Time Series Analysis

ARMA Models: Data Examples

Nicoleta Serban, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering

U.S. Fuel Consumption:

Exploratory Data Analysis

Georgia Tech

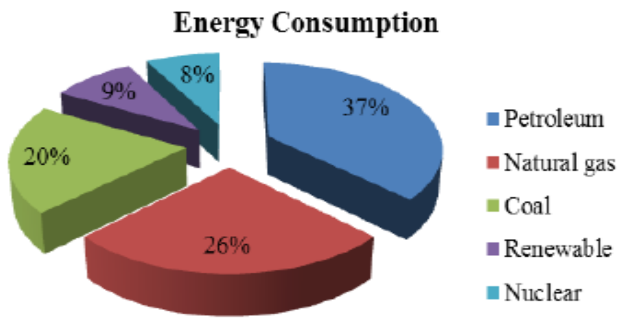
About This Lesson

Energy Consumption

- **Data Source:** U.S. Department of Energy, The Energy Information Administration (EIA): Monthly Electricity Consumption in million megawattshours (United States Lower 48 region) over 1985-2019

Questions of interest:

- What are common characteristics in energy consumption by source?
- Can we predict energy consumption over the course of a year?



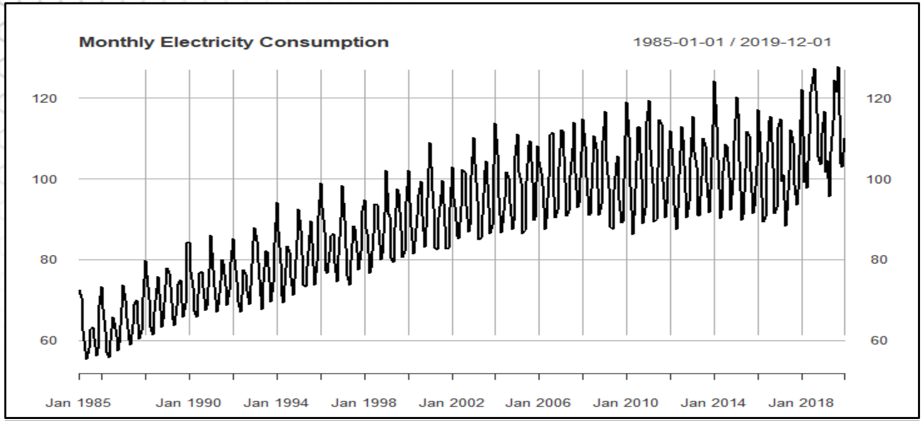
Georgia Tech

Time Series Plots

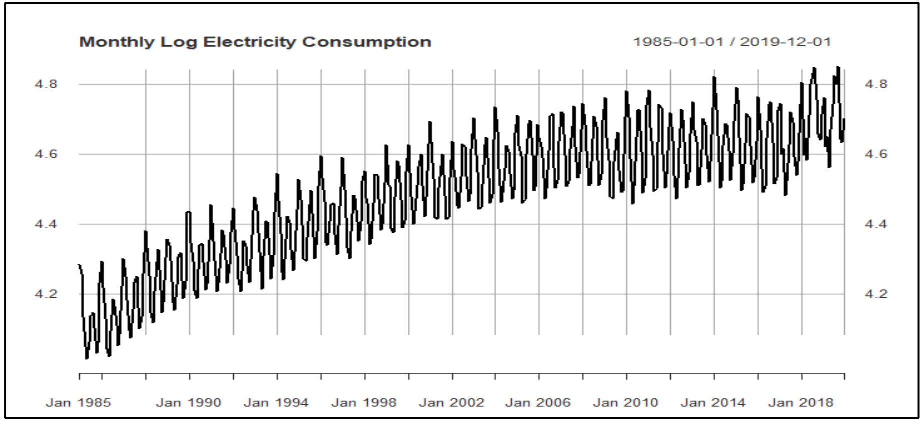
```
## Read the data
US_Energy_Monthly_Consump = read.csv("Electric_Monthly_Consump.csv")
dates = as.character(US_Energy_Monthly_Consump$DATE)
dates = as.Date(dates,format = "%m/%d/%y")
consumption = xts(US_Energy_Monthly_Consump[,2], as.Date(dates))
plot(consumption)

#The time series displays an increasing variability over time
lconsumption = log(consumption)
lconsumption = xts(lconsumption, as.Date(dates))
plot(lconsumption)
```

