Spring 2013 Statistics 153 (Time Series) : Lecture Seventeen

Aditya Guntuboyina

21 March 2013

1 Overfitting as a Diagnostic Tool

After fitting an adequate model to the data, fit a slightly more general model. For example, if an AR(2) model seems appropriate, overfit with an AR(3) model. The original AR(2) model can be confirmed if while fitting the AR(3) model:

1. The estimate of the additional ϕ_3 parameter is not significantly different from zero.
2. The estimates of the common parameters, ϕ_1 and ϕ_2, do not change significantly from their original estimates.

How does one choose this general model to overfit? While fitting a more general model, one should not increase the order of both the AR and MA models. Because it leads to lack of identifiability issues. For example: consider the MA(1) model: $X_t = (1+\theta B)Z_t$. Then by multiplying by the polynomial $1-\phi z$ on both sides: we see that X_t also satisfies the ARMA(1, 2) model: $X_t - \phi X_{t-1} = Z_t + (\theta - \phi)Z_{t-1} + \phi\theta Z_{t-2}$. But note that the parameter ϕ is not unique and thus if we fit an ARMA(1, 2) model to a dataset that is from MA(1), we might just get an arbitrary estimate for ϕ.

In general, it is a good idea to find the general overfitting model based on the analysis of the residuals. For example, if after fitting an MA(1) model, a not too small correlation remains at lag 2 in the residuals, then overfit with an MA(2) and not ARMA(1, 1) model.

2 Sines and Cosines

Frequency domain techniques: Using sines and cosines to study time series data.

Sinusoid: $R\cos(2\pi ft + \Phi)$. The following terminology is standard. R is called the *amplitude*, f is called the *frequency* and Φ is called the *phase*. The quantity $1/f$ is called the *period* and $2\pi f$ is termed the *angular frequency*. Note that three parameters R, f and Φ are involved in the definition of the sinusoid.

The function can also be written as $A\cos 2\pi ft + B\sin 2\pi ft$ where $A = R\cos\Phi$ and $B = R\sin\Phi$. This parametrization also has three parameters: f, A and B.