

Time Series Lab Handout: Predictive Analytics

Data Matters, Odum Institute, UNC/Chapel Hill, 25 June 2014

Instructor: Philip Schrodt, Parus Analytical Systems [schrodt735@gmail.com]

References

My favorite R crib site: <http://www.statmethods.net/index.html>

Nice R style guide: <https://google-styleguide.googlecode.com/svn/trunk/Rguide.xml>

A good time series reference:

<http://a-little-book-of-r-for-time-series.readthedocs.org/en/latest/src/timeseries.html>

Read the data and convert some variables to a time series

Get into the appropriate directory

```
ngadata <- read.table("Nigeria_1960-99.csv", header=TRUE, sep=",")
```

```
names(ngadata)
```

```
ngadata
```

```
imr.ts <- ts(ngadata$IMR, start = 1960)
```

```
imr.ts
```

```
fmlit.ts <- ts(100 - ngadata$ILLITFM, start=1960)
```

```
export.ts <- ts(ngadata$EXPORTS, start=1960)
```

A couple of basic time series plots

```
plot.ts(export.ts)
```

```
plot.ts(fmlit.ts)
```

```
twovar.ts <- cbind(fmlit.ts, imr.ts*10)
```

```
plot(twovar.ts, plot.type="single", col=c("blue", "red"), , ylab="", lty=1:2)
```

```
legend(1962, 50, legend=c("Female literacy", "Infant mortality"), col=c("blue", "red"), 1
```

Plot the ACF and CCF

```
acf(export.ts)
```

```
acf(fmlit.ts)
```

```
ccf(fmlit.ts, import.ts)
```

```
ccf(import.ts, fmlit.ts)
```

Now, the only way I can ever be sure that I'm interpreting the CCF correctly

```
export.l3 <- lag(export.ts,3)
lagdat <- cbind(export.ts,export.l3)
plot.ts(lagdat, plot.type = "single")
lagdat[1:16,]
ccf(export.ts,export.l3)
```

Running a basic predictive model

```
ngadata$riots.f2 <- c(ngadata$RIOTS[3:length(ngadata$RIOTS)], NA, NA)
ngadata[1:8,]
fit <- lm(riots.f2 ~ DEMOC + ILLITFM + GDPCAP + RIOTS, data=ngadata)
summary(fit)
fit <- lm(riots.f2 ~ DEMOC + ILLITFM + GDPCAP, data=ngadata)
summary(fit)
acf(ngadata$RIOTS)

ngadata$gdpcap.f2 <- c(ngadata$GDPCAP[3:length(ngadata$GDPCAP)], NA, NA)
ngadata[1:8,]
fit <- lm(gdpcap.f2 ~ DEMOC + ILLITFM + EXPORTS + GDPCAP, data=ngadata)
summary(fit)
fit <- lm(gdpcap.f2 ~ DEMOC + ILLITFM + EXPORTS, data=ngadata)
summary(fit)
acf(ngadata$GDPCAP)
```

Let's look at some diagnostics

```
resid.ts <- ts(residuals(fit), start = 1960)
acf(resid.ts)
pred.ts <- ts(fitted(fit), start = 1960)
gd.ts <- ts(ngadata$gdpcap.f2,start = 1960)
twovar <-cbind(gd.ts,pred.ts)
plot(twovar, plot.type="single", col=c("blue", "red"), , ylab="", lty=1:2)
```

Now, try this with the riots.f2 model and note the differences in the acf(resid)