

Spring 2013 Statistics 153 (Time Series) : Lecture Sixteen

Aditya Guntuboyina

19 March 2013

1 Seasonal ARMA Models

The doubly infinite sequence $\{X_t\}$ is said to be a seasonal ARMA(P, Q) process with period s if it is stationary and if it satisfies the difference equation $\Phi(B^s)X_t = \Theta(B^s)Z_t$ where $\{Z_t\}$ is white noise and

$$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$$

and

$$\Theta(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}.$$

Note that these can also be viewed as ARMA(Ps, Qs) models. However note that these models have $P + Q + 1$ (the 1 is for σ^2) parameters while a general ARMA(Ps, Qs) model will have $Ps + Qs + 1$ parameters. So these are much sparser models.

Unique Stationary solution exists to $\Phi(B^s)X_t = \Theta(B^s)Z_t$ if and only if every root of $\Phi(z^s)$ has magnitude different from one. Causal stationary solution exists if and only if every root of $\Phi(z^s)$ has magnitude strictly larger than one. Invertible stationary solution exists if and only if every root of $\Theta(z^s)$ has magnitude strictly larger than one.

The ACF and PACF of these models are **non-zero** only at the seasonal lags $h = 0, s, 2s, 3s, \dots$. At these seasonal lags, the ACF and PACF of these models behave just as the case of the unseasonal ARMA model: $\Phi(B)X_t = \Theta(B)Z_t$.

2 Multiplicative Seasonal ARMA Models

In the co2 dataset, for the first and seasonal differenced data, we needed to fit a stationary model with non-zero autocorrelations at lags 1, 11, 12 and 13 (and zero autocorrelation at all other lags). We can use a MA(13) model to this data but that will have 14 parameters and therefore will likely overfit the data. We can get a much more parsimonious model for this dataset by *combining* the MA(1) model with a seasonal MA(1) model of period 12. Specifically, consider the model

$$X_t = (1 + \Theta B^{12})(1 + \theta B)Z_t.$$

This model has the autocorrelation function:

$$\rho_x(1) = \frac{\theta}{1 + \theta^2} \quad \text{and} \quad \rho_x(12) = \frac{\Theta}{1 + \Theta^2}$$

and

$$\rho_x(11) = \rho_x(13) = \frac{\theta\Theta}{(1 + \theta^2)(1 + \Theta^2)}.$$

At every other lag, the autocorrelation $\rho_X(h)$ equals zero. This is therefore a suitable model for the first and seasonal differenced data in the co2 dataset.

More generally, we can combine, by multiplication, ARMA and seasonal ARMA models to obtain models which have special autocorrelation properties with respect to seasonal lags:

The **Multiplicative Seasonal Autoregressive Moving Average Model** $\text{ARMA}(p, q) \times (P, Q)_s$ is defined as the stationary solution to the difference equation:

$$\Phi(B^s)\phi(B)X_t = \Theta(B^s)\theta(B)Z_t.$$

The model we looked at above for the co2 dataset is $\text{ARMA}(0, 1) \times (0, 1)_{12}$.

Another example of a multiplicative seasonal ARMA model is the $\text{ARMA}(0, 1) \times (1, 0)_{12}$ model:

$$X_t - \Phi X_{t-12} = Z_t + \theta Z_{t-1}.$$

The autocorrelation function of this model can be checked to be $\rho_X(h) = \Phi^h$ for $h \geq 0$ and

$$\rho_X(12h - 1) = \rho_X(12h + 1) = \frac{\theta}{1 + \theta^2} \Phi^h \quad \text{for } h = 0, 1, 2, \dots$$

and $\rho_X(h) = 0$ at all other lags.

When you get a stationary dataset whose correlogram shows interesting patterns at seasonal lags, consider using a multiplicative seasonal ARMA model. You may use the R function *ARMAacf* to understand the autocorrelation and partial autocorrelation functions of these models.

3 SARIMA Models

These models are obtained by combining differencing with multiplicative seasonal ARMA models. These models are denoted by $\text{ARIMA}(p, d, q) \times (P, D, Q)_s$. This means that after differencing d times and seasonal differencing D times, we get a multiplicative seasonal ARMA model. In other words, $\{Y_t\}$ is $\text{ARIMA}(p, d, q) \times (P, D, Q)_s$ if it satisfies the difference equation:

$$\Phi(B^s)\phi(B)\nabla_s^D\nabla^d Y_t = \delta + \Theta(B^s)\theta(B)Z_t.$$

Recall that $\nabla_s^d = (1 - B^s)^d$ and $\nabla^d = (1 - B)^d$ denote the differencing operators.

In the co2 example, we wanted to use the model $\text{ARMA}(0, 1) \times (0, 1)_{12}$ to the seasonal and first differenced data: $\nabla\nabla_{12}X_t$. In other words, we want to fit the SARIMA model with nonseasonal orders 0, 1, 1 and seasonal orders 0, 1, 1 with seasonal period 12 to the co2 dataset. This model can be fit to the data using the function *arima()* with the *seasonal* argument.

4 AIC

AIC stands for Akaike's Information Criterion. It is a model selection criterion that recommends choosing a model for which:

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

is the smallest. Here k denotes the number of parameters in the model. For example, in the case of an $\text{ARMA}(p, q)$ model with a non-zero mean μ , we have $k = p + q + 2$.

The first term in the definition of AIC measures the fit of the model i.e., the model performance on the given data set. The term $2k$ serves as a penalty function which penalizes models with too many parameters.

While comparing a bunch of models for a given dataset, you may use the AIC. There are other criteria as well. For example, the BIC (Bayesian Information Criterion) looks at:

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

Note that the penalty above is larger than that of AIC. Consequently, BIC selects more parsimonious models compared to AIC.

5 Time Series Cross Validation

Read the two articles on Rob Hyndman's blog: <http://robjhyndman.com/hyndsight/crossvalidation/> for a simple introduction to cross validation in general and <http://robjhyndman.com/hyndsight/tscvexample/> for cross validation specific to time series.

There are many ways to do cross validation for time series. Suppose we have monthly data for m years x_1, \dots, x_n where $n = 12m$ and the objective is to predict the data for the next year (This is similar to the midterm problem which has weekly data instead of monthly). Suppose we have ℓ competing models M_1, \dots, M_ℓ for the dataset. We can use cross-validation in order to pick one of these models in the following way:

1. Fix a model M_i . Fix $k < m$.
2. Fit the model M_i to the data from the first k years.
3. Using the fitted model, predict the data for the $(k + 1)$ st year.
4. Calculate the sum of squares of errors of prediction for the $(k + 1)$ st year.
5. Repeat these steps for $k = k_0, \dots, m - 1$ where k_0 is an arbitrary value of your choice.
6. Average the sum of squares of errors of prediction over $k = k_0, \dots, m - 1$. Denote this value by CV_i and call it the Cross Validation score of model M_i .
7. Calculate CV_i for each $i = 1, \dots, \ell$ and choose the model with the smallest Cross-Validation score.