Time Series Analysis of the Level of Lake Huron

Alex Trindade

November 21, 2005

Abstract

A time series plot of the levels of Lake Huron from 1875 to 1972 suggests a steady decline. Fitting a linear regression trend in time with an ARMA covariance structure for the residual errors, leads to the conclusion that the decline is indeed statistically significant. The fitted model is used to forecast the next 3 years of lake levels.

1   Introduction

Scientists have historically used the North American Great Lakes region as an indicator of long term environmental damage on the continent. Etc...

In this paper, we take a close look at the levels of Lake Huron, recorded as yearly averages in feet between the years 1875-1972. Figure 1 shows a time series plot of the data. As is immediately apparent, there seems to be a declining trend in lake levels. Our primary goal is to fit an appropriate model to the data in order to determine if this trend is statistically significant.

2   Methods & Results

Let $\{Y_t\}$ denote the stochastic process in time t, from which the observed lake data is considered to be a realization. To answer the primary question of interest, we will fit the model,

$$Y_t = \beta_1 + \beta_2 t + W_t, \tag{1}$$

where $\{W_t\}$ is a zero mean ARMA(p, q) process to be determined. To answer the question we will construct a 95% confidence interval for β_2, taking into account the correlation structure of $\{W_t\}$.

The correlation structure of $\{W_t\}$ can be incorporated into an ordinary least squares (OLS) regression model via generalized least squares (GLS), as described in Brockwell and Davis (2002), section 6.6. After obtaining the OLS estimates, the residuals from the fit comprise our first estimate of $\{W_t\}$, $\{\hat{W}_t\}$. Inspection of the sample ACF and PACF of $\{\hat{W}_t\}$, suggests that WN is not an appropriate model (Figure 2). An in-depth exploration suggests that an AR(2) might be a reasonable model.

With an AR(2) as a candidate model for $\{W_t\}$, we refine the OLS estimates by iterating between GLS estimation and ML estimation, until the model coefficients stabilize. This leads to the following fitted model for $\{W_t\}$,

$$W_t = 1.008 W_{t-1} - 0.295 W_{t-2} + Z_t, \qquad \{Z_t\} \sim \text{WN}(0, 0.4571),$$

