Time Series Analysis

Basics of Time Series Analysis:
Data Example

Nicoleta Serban, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

Trend and Seasonality Estimation of ED Volume Time Series

Georgia Tech

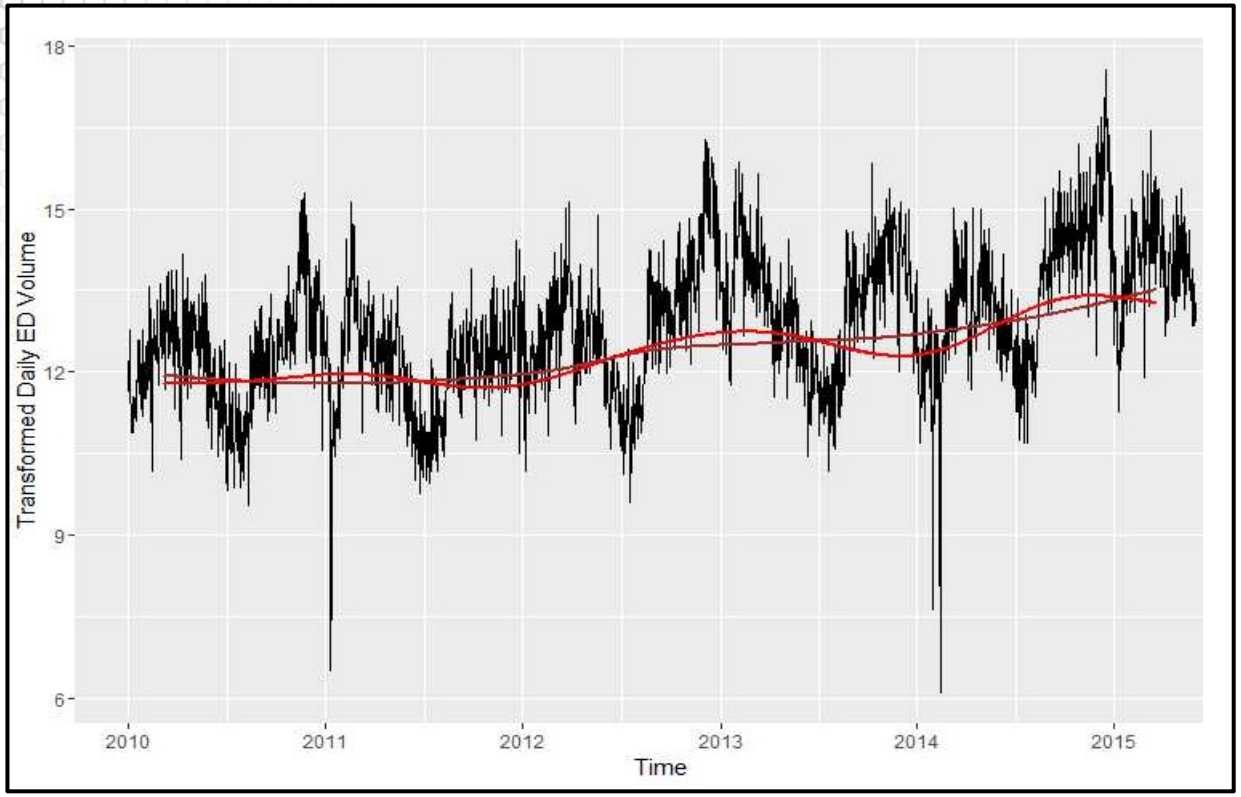
About This Lesson

Trend Estimation

```
## Equally spaced time points
time.pts = c(1:length(Volume))
time.pts = c(time.pts - min(time.pts))/max(time.pts)
## Local Polynomial Trend Estimation
loc.fit = loess(Volume.tr~time.pts)
vol.fit.loc = fitted(loc.fit)
## Splines Trend Estimation
library(mgcv)
gam.fit = gam(Volume.tr~s(time.pts))
vol.fit.gam = fitted(gam.fit)
## Is there a trend?
plot(dates,Volume.tr,ylab="ED Volume")
lines(dates,vol.fit.loc,lwd=2,col="brown")
lines(dates,vol.fit.gam,lwd=2,col="red")
```

Georgia Tech

Trend Estimation



Georgia Tech

Trend Estimation (cont'd)

```
Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.85769    0.02441   526.7   <2e-16
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05

Approximate significance of smooth terms:
              edf Ref.df     F p-value
s(time.pts) 8.628   8.96 93.39 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05

R-sq.(adj) =  0.296   Deviance explained = 29.9%
```

