Spring 2013 Statistics 153 (Time Series) : Lecture Fifteen

Aditya Guntuboyina

14 March 2013

1 ARIMA Forecasting

Forecasting for a future observation, x_{n+m}, is done using the best linear predictor of X_{n+m} in terms of $X_1, \dots, X_n$. The coefficients of the best linear predictor involve the parameters of the ARIMA model used for $x_1, \dots, x_n$. These paramters are estimated from data.

We have already seen how the best linear predictor of a random variable Y in terms of $W_1, \dots, W_m$ is calculated.

Suppose that all the random variables $Y, W_1, \dots, W_m$ have mean zero. Then the best linear predictor is $a_1 W_1 + \dots + a_m W_m$ where $a_0, \dots, a_m$ are characterized by the set of equations:

$$\mathrm{cov}(Y - a_1 W_1 - \dots - a_m W_m, W_i) = 0 \qquad \text{for } i = 1, \dots, m.$$

The above gives m equations in the m unknowns $a_1, \dots, a_m$. The equations can be written in a compact form as $\Delta a = \zeta$ where $\Delta(i,j) = \mathrm{cov}(W_i, W_j)$ and $\zeta_i = \mathrm{cov}(Y, W_i)$.

If the random variables $Y, W_1, \dots, W_m$ have different means: $\mathbb{E}Y = \mu_Y$ and $\mathbb{E}W_i = \mu_i$, then the best linear predictor of $Y - \mu_Y$ in terms of $W_1 - \mu_1, \dots, W_m - \mu_m$ is given by $a_1(W_1 - \mu_1) + \dots + a_m(W_m - \mu_m)$ where $a_1, \dots, a_m$ are given by the same equation $\Delta a = \zeta$. Thus, in these non-zero mean case, the best linear predictor of Y in terms of $W_1, \dots, W_m$ is

$$\mu_Y + a_1(W_1 - \mu_1) + \dots + a_m(W_m - \mu_m).$$

The prediction error is measured by

$$\mathbb{E}\left(Y - \mu_Y - a_1(W_1 - \mu_1) - \dots - a_m(W_m - \mu_m)\right)^2.$$

For ARMA models, there exist iterative algorithms for quickly calculating the best linear predictors of X_{n+m} based on $X_1, \dots, X_n$ and the corresponding prediction errors recursively over n and m e.g., Durbin-Levinson and Innovations. These do not explicit inversion of the matrix Δ.

2 Time Series Data Analysis

1. Exploratory analysis.
2. Decide if it makes sense to transform the data (either for better interpretation or for stabilizing the variance).
3. Deal with trend or seasonality. Either by fitting deterministic models or by smoothing or differencing.

4. Fit an ARMA model to the residuals obtained after trend and seasonality are removed.
5. Check if the fitted ARMA model is adequate (Model Diagnostics).
6. Forecast.

3 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 residuals and their correlogram. Departures from gaussianity also need to be assessed (this is done by looking at the Q-Q plot).

Let $r_e(h)$ denote the sample acf of the residuals from an ARMA fit. For the fit to be good, the 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.

4 Seasonal ARIMA Models

These provide models that have non-zero autocorrelations for small lags (say, 0 and 1) and also at some seasonal lag (say, 12) and zero autocorrelation for all other lags.

Consider the MA model: $X_t = Z_t + \Theta Z_{t-12} = (1 + \Theta B^{12}) Z_t$. This can be thought of as an MA(12) model with $\theta_1 = \dots = \theta_{11} = 0$ and $\theta_{12} = \Theta$. This is a stationary model whose autocovariance function is non-zero only at lags 0 and 12. It is therefore called a seasonal MA(1) model with seasonal period 12.

Generalizing, a seasonal MA(Q) model with seasonal period s is defined by

$$X_t = Z_t + \Theta_1 Z_{t-s} + \Theta_2 Z_{t-2s} + \dots + \Theta_Q Z_{t-Qs}.$$

This stationary model has autocorrelation that is non-zero only at lags $0, s, 2s, \dots, Qs$. Note that this is just a MA(Qs) model with the MA polynomial $1 + \Theta_1 z^s + \Theta_2 z^{2s} + \dots + \Theta_Q z^{Qs}$.

For the co2 dataset, we need a stationary model with non-zero autocorrelations at lags 1, 11, 12 and 13 (and zero autocorrelation at all other lags). An example of such a model is given by:

$$X_t = Z_t + \theta Z_{t-1} + \Theta Z_{t-12} + \theta\Theta Z_{t-13}.$$

More compactly, this model can be written as

$$X_t = (1 + \theta B + \Theta B^{12} + \theta\Theta B^{13})Z_t = (1 + \theta B)(1 + \Theta B^{12})Z_t.$$

This is just a MA(12) model with the MA polynomial $(1 + \theta z)(1 + \Theta z^{12})$.

It is easy to check that for this model:

$$\gamma_X(h) = (1 + \theta^2)(1 + \Theta^2)\sigma_Z^2,$$

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

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.

More generally, we can consider ARMA models with AR polynomial $\phi(z)\Phi(z)$ and MA polynomial $\theta(z)\Theta(z)$ where

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p,$$
$$\Phi(z) = 1 - \Phi_1 z^s - \Phi_2 z^{2s} - \cdots - \Phi_p z^{Ps}$$

and

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q,$$
$$\Theta(z) = 1 + \Theta_1 z^s + \Theta_2 z^{2s} + \cdots + \Theta_Q z^{Qs}.$$

This is called the **multiplicative seasonal ARMA**$(p, q) \times (P, Q)_s$ **model with seasonal period** s.

In the co2 example, we wanted to use such a model to the first and seasonal differenced data. Specifically, we want to use the model ARMA$(0, 1) \times (0, 1)_{12}$ to the seasonal and first differenced data: $\nabla\nabla_{12}X_t$. A sequence $\{Y_t\}$ is said to be a **multiplicative seasonal ARIMA model** with nonseasonal orders p, d, q, seasonal orders P, D, Q and seasonal period s if the differenced series $\nabla^d\nabla_s^d Y_t$ satisfies an ARMA$(p, q) \times (P, Q)_s$ model with seasonal period s.

Therefore, we want to fit the multiplicative seasonal ARIMA 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 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.