Tutorial I : Detecting a breakpoint in a time series

Jan Verbesselt, Achim Zeileis

November 25, 2014

Abstract

This tutorial explain the basics of structural change test using with the BFAST concept. Basic principles are illustrated using the *Nile* data set (add reference).

1 Introduction

First load the BFAST package, which also loads the required packages for this small tutorial (i.e. strucchange, zoo).

```
> library("bfast")
> library("zoo")
> library("strucchange")
```

We will illustrate basic principles using the following time series of Annual Flow data of the river Nile. For more information also check the papers of Zeileis et al. (2005).

```
> plot(Nile, ylab="Annual Flow of the river Nile")
> abline(h= mean(Nile),col='blue')
```

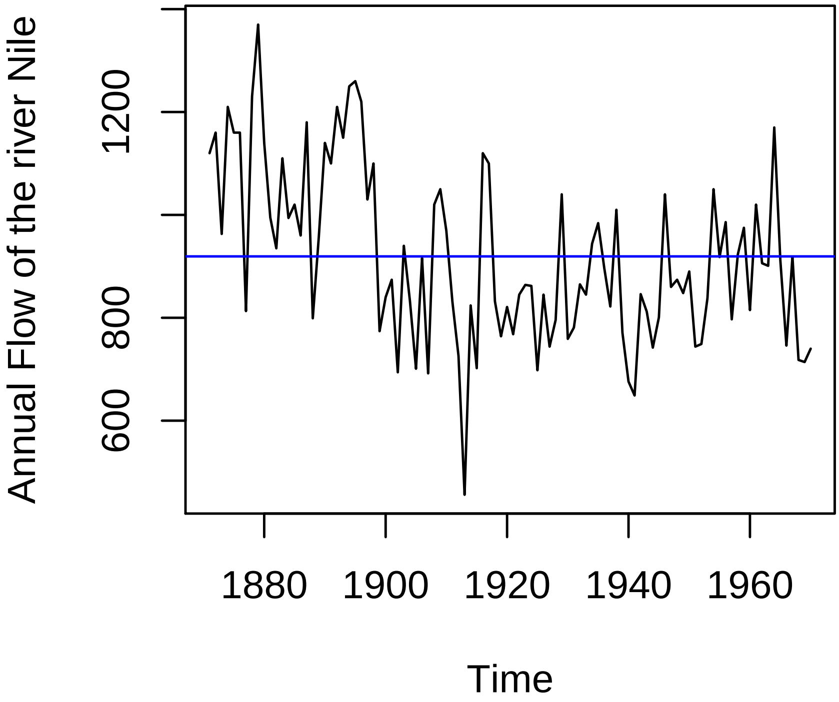


Figure 1: Annual river flow of the river Nile. The mean river flow (m3/sec) is plotted as a blue horizontal line.

2 CUMSUM and MOSUM illustrated

Here the CUSUM and MOSUM of the residuals of the Nile data using a constant (intercept) as explanatory variable in the model.

```
> plot(merge(
+      Nile = as.zoo(Nile),
+      zoo(mean(Nile), time(Nile)),
+      CUSUM = cumsum(Nile - mean(Nile)),
+      zoo(0, time(Nile)),
+      MOSUM = rollapply(Nile - mean(Nile), 15, sum),
+      zoo(0, time(Nile))
+    ), screen = c(1, 1, 2, 2, 3, 3), main = "", xlab = "Time",
+    col = c(1, 4, 1, 4, 1, 4)
+  )
```