Time Series Plots

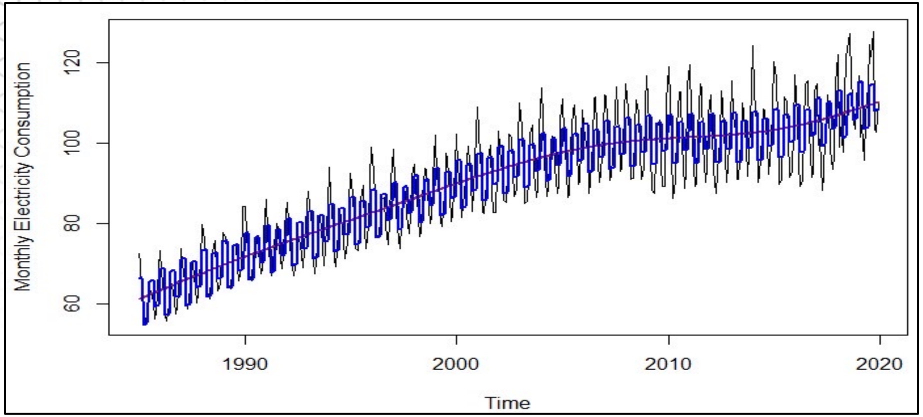


Time-Varying Variability

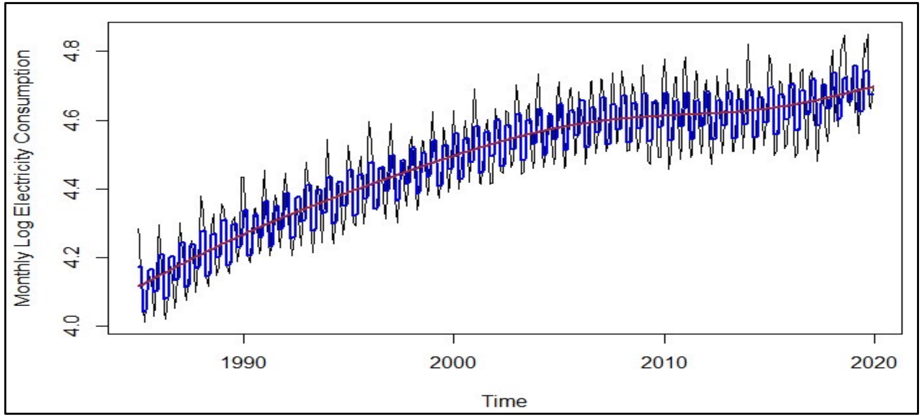


Log: Variance Stabilizing Transformation

Why transforming?

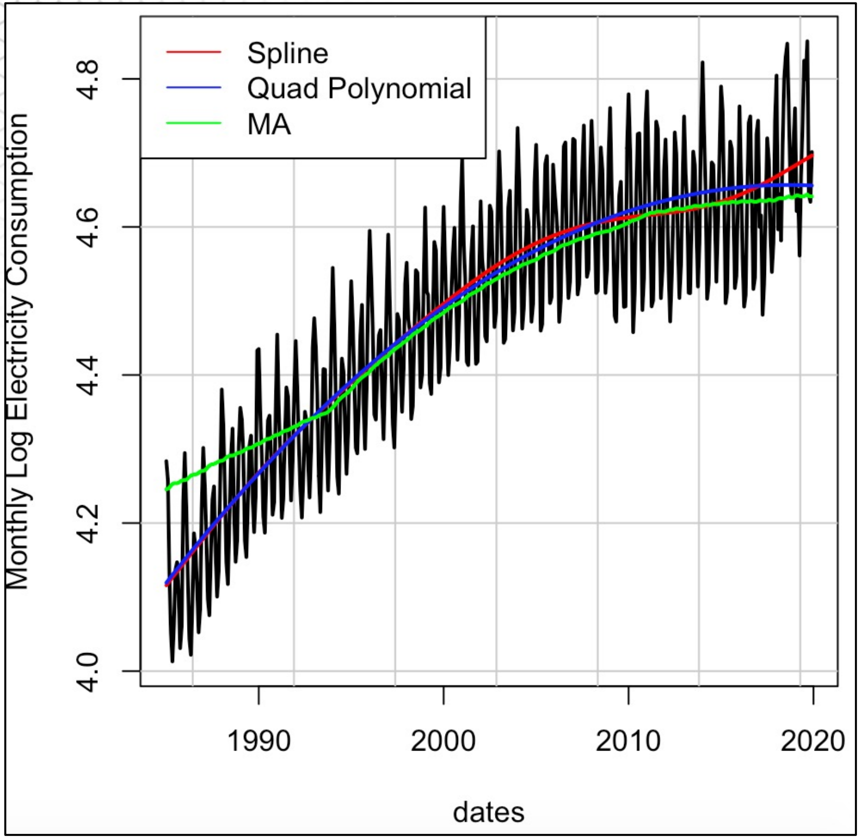


Time-Varying Variability: The residuals (difference between data and fitted blue curve) gradually increases as time increases



