Session 9: Time Series Intro pt 2

Devesh Tiwari

July 6, 2016

Background

For any given time series, we can think of it as having two parts: The longer term trend and the error, each of which we can model each separately. For the former, we can use smoothing/filtering (among other techniques) and after using a smoothing technique, we can extract the residuals in the same way we did when we were working with the CLM.

The resulting residuals are called the residual series. With time series data, we often worry about serial autocorrelation (aka dependency structure) and the bulk of this course deals with two types of temporal relationships: The auto regressive (AR) and the moving average (MA) processes. Residual series with no trend, a constant variance, and nodependency structure is called White Noise (these time series are called strictly stationary). White Noise is also the type of residuals that should be produced from a CLM where are all of the assumptions have been satisfied (normally dist, homoskedastic, zero-conditional mean).

High level overview of modeling time series

At a high level of abstraction, modeling time series data requires two major steps.

(1) Model the longer term trend using smoothing, regression, and other techniques.

(2) Model the resulting residual series, which ought to be stationary in the mean but not neccessarily white noise. Residuals are often correlated with each other.

Step two is what this course focusees on. We assume that a given observation in the residual series is a function of its past observations. We also assume that the residuals are generated by an AR and/or MA process. What we do not know is:

(1) How many of the lagged observations influence a given data point.

(2) The *true* underlying process generating the residual series. It could be an AR or MA process, a combination of the two (ARMA), whether differencing is needed (ARIMA), whether or not there is a seasonal component to the residual series (SARIMA), or whether the residual series has constant variance or not (GARCH).

Because we do not know these things, we have to examine several models and choose the best one. When modeling the residual series, we aim to:

(1) Generate a residual series that resembles white noise.

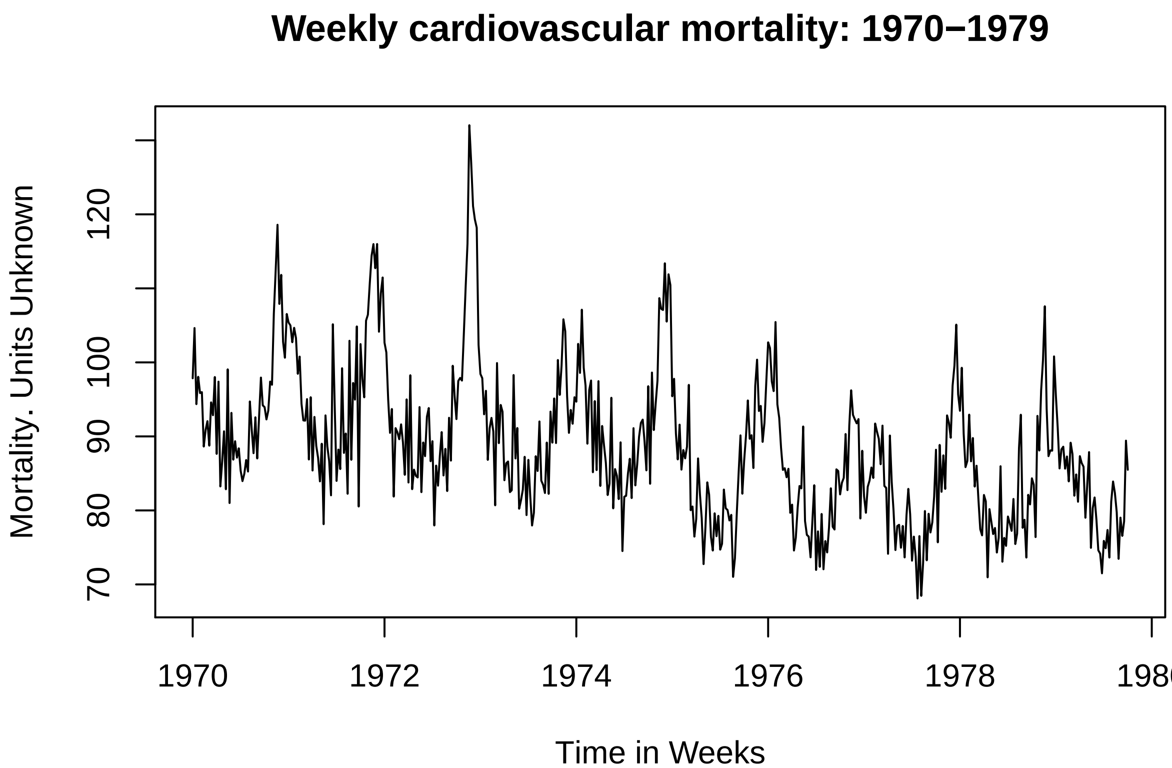
(2) Maximize predicive accuracy with the least complex model possible (parsimony principle).

Exericse 1: Smoothing, bandwidth size, and accuracy

Examine the cardiovascular mortaility data from last week and plot it.

```
rm(list = ls())
library(astsa)

plot(cmort, xlab= "Time in Weeks", ylab="Mortality. Units Unknown")
title(main="Weekly cardiovascular mortality: 1970-1979")
```



(1) Is this time series stationary?

(2) Apply the moving average smoothing function to the data and begin by choosing whatever window size you wish.

(3) Plot the moving average series. What do you see? What notable features can you see here that you could not see clearly in the raw data?

(4) Apply the moving average smoothing function to the raw data again, this time choosing a different window size.

(5) Plot the raw data and two moving averages on the same plot. Based on what you see, what is the relationship between the size of the bandwidth and how well if fits the data? Which moving average series is more "complicated"?

Exericise 2: Residual Time Series

After we have detected a trend, it is useful to de-trend the data. We do this by taking every value of *cmort* and subtract our trended data from it. This is very similar to how we calculated the residuals from OLS regressions. The resulting residuals is called the *residual error series*. Return to your moving averages from above.

(1) Choose one of the moving average objects you created in part 1 and create an object that contains your residuals.

(2) Calculate the mean and sd of the error series, along with a histogram.

(3) Plot the residual time series just as you would if it were a regular time series object. Plot a moving average of the residual time series as well. What do you notice? Would you characterize your residuals as being white noise? What does that mean?

(4) Create correlograms for the residual time series and plot it next to a correlogram for the underlying notice. What do you notice?

(5) Repeat this exercise with your other moving average object. What do you notice about the relationship between how well your moving average model fits the data and the nature of the residuals?

(6) For a given filter, plot the residual series and draw confidence bands at +/- 2 se. For points that are outside of these bounds, what does that mean? How does your interpretation change if your residual series had a dependency structure that you did not model properly?

Helpful R-commands

- filter
- plot
- acf
- pacf

Also, beware that creating a moving average object generates missing values. In order to deal with this issue when calcuating summary stats: mean(X, na.rm = TRUE); in order to deal with this issue when creating an acf plot: acf(X[!is.na(X)]).