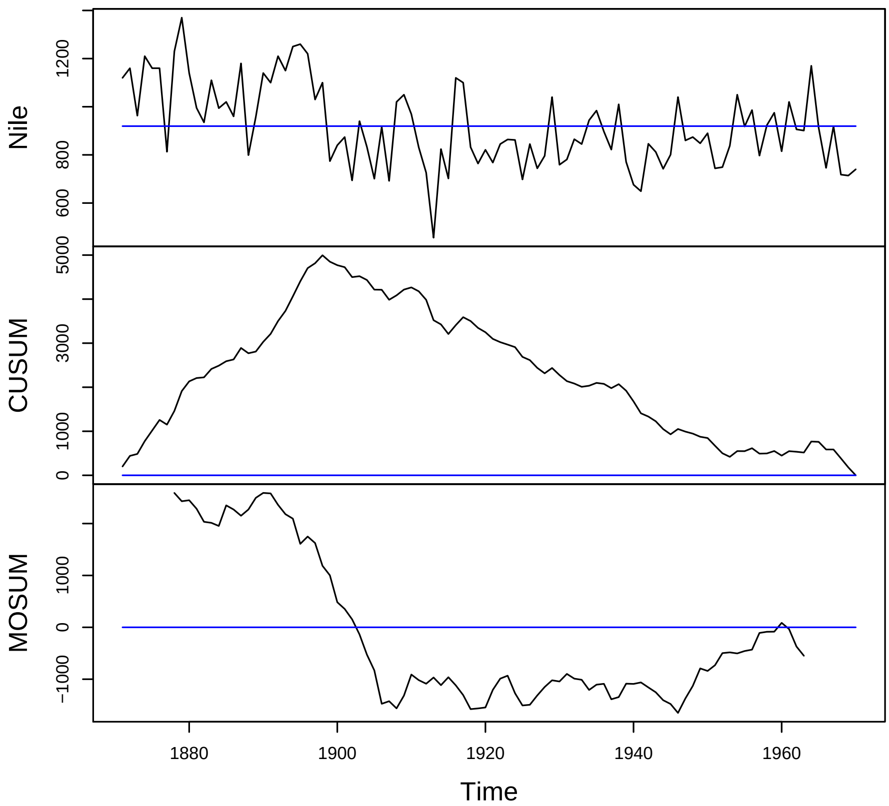
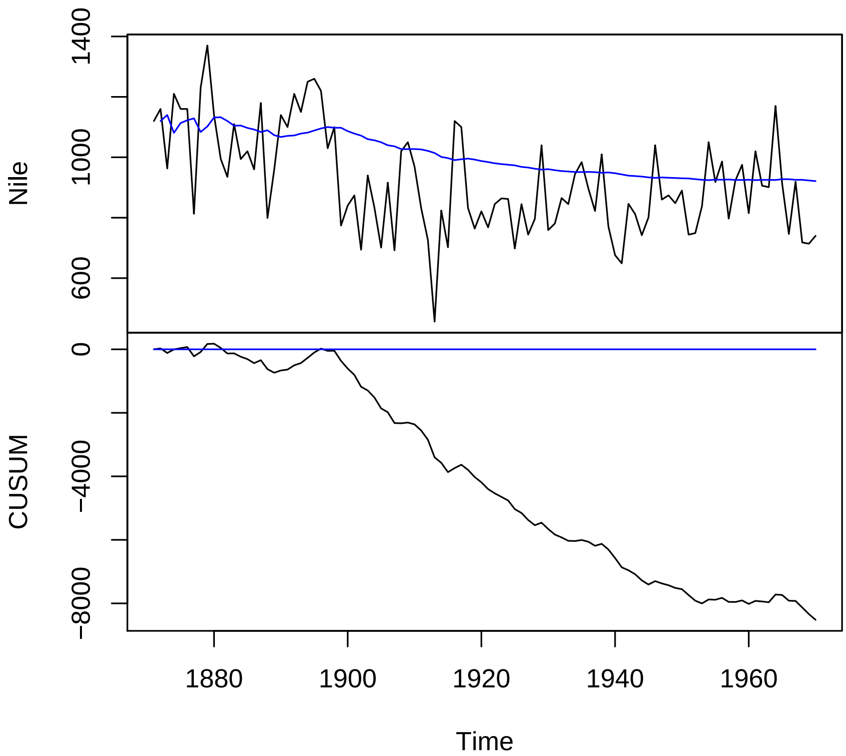


Figure 2: Nile data. MOSUM test. CUSUM test.

