Time Series Analysis

ARMA Models

Nicoleta Serban, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering

Order Selection & Residual Analysis

Georgia Tech

About This Lesson

ARMA Model

$\{X_t, t \in \mathbb{Z}\}$ an ARM$A(p, q)$ process,

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$$

where $Z_t \sim$ WN$(0, \sigma^2)$.

- Model Parameters: ϕ_1, .., ϕ_p (AR coefficients), $\theta_1, .., \theta_q$ (MA coefficients), μ (mean) and σ^2 (variance) unknown
- Commonly, de-mean the process X_t (hence X_t=0)
- Orders: p, q set fixed but they need to be selected

Order Selection: AICC

The AICC is a modified version of the Akaike Information Criterion (AIC):

- Let $\bar{\phi}=(\phi_1,\ldots,\phi_p)^T$ and $\bar{\theta}=(\theta_1,\ldots,\theta_q)^T$. The AICC is defined for an ARMA(p,q) model with coefficients $\bar{\phi}$, $\bar{\theta}$ and σ^2 by

$$\text{AICC}(\bar{\phi},\bar{\theta},\sigma^2) = -2\ln L(\bar{\phi},\bar{\theta},\sigma^2) + \boxed{\frac{2(p+q+1)n}{n-p-q-2}}.$$

- Approach to order selection:
    1. Fit models with varying orders 0,1,…p (AR) and 0,1,..,q (MA)
    2. Compute AICC for each combination of orders and select the orders with the smallest AICC

Residual Analysis

For a time series $\{X_t, t = 1, \ldots, n\}$ and an ARMA model for the time series, the residuals are defined to be

$$\hat{W}_t = \frac{\left(X_t - \hat{X}_t\right)}{r_{t-1}^{1/2}}, \qquad t = 1, 2, \ldots, n,$$

where $r_{t-1} = v_{t-1}/\sigma^2$ and $v_{t-1} = \mathrm{E}\left[\left(X_t - \hat{X}_t\right)^2\right]$.

The *rescaled residuals* are

$$\hat{R}_t = \hat{W}_t/\hat{\sigma}\,.$$

Residual Analysis (cont'd)

If $\{X_t\}$ really is a realization of the ARMA process used to generate the residuals, then properties of $\{\widehat{W}_t\}$ should reflect properties of the underlying process $\{Z_t\}$. In particular, $\{\widehat{W}_t\}$ should be approximately

1. uncorrelated if $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$, and
2. normal if $\{Z_t\} \sim \mathrm{N}(0, \sigma^2)$.

Georgia Tech

Diagnostics: Uncorrelated Residuals

The following tests can be used to test the hypothesis H_0: $\{W_t\} \sim$ WN$(0, \sigma^2)$ (uncorrelated data)

1. **The Sample ACF and PACF**

   If H_0 holds, then (at least, for large n), the sample ACF has the property: $\hat{\rho}_W(h) \sim$ IIDN$(0, 1/n)$ for $h > 1$. Also, the sample PACF has the property $\hat{\alpha}_W(h) \sim$ IIDN$(0, 1/n)$ for $h \geq 1$.

2. **The Portmanteau Test**

   Let $Q = n\sum_{j=1}^{h} \hat{\rho}_W^2(j)$. If H_0 holds, then Q is approximately χ^2 with h degrees of freedom. Reject H_0 if $Q > \chi^2_{1-\alpha}(h)$.

Uncorrelated residuals – Do not reject null hypothesis

Large P-value!!

Diagnostics: Uncorrelated Residuals (cont'd)

3. **The Ljung-Box Test**

$$Q_{LB} = n(n+2)\sum_{j=1}^{h} \hat{\rho}_W^2(j)/(n-j)\,.$$

Reject H_0 if $Q_{LB} > \chi^2_{1-\alpha}(h)$.

4. **The McLeod-Li Test**

$$Q_{ML} = n(n+2)\sum_{j=1}^{h} \hat{\rho}_{WW}^2(j)/(n-j)\,.$$

where $\hat{\rho}_{WW}(j)$ is the sample ACF of the squared residuals $\{W_1^2, \ldots, W_n^2\}$ at lag j.
Reject H_0 if $Q_{ML} > \chi^2_{1-\alpha}(h)$.

Diagnostics: Uncorrelated Residuals (cont’d)

Q-Q Plots

(These test for normality.) Let $W_{(1)} < W_{(2)} < \ldots < W_{(n)}$ be the order-statistics of a random sample $W_1, \ldots, W_n$ from a $N(\mu, \sigma^2)$ distribution. If $X_{(1)}, \ldots, X_{(n)}$ are the order statistics from a $N(0,1)$ sample of size n, then

$$\mathrm{E}\left[W_{(j)}\right] = \mu + \sigma m_j, \qquad m_j = \mathrm{E}\left[X_{(j)}\right]$$

so a plot of the points $\left(m_1, W_{(1)}\right), \ldots, \left(m_n, W_{(n)}\right)$ should be approximately linear.

Histogram

Evaluate the shape of the distribution in terms of skewedness, tails, gaps and outliers

Diagnostics: Normality Assumption (cont'd)

Hypothesis Testing Example:

- $\mathbf{H_0}$: **Residuals are not significantly different than a normal population.**
- $\mathbf{H_a}$: **Residuals are significantly different than a normal population**

Shapiro Wilk (W):

Fairly powerful omnibus test. Not good with small samples.

Good power with symmetrical, short and long tails. Good with asymmetry.

Jarque-Bera(JB):

Good with symmetric and long-tailed distributions.

Less powerful with asymmetry, and poor power with bimodal data.

D'Agostino(D or Y):

Good with symmetric and very good with long-tailed distributions.

Less powerful with asymmetry.

Anderson-Darling (A):

Similar in power to Shapiro-Wilk but has less power with asymmetry.

Works well with discrete data.

Outliers in Time Series

A data point far from the majority of the time series data may be called an *outlier*, especially if it does not follow the general trend of the rest of the data.

- A data point is called an *influential point* if it influences the fit of the time series model.
- Excluding an outlier point may or may not influence the model fit significantly.

The upshot: Sometimes there are good reasons to exclude subsets of data (e.g., errors in data entry or experimental errors). Sometimes an outlier belongs in the data. Outliers should always be examined.

Variance Stabilizing Transformation

Transform the response variable from X to $\overset{*}{\boldsymbol{X}}$ via

$$\overset{*}{\boldsymbol{X}} = \boldsymbol{X}^{\lambda}$$

where the value of λ depends on how $\mathrm{Var}(X)$ changes as x changes.

$\sigma_y(x) \propto const$	$\lambda = 1$	(don't transform)
$\sigma_y(x) \propto \sqrt{\mu_x}$	$\lambda = 1/2$	$\overset{*}{\boldsymbol{X}} = \sqrt{\boldsymbol{X}}$
$\sigma_y(x) \propto \mu_x$	$\lambda = 0$	$\overset{*}{\boldsymbol{X}} = \ln(X)$
$\sigma_y(x) \propto \mu_x^2$	$\lambda = -1$	$\overset{*}{\boldsymbol{X}} = \frac{1}{\boldsymbol{X}}$

Summary