljung-Box", fitdf = 2)

        Box-Ljung test

data:  resids
X-squared = 3.7882, df = 8, p-value =
0.8757
```

- Use confidence intervals for the parameters. Intervals that contain zero may indicate that we can simplify the model

- We can also use model selection criteria, such as AIC, to compare between different models

Diagnostics via the Time Series Residuals

- Recall the innovations are given by

$$U_t = X_t - \hat{X}_t$$

Diagnostics via the Time Series Residuals

- Recall the innovations are given by

$$U_t = X_t - \hat{X}_t$$

- Under a Gaussian model, $\{U_t : t = 1, \cdots, T\}$ is an independent set of RVs with

$$U_t \sim \mathrm{N}(0, \nu_{t-1}) \overset{d}{=} \sigma \mathrm{N}(0, r_{t-1}).$$

Diagnostics via the Time Series Residuals

- Recall the innovations are given by

$$U_t = X_t - \hat{X}_t$$

- Under a Gaussian model, $\{U_t : t = 1, \cdots, T\}$ is an independent set of RVs with

$$U_t \sim \mathrm{N}(0, \nu_{t-1}) \overset{d}{=} \sigma \mathrm{N}(0, r_{t-1}).$$

- Define the residuals $\{R_t\}$ by

$$R_t = \frac{U_t}{\sqrt{r_{t-1}}} = \frac{X_t - \hat{X}_t}{\sqrt{r_{t-1}}}$$

Under Gaussian model $R_t \overset{i.i.d}{\sim} \mathrm{N}(0, \sigma^2)$

ARMA Order Selection

- We would prefer to use models that compromise between a small residual error $\hat{\sigma}^2$ and a small number of parameters $(p + q + 1)$

ARMA Order Selection

- We would prefer to use models that compromise between a small residual error $\hat{\sigma}^2$ and a small number of parameters $(p + q + 1)$
- To choose the order (p and q) of ARMA model it makes sense to penalize models with a large number of parameters

ARMA Order Selection

- We would prefer to use models that compromise between a small residual error $\hat{\sigma}^2$ and a small number of parameters $(p + q + 1)$

- To choose the order (p and q) of ARMA model it makes sense to penalize models with a large number of parameters

- Here we consider an information based criteria to compare models

Akaike Information Criterion (AIC)

- The Akaike information criterion (AIC) is defined by

$$\mathrm{AIC} = -2\ell_n(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2) + 2(p+q+1)$$

- We choose the values of p and q that minimizes the AIC value

- For AR(p) models, AIC tends to overestimate p. The bias corrected version is

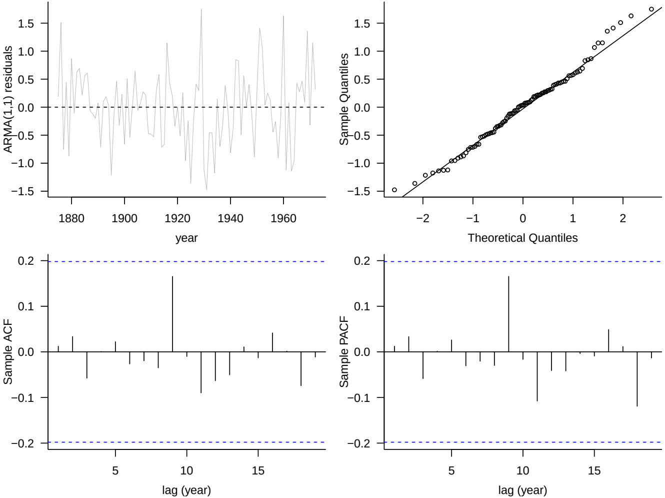
$$\mathrm{AICc} = 2\ell_n(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2) + \frac{2n(p+q+1)}{(n-1)-(p+q+1)}$$

Lake Huron Example: AIC and AICc

```
m1 <- arima(LakeHuron, order = c(1, 0, 0), xreg = yr)
m2 <- arima(LakeHuron, order = c(1, 0, 1), xreg = yr)
m3 <- arima(LakeHuron, order = c(2, 0, 0), xreg = yr)
m4 <- arima(LakeHuron, order = c(2, 0, 1), xreg = yr)
AIC(m1); AIC(m2); AIC(m3); AIC(m4)
library(MuMIn)
AICc(m1); AICc(m2); AICc(m3); AICc(m4)
```
```

```
[1] 218.4501
[1] 212.3954
[1] 212.3965
[1] 214.0638
[1] 218.8803
[1] 213.0476
[1] 213.0487
[1] 214.9868
```
```

Lake Huron Model Diagnostics




```
> Box.test(resids, lag = 10, type = "Ljung-Box", fitdf = 2)

        Box-Ljung test

data:  resids
X-squared = 3.7882, df = 8, p-value =
0.8757
```