Spring 2015 Statistics 153 (Time Series) : Lecture Eighteen

Aditya Guntuboyina

02 March 2015

1 Model Diagnostics

After fitting an ARIMA model to data, one can form the residuals: $x_i - \hat{x}_i^{i-1}$ by looking at the difference between the ith observation and the best linear prediction of the ith observation based on the previous observations $x_1, \dots, x_{i-1}$. One usually standardizes this residual by dividing by the square-root of the corresponding prediction error.

If the model fits well, the standardized residuals should behave as an iid sequence with mean zero and variance one. One can check this by looking at the plot of the standardized residuals and their correlogram.

Let $r_e(h)$ denote the sample acf of the standardized residuals from an ARMA fit. For the fit to be good, the standardized residuals have to be iid with mean zero and variance one which implies that $r_e(h)$ for $h = 1, 2, \dots$ have to be i.i.d with mean 0 and variance $1/n$.

In addition to plotting $r_e(h)$, there is a formal test that takes into account the magnitudes of $r_e(h)$ together. This is the Ljung-Box-Pierce test that is based on the so-called Q-statistic:

$$Q := n(n+2) \sum_{h=1}^{H} \frac{r_e^2(h)}{n-h}.$$

Under the null hypothesis of model adequacy, the distribution of Q is asymptotically χ^2 with degrees of freedom $H - p - q$. The maximum lag H is chosen arbitrarily (typically 20). Thus, one would reject the null at level α if the observed value of Q exceeds the $(1-\alpha)$ quantile of the χ^2 distribution with $H-p-q$ degrees of freedom.

2 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

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

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: $\Phi(B)X_t = \Theta(B)Z_t$.

3 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** 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 ARMA$(0, 1) \times (0, 1)_{12}$.

Another example of a multiplicative seasonal ARMA model is the 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 $\rho_X(12h) = \Phi^h$ for $h \geq 0$ and

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

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.

4 SARIMA Models

These models are obtained by combining differencing with multiplicative seasonal ARMA models. These models are denoted by 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 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 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.

5 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 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.