Fig. 2 illustrates how the CUSUM and MOSUM test evolve over time where CUSUM is a cumulative sum of the residuals and MOSUM is a moving sum (using a window of 15 years) of the residuals. The cumulative and moving sum of the residuals should fluctuate around zero (blue line) however significant deviation occurs from 1918 (?) onwards.

The recursive CUSUM illustrated.

```
> plot(merge(
+      Nile = as.zoo(Nile),
+      zoo(c(NA, cumsum(head(Nile, -1))/1:99), time(Nile)),
+      CUSUM = cumsum(c(0, recresid(lm(Nile ~ 1)))),
+      zoo(0, time(Nile))
+    ), screen = c(1, 1, 2, 2), main = "", xlab = "Time",
+    col = c(1, 4, 1, 4)
+  )
```



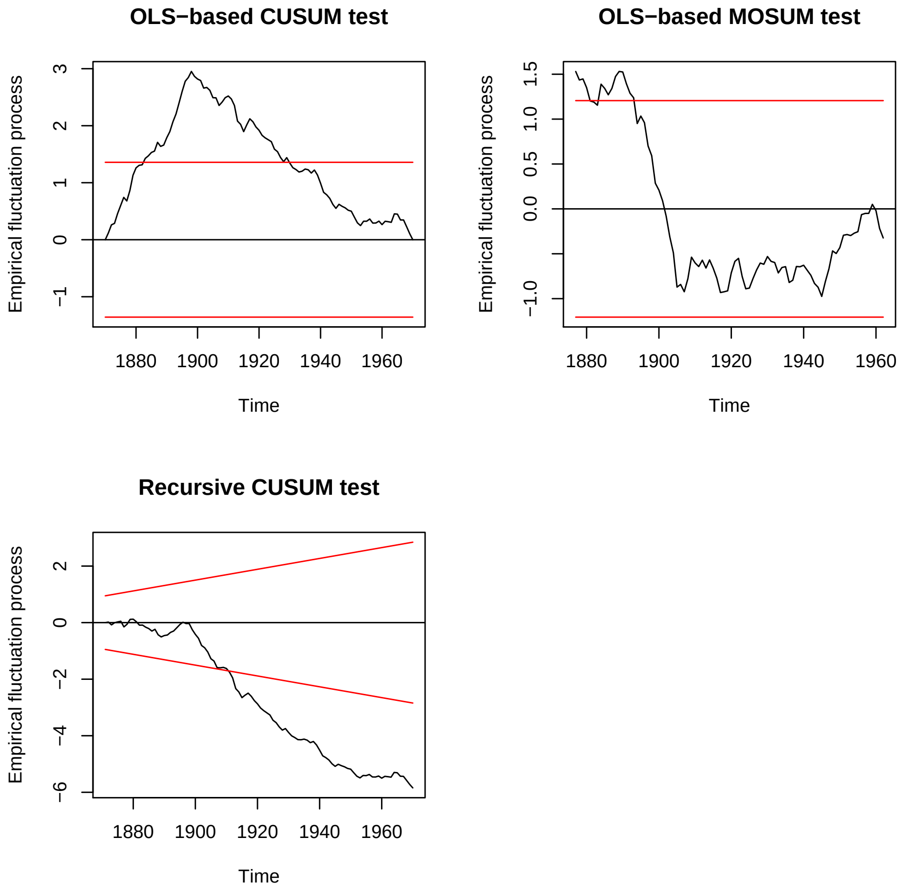
Rec-CUMSUM. The blue line in the first panel simply shows the mean of all observations prior to it, i.e., prediction[t] = mean(Nile[1:(t-1)]). The black line in the second panel shows the cumulated recursive residuals (= standardized one-step-prediction errors). You can see that up to the building of the dam, the residuals are approximately zero and then they start deviating from zero.