Smooth trend is statistically significant

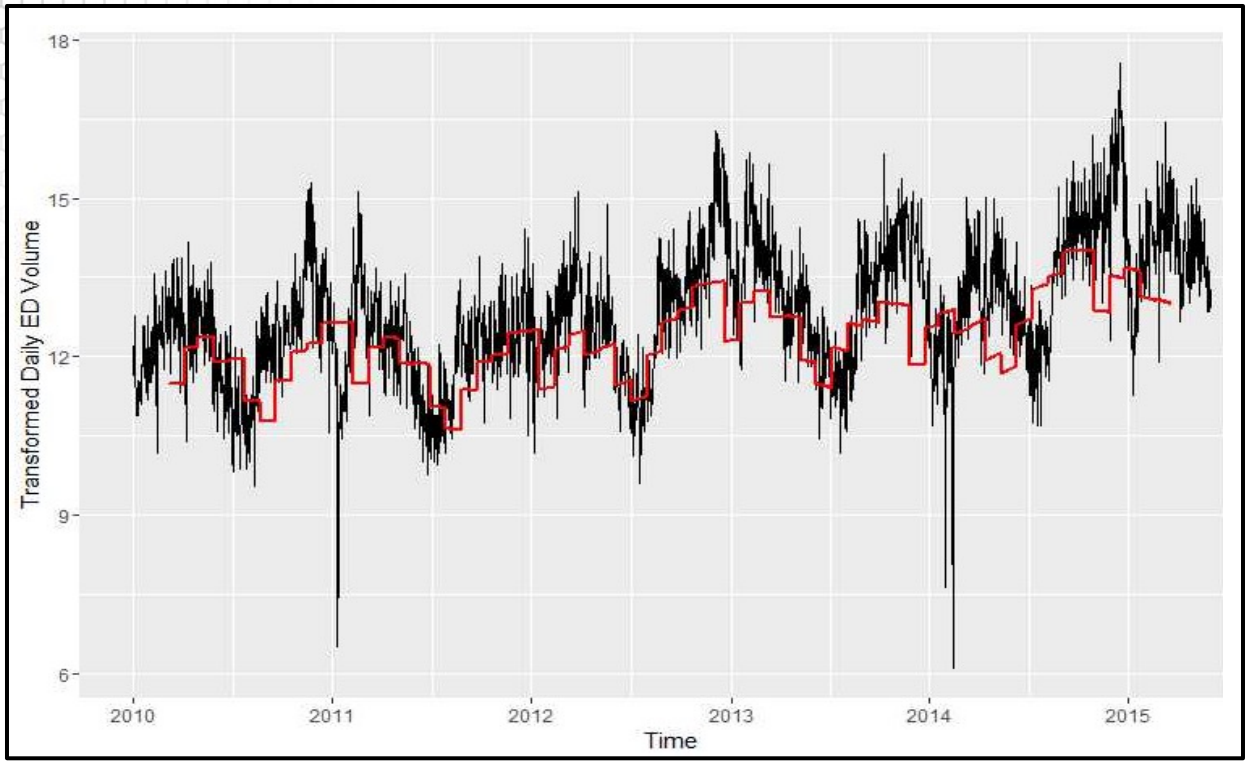
29.6% of variability explained

Trend and Seasonality Estimation

```
## Model Trend + Monthly Seasonality
month = as.factor(format(dates,"%b"))
gam.fit.seastr.1 = gam(Volume.tr~s(time.pts)+month-1)
summary(gam.fit.seastr)
vol.fit.gam.seastr.1 = fitted(gam.fit.seastr.1)
ggplot(edvoldata, aes(dates, sqrt(Volume+3/8))) + geom_line() + xlab("Time") +
ylab("Transformed Daily ED Volume")
lines(dates,vol.fit.gam.seastr.1,lwd=2,col="red")
```

