INSEAD

The Business School for the World®

Prof. Anton Ovchinnikov

Prof. Spyros Zoumpoulis

DSB Classes 03-04, January 26, 2018

- Time Series Models

Plan for the day
Learning objectives

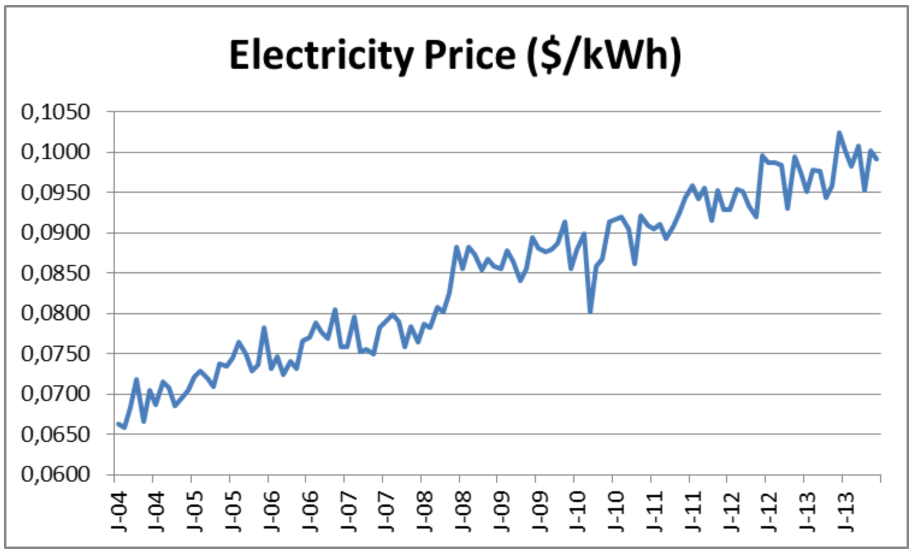
INSEAD
The Business School for the World®

- Session 03-04 learning objectives:
  - Conceptual introduction to Time Series modeling: noise, trend and seasonality
  - Methods for Time Series modeling, and their R implementations

Timeseries

INSEAD
The Business School for the World®

- **Definition**: a time series is a series of data points indexed (or listed or graphed) in time order
- Example: monthly electricity prices in the state of California