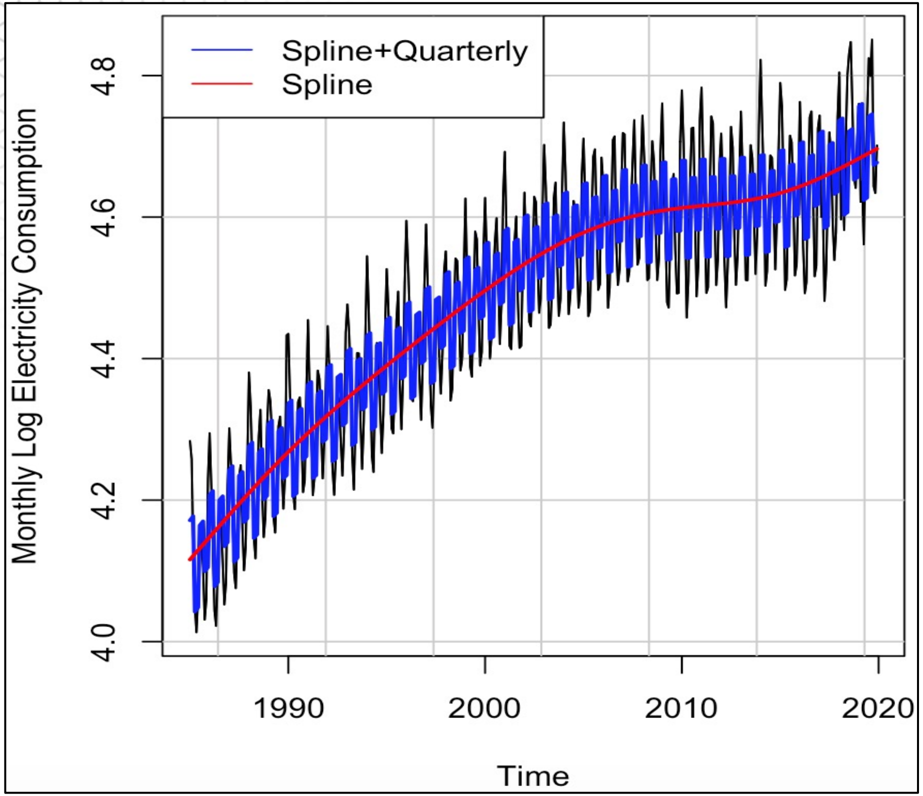
The residuals seem to be constant over time. Thus, log transform is needed to eliminate the “horn” shape in the times series.

Trend Estimation



- Observe a positive log trend in the log-transformed time series.
- Among the three trend fitting methods, spline seems to best capture the trend.

Trend & Quarterly Seasonality



	Estimate	Std. Error	t value	Pr(>ltl)
seasonsQ1	4.549463	0.006442	706.2	<2e-16 ***
seasonsQ2	4.411546	0.006438	685.2	<2e-16 ***
seasonsQ3	4.524308	0.006438	702.7	<2e-16 ***
seasonsQ4	4.450629	0.006442	690.9	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

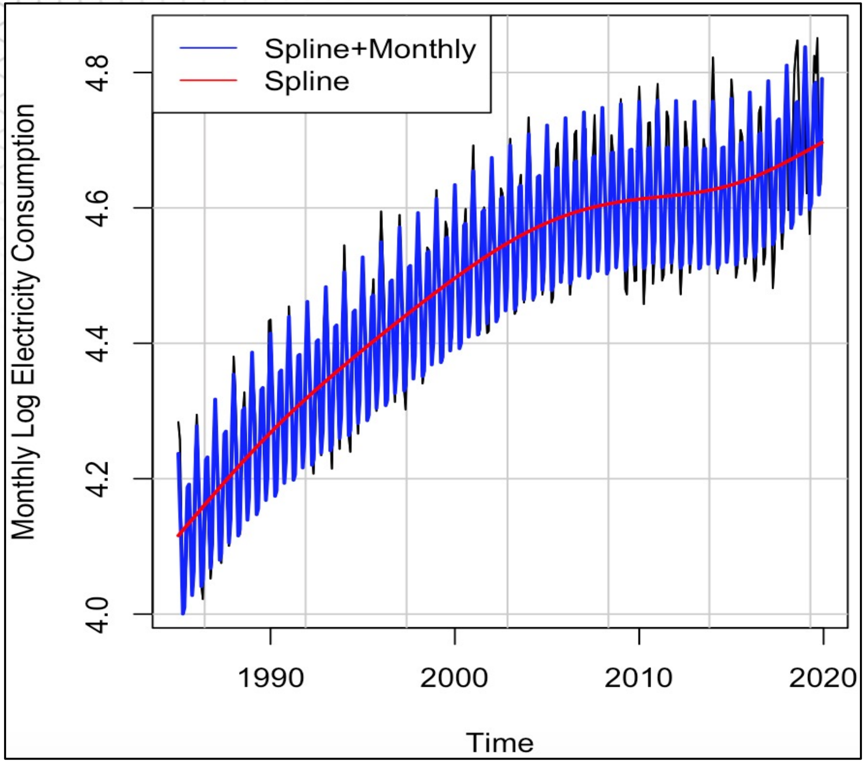
Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(time.pts)	6.119	7.279	349.5	<2e-16 ***