Georgia Tech

Trend and Seasonality Estimation



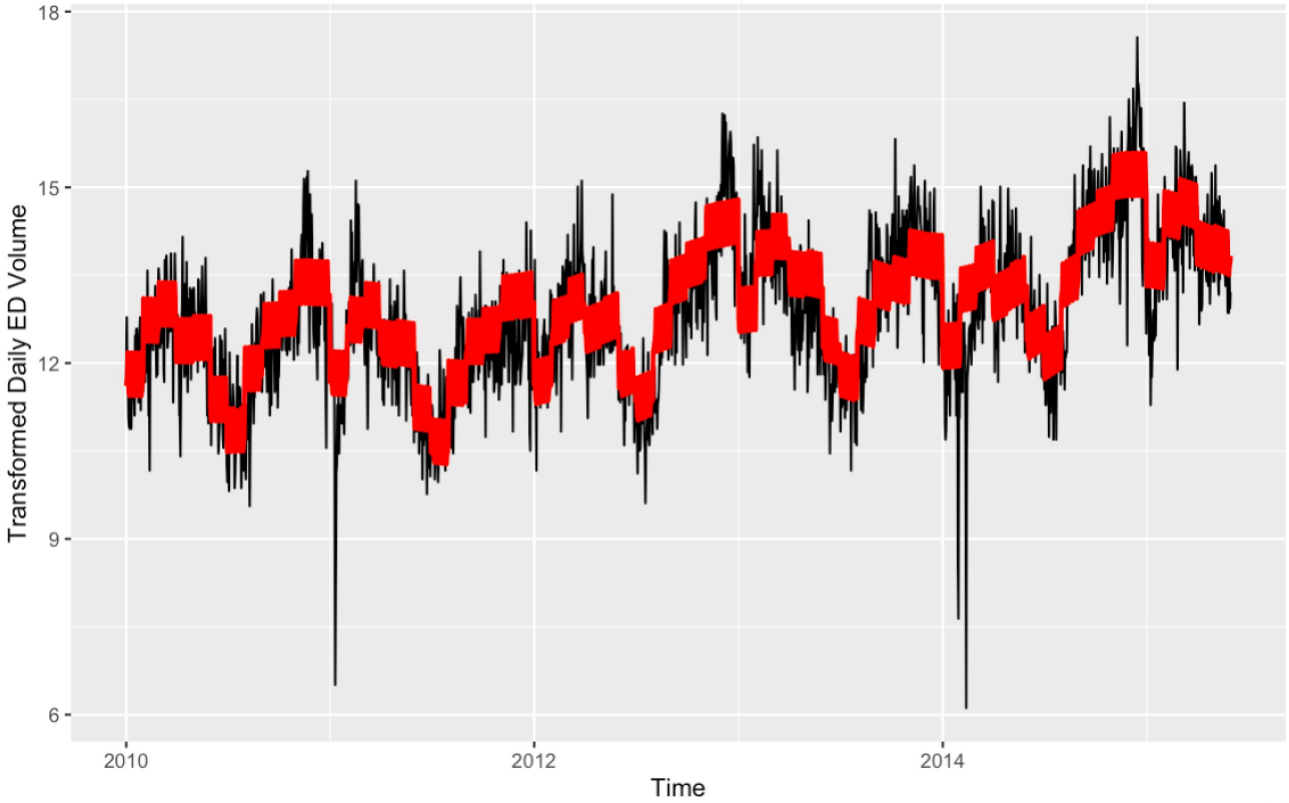
Georgia Tech

Trend and Seasonality Estimation (cont'd)

```
## Add day-of-the-week seasonality
week = as.factor(weekdays(dates))
gam.fit.seastr.2 = gam(Volume.tr~s(time.pts)+month+week)
summary(gam.fit.seastr.2)
vol.fit.gam.seastr.2 = fitted(gam.fit.seastr.2)
## Compare the two fits
ggplot(edvoldata, aes(dates, vol.fit.gam.seastr.2)) + geom_line() + xlab("Time") +
ylab("Seasonality and Trend: Daily ED Volume")
lines(dates,vol.fit.gam.seastr.1,lwd=2,col="red")
```

Georgia Tech

Trend and Seasonality Estimation (cont'd)



Trend and Seasonality Estimation (cont'd)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	12.77772	0.07383	173.063	< 2e-16
monthAug	-0.50061	0.08852	-5.655	1.79e-08
monthDec	0.93273	0.08791	10.611	< 2e-16
monthFeb	0.33589	0.08509	3.948	8.17e-05
monthJan	-0.59767	0.08365	-7.145	1.26e-12
monthJul	-1.53530	0.08831	-17.385	< 2e-16
monthJun	-1.00553	0.08880	-11.324	< 2e-16
monthMar	0.61815	0.08272	7.472	1.18e-13
monthMay	0.05271	0.08272	0.637	0.52403
monthNov	0.94163	0.08900	10.580	< 2e-16
monthOct	0.41557	0.08849	4.696	2.84e-06
monthSep	0.22391	0.08935	2.506	0.01229
weekMonday	0.57169	0.06647	8.601	< 2e-16
weekSaturday	0.04589	0.06641	0.691	0.48967
weekSunday	0.17538	0.06641	2.641	0.00834
weekThursday	-0.16031	0.06647	-2.412	0.01597
weekTuesday	0.08099	0.06647	1.218	0.22323
weekWednesday	-0.11050	0.06647	-1.662	0.09662

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	
s(time.pts)	8.792	8.987	151.8	<2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

R-sq.(adj) = 0.627 Deviance explained = 63.2%

GCV = 0.63263 Scale est. = 0.62405 n = 1977

Most regression coefficients are statistically significant

Some regression coefficients are statistically significant

Smooth trend is statistically significant

62.7% of variability explained

Georgia Tech

Trend and Seasonality Estimation (cont'd)