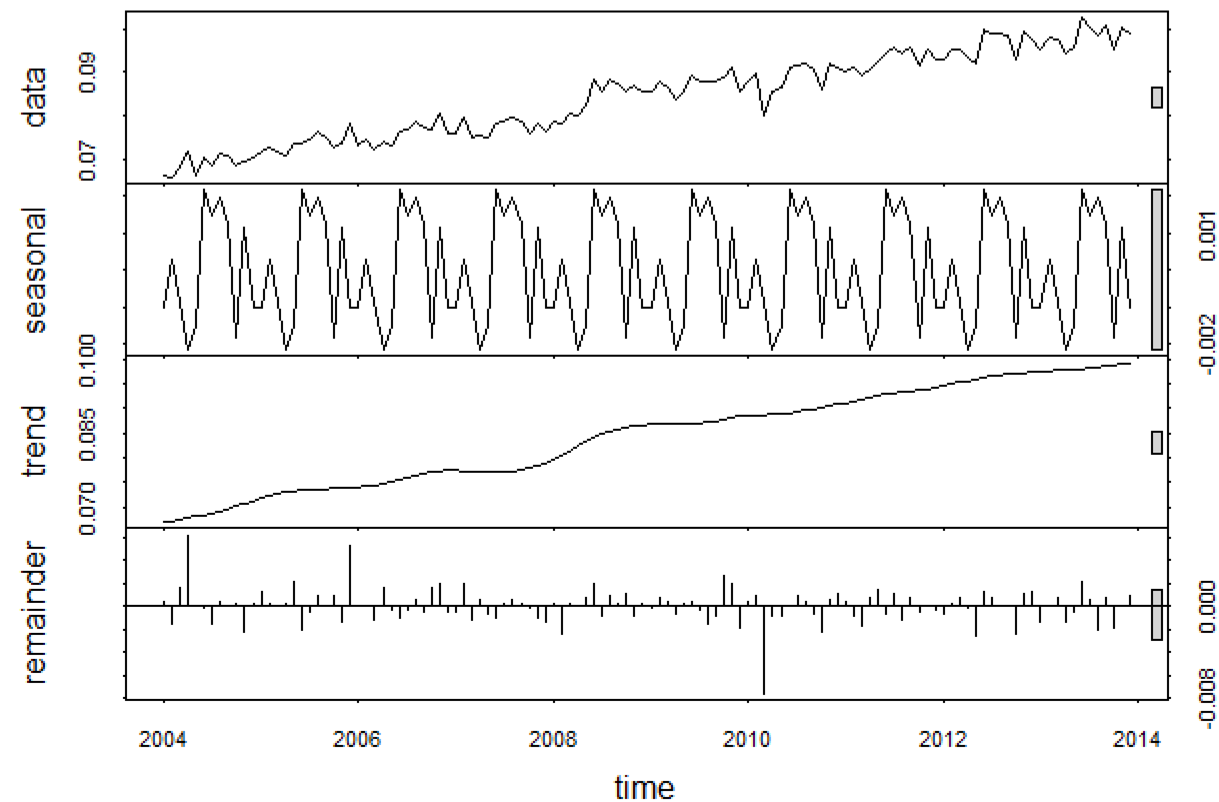


  - What do you “see” from/on the graph?
  - BTW, why do we need any special methods for time series? Why will regression not be quite sufficient?

This plot: file “0304 Time Series Examples…” on portal

Understanding timeseries: noise, level, trend, seasonality

INSEAD
The Business School for the World®



Date: 0304 electric rates.csv on portal

Conceptual understnaing of noise, level, and trend

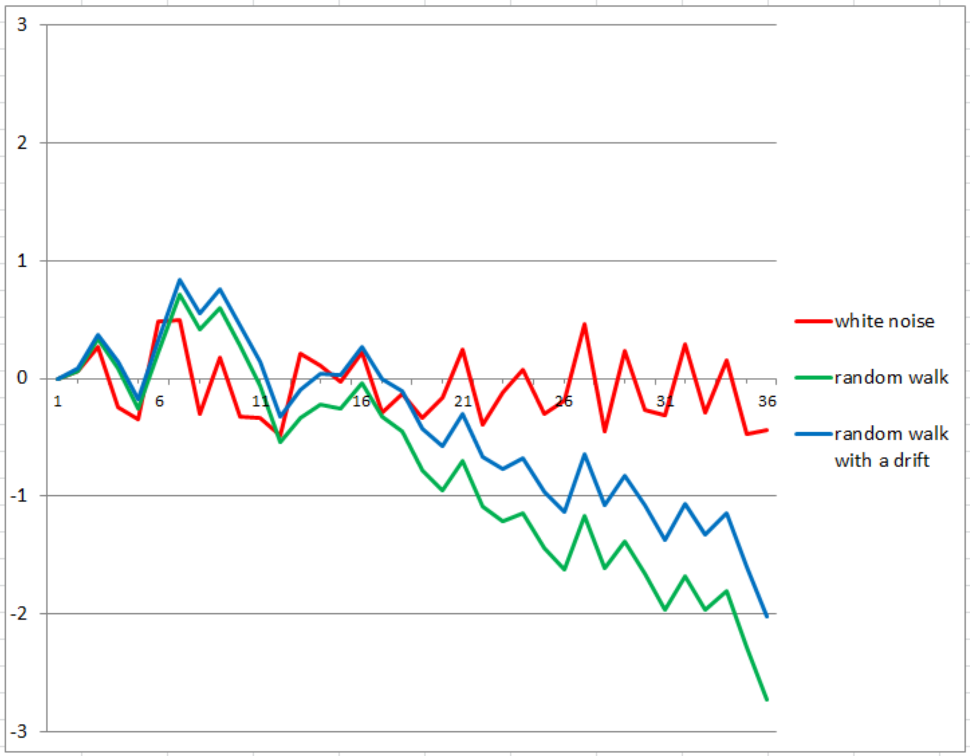
INSEAD
The Business School for the World®

	A	B	C	D
1				0,02
2	month	white noi:	random w	random walk with a drift
3	1	0	0	0
4	2	0,063765	0,063765	0,083765
5	3	0,268283	0,332048	0,372048
6	4	-0,24494	0,087109	0,147109
7	5	-0,34402	-0,25692	-0,17692
8	6	0,480934	0,224019	0,324019
9	7	0,49198	0,715999	0,835999
10	8	-0,29916	0,416842	0,556842
11	9	0,179109	0,595951	0,755951
12	10	-0,3208	0,275148	0,455148
13	11	-0,33511	-0,05997	0,140035
14	12	-0,48094	-0,5409	-0,3209
15	13	0,208396	-0,33251	-0,09251
16	14	0,112326	-0,22018	0,03982
17	15	-0,03069	-0,25087	0,029125
18	16	0,216677	-0,0342	0,265802
19	17	-0,2879	-0,3221	-0,0021
20	18	-0,12668	-0,44877	-0,10877
21	19	-0,33757	-0,78634	-0,42634
22	20	-0,16417	-0,95051	-0,57051
23	21	0,247773	-0,70273	-0,30273
24	22	-0,38823	-1,09097	-0,67097
25	23	-0,12052	-1,21148	-0,77148
26	24	0,069519	-1,14197	-0,68197
27	25	-0,30347	-1,44543	-0,96543
28	26	-0,18409	-1,62952	-1,12952
29	27	0,466844	-1,16268	-0,64268
30	28	-0,45246	-1,61514	-1,07514
31	29	0,231228	-1,38391	-0,82391
32	30	-0,27339	-1,6573	-1,0773
33	31	-0,31254	-1,96983	-1,36983
34	32	0,285999	-1,68383	-1,06383
35	33	-0,28586	-1,9697	-1,3297
36	34	0,159024	-1,81067	-1,15067
37	35	-0,47562	-2,28629	-1,60629
38	36	-0,43901	-2,7253	-2,0253
39				