The following R lines reproduce the OLS-CUSUM, -MOSUM, and -Rec-CUSUM plots which are suitably scaled (see Zeileis et al. 2005)

```
> ocus.nile <- efp(Nile ~ 1, type = "OLS-CUSUM")
> omus.nile <- efp(Nile ~ 1, type = "OLS-MOSUM")
> rocus.nile <- efp(Nile ~ 1, type = "Rec-CUSUM")
```

These plots illustrate that a significant structural change is detected by the CUSUM and the MOSUM test. Especially the CUSUM illustrate that around 1910 the empiral fluctuation process based on the CUSUM residuals goes outside the boundaries. The MOSUM based plot less clearly illustrates the potential structural change that occurs (i.e. efp does not go outside the significance boundaries expect around early 1900. There is however a clear drop in the scaled MOSUM residuals visual which indicates a significant change in the data variation of the residuals.

```
> opar <- par(mfrow=c(2,2))
> plot(ocus.nile)
> plot(omus.nile)
> plot(rocus.nile)
> par(opar)
```

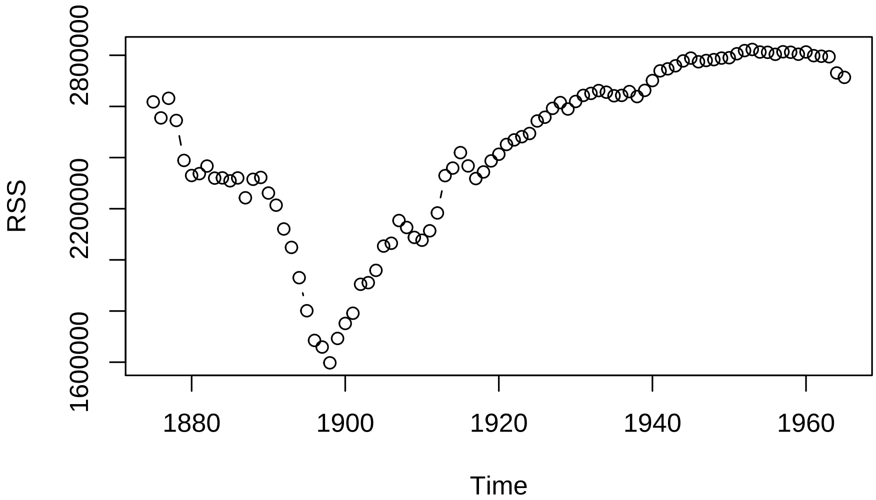


3 Fitting a piecewise linear model

The following section illustrate how the location of the structural change in the Nile time series is determined.

First, determine the the minimum Residual Sum of Squares RSS to determine the position of the breakpoint:

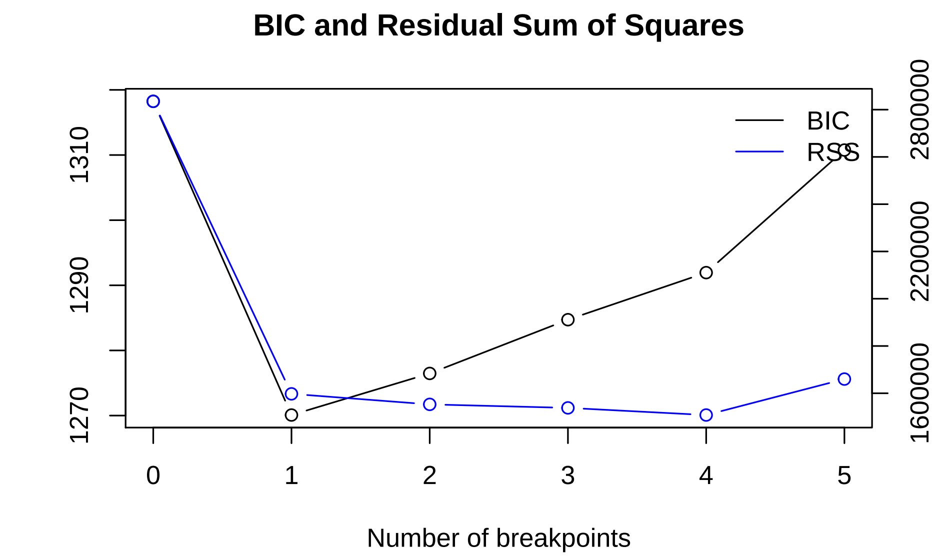
```
> plot(1870 + 5:95, sapply(5:95, function(i) {
+    before <- 1:i
+    after <- (i+1):100
+    res <- c(Nile[before] - mean(Nile[before]), Nile[after] - mean(Nile[after]))
+    sum(res^2)
+ }), type = "b", xlab = "Time", ylab = "RSS")
```