Does the addition of seasonality of day of the week adds predictive power?

lm.fit.seastr.1 = lm(Volume.tr~month)
lm.fit.seastr.2 = lm(Volume.tr~month+week)
anova(lm.fit.seastr.1,lm.fit.seastr.2)

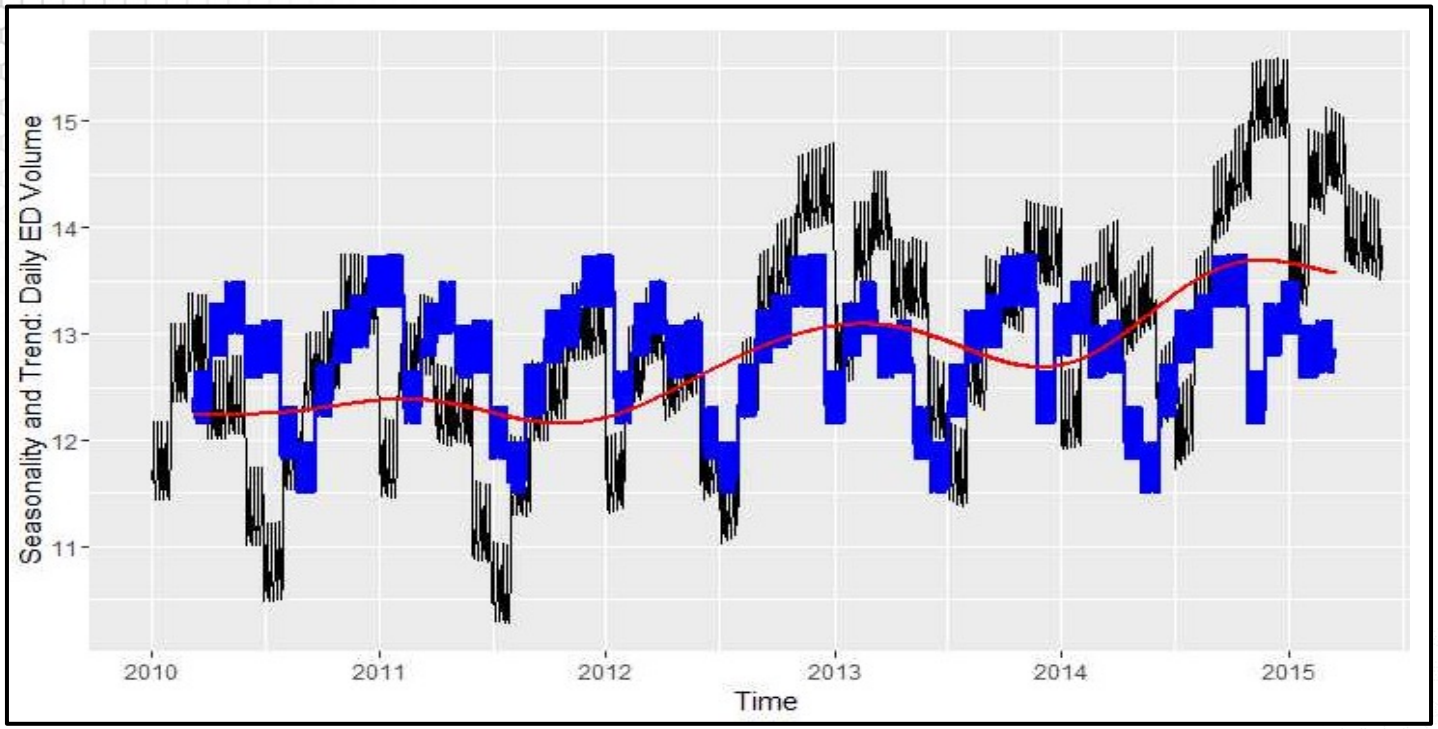
```
Analysis of Variance Table

Model 1: Volume.tr ~ month
Model 2: Volume.tr ~ month + week
  Res.Df    RSS Df Sum of Sq     F    Pr(>F)
1   1965 2169.9
2   1959 2071.1  6    98.826 15.58 < 2.2e-16
```

The seasonality due to day of the week improves the prediction power of the model

Seasonality vs Trend

```
## Compare with & without trend
ggplot(edvoldata, aes(dates, vol.fit.gam.seastr.2)) + geom_line() + xlab("Time") +
ylab("Seasonality and Trend: Daily ED Volume")
lines(dates,vol.fit.lm.seastr.2,lwd=2,col="blue")
lines(dates,vol.fit.gam,lwd=2,col="red")
```

Seasonality vs Trend



Summary