and the following estimates (standard errors) for the regression coefficients,

$$\hat{\beta}_1 = 10.091 \ (0.463), \qquad \hat{\beta}_2 = -0.022 \ (0.008).$$

1

Figure 1: Level of Lake Huron 1875-1972 (feet), showing a trend line fitted via GLS with an AR(2) model for the residuals.
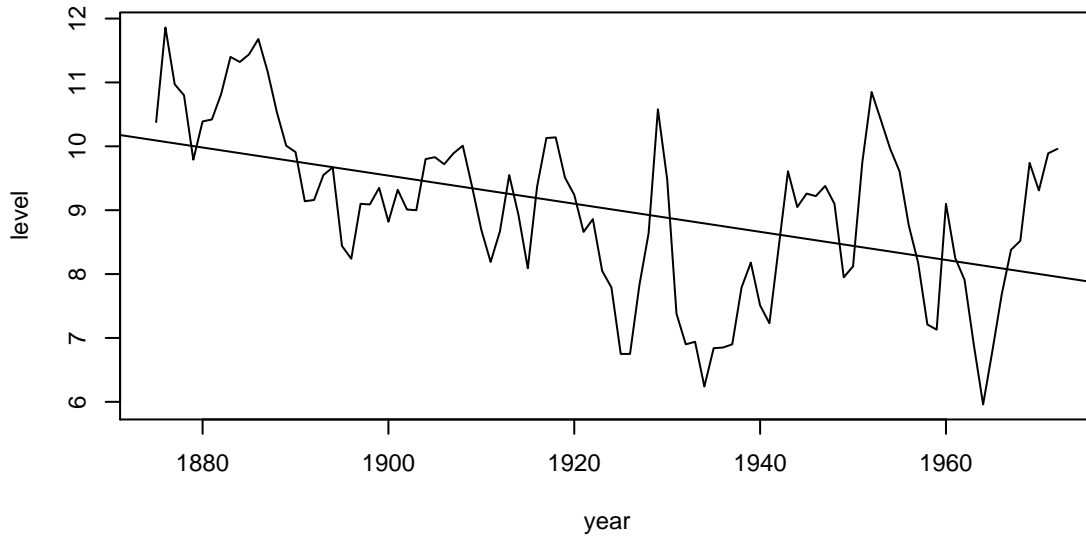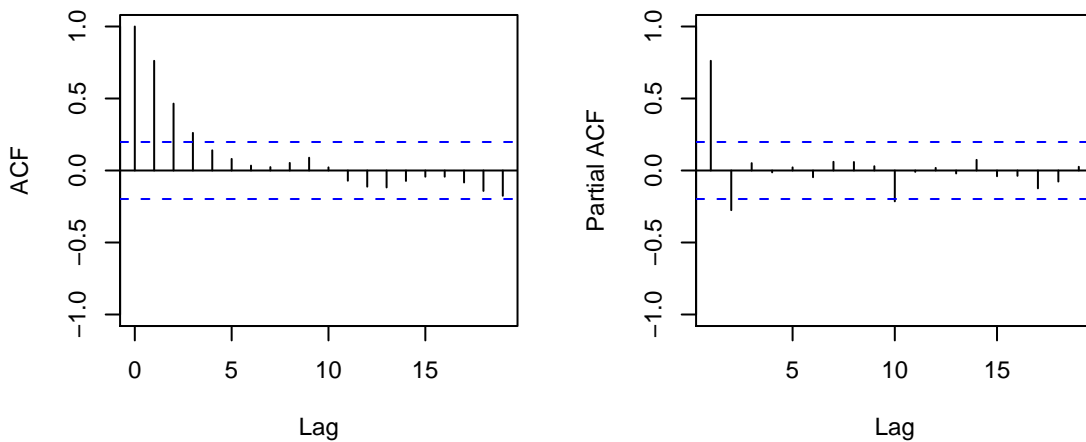

Figure 2: Sample ACF and PACF of the residuals $\{\hat{W}_t\}$ from an OLS regression fit to the Lake Huron data.

We therefore obtain the following 95% confidence interval for β_2, (-0.22 ± 0.0048), indicating a significant decline in the level of Lake Huron during the years 1875-1972.

Finally, we use the fitted model to forecast annual average lake levels for the next 3 years. The point forecasts as well as associated 95% confidence intervals are presented in Table 1.

Table 1: Point forecasts and associated 95% confidence intervals for the next three years of Lake Huron levels.

Year	Point forecast	95% confidence interval
1973	9.40	(8.07, 10.72)
1974	8.80	(6.93, 10.68)
1975	8.37	(6.27, 10.47)

References

[1] Brockwell, P.J. and Davis, R.A. (2002). *Introduction to Time Series and Forecasting* (2nd edn). New York: Springer-Verlag.

A R Code for the Figures

```
########## fitted line and time series plot of lake.tsm
library(ts)
#pdf(file="lake.pdf",horizontal=FALSE,width=6,height=3.5,pointsize=9)
postscript(file="lake.eps",horizontal=FALSE,width=6,height=3.5,pointsize=9)
x <- scan("lake.tsm")
myts <- ts(matrix(x, length(x), 1), start=1875)
ts.plot(myts, gpars=list(main="", xlab="year", ylab="level", lty=1))
abline(10.091+1875*0.022,-0.022)
dev.off()


### ACF/PACF of residuals from OLS model fit to lake.tsm
#pdf(file="lakeOLSres.pdf",horizontal=FALSE,width=6,height=3,pointsize=9)
postscript(file="lakeOLSres.eps",horizontal=FALSE,width=6,height=3,pointsize=9)
x <- scan("lake.tsm")
time <- seq(1,length(x))
myfit <- lm(x~time+1)
w <- myfit$residuals
par(mfrow=c(1,2))
acf(w,main="", ylim=c(-1,1))
pacf(w,main="", ylim=c(-1,1))
dev.off()
```