Trend and seasonality are both statistically significant

Georgia Tech

Trend & Monthly Seasonality



	Estimate	Std. Error	t value	Pr(>|t|)
seasons_month1	4.623954	0.005777	800.4	<2e-16 ***
seasons_month2	4.544303	0.005776	786.8	<2e-16 ***
seasons_month3	4.480860	0.005775	775.9	<2e-16 ***
seasons_month4	4.377087	0.005775	758.0	<2e-16 ***
seasons_month5	4.383215	0.005774	759.1	<2e-16 ***
seasons_month6	4.474589	0.005774	775.0	<2e-16 ***
seasons_month7	4.553869	0.005774	788.7	<2e-16 ***
seasons_month8	4.554755	0.005774	788.8	<2e-16 ***
seasons_month9	4.464061	0.005775	773.1	<2e-16 ***
seasons_month10	4.384087	0.005775	759.1	<2e-16 ***
seasons_month11	4.416126	0.005776	764.6	<2e-16 ***
seasons_month12	4.550933	0.005777	787.8	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

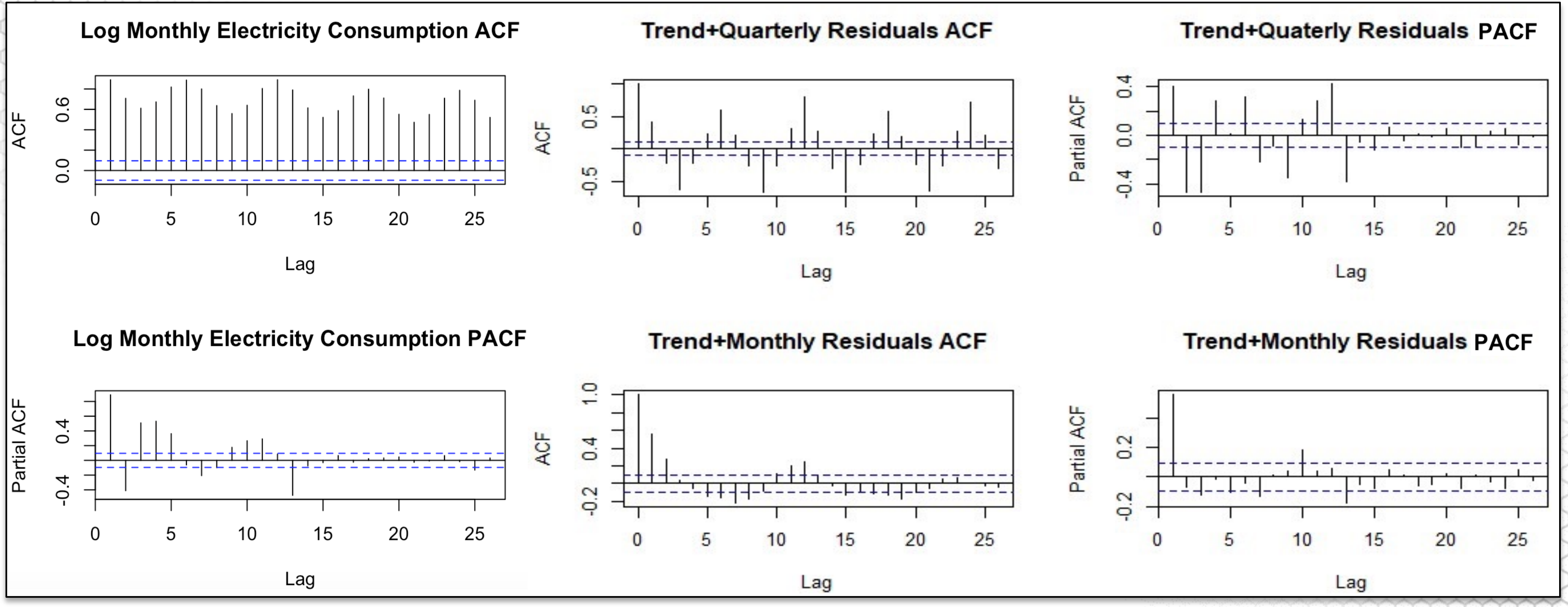
Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(time.pts)	7.969	8.72	1090	<2e-16 ***

Trend and seasonality are both statistically significant

Georgia Tech

Trend & Seasonality: Residual Analysis



Summary