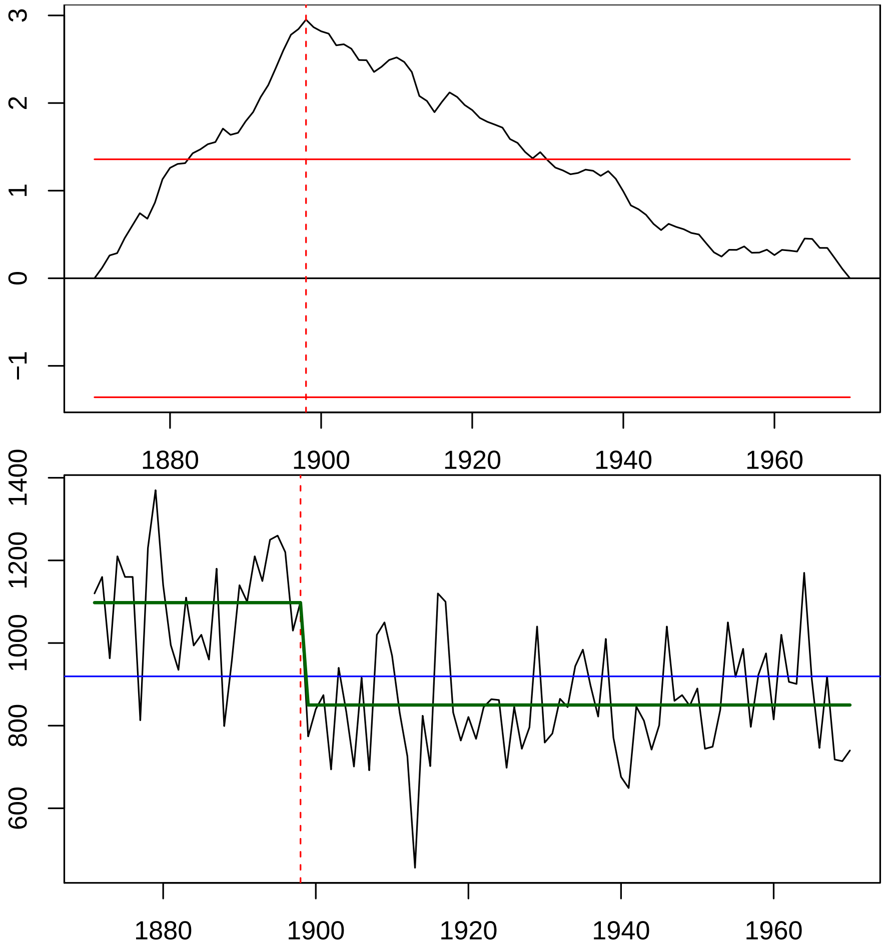
Second, use the strucchange functionality to determine the *Date* of change.

```
> bp.nile <- breakpoints(Nile ~ 1)
> nile.fac <- breakfactor(bp.nile, breaks = 1 )
> fm1.nile <- lm(Nile ~ nile.fac - 1)

> plot(bp.nile)
```



```
> opar <- par(mfrow=c(2,1), mar=c(2,2,0,2))
> plot(ocus.nile, alt.boundary=F,main="")
> abline(v= 1898, lty=2, col='red')
> plot(Nile, ylab="Annual Flow of the river Nile")
> abline(h= mean(Nile),col='blue')
> abline(v= 1898, lty=2, col='red')
> lines(ts(predict(fm1.nile),start=1871,freq=1), col='darkgreen',lwd=2)
> par(opar)
```



4 More information

More information can be found on the following website http://bfast.r-forge.r-project.org/ and in the BFAST papers mentioned on that website.