=RAND()-0,5

=C3+B4

=D3+B4+$D$1



What is your forecast for months 37? 38? 39?

File “0304 Time Series Examples….xls” on portal (press F9 to refresh

Conceptual understnaing of seasonality

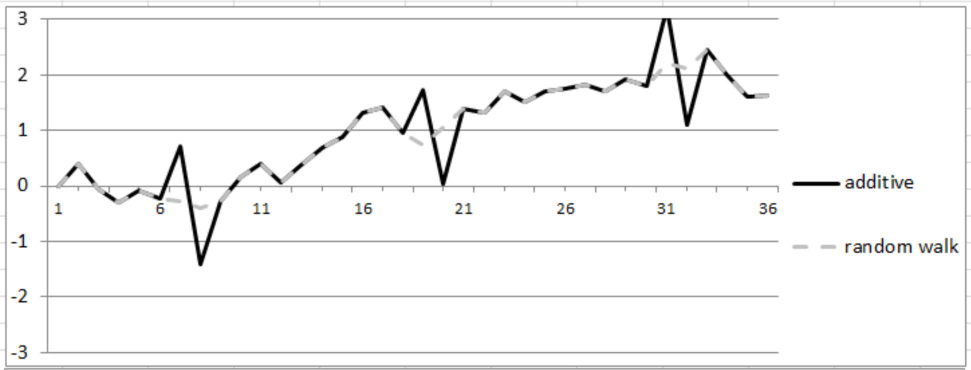
INSEAD
The Business School for the World®

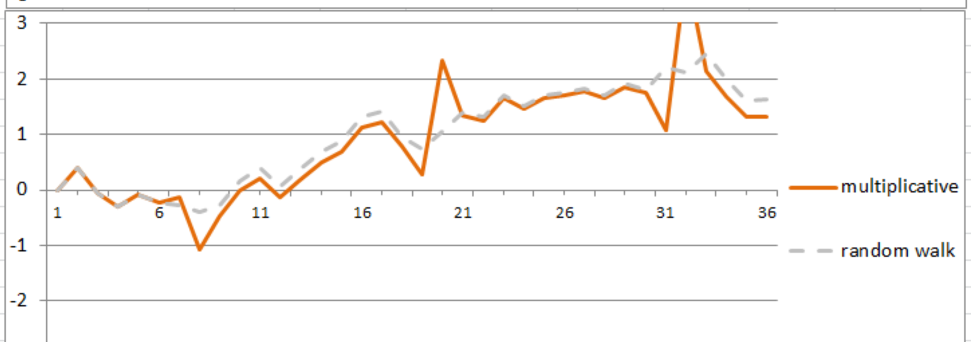
	A	B	D	E	F
1				seasonality	
2	month	month inc	random w	additive	multiplicative
3	1	Jan	0	0	0
4	2	Feb	0,399443	0,399443	0,399443
5	3	March	-0,05734	-0,05734	-0,05734
6	4	April	-0,31371	-0,31371	-0,31371
7	5	May	-0,08643	-0,08643	-0,08643
8	6	June	-0,22046	-0,22046	-0,22046
9	7	July	-0,27772	0,722277	-0,13886
10	8	Aug	-0,40519	-1,40519	-1,06533
11	9	Sept	-0,28733	-0,28733	-0,47373
12	10	Oct	0,167582	0,167582	-0,01882
13	11	Nov	0,400684	0,400684	0,214284
14	12	Dec	0,056892	0,056892	-0,12951
15	13	Jan	0,37146	0,37146	0,185059
16	14	Feb	0,692372	0,692372	0,505972
17	15	March	0,875074	0,875074	0,688674
18	16	April	1,304348	1,304348	1,117947
19	17	May	1,414004	1,414004	1,227604
20	18	June	0,962493	0,962493	0,776093
21	19	July	0,727442	1,727442	0,270521
22	20	Aug	1,040654	0,040654	2,334931
23	21	Sept	1,395114	1,395114	1,344696
24	22	Oct	1,305113	1,305113	1,254695
25	23	Nov	1,705179	1,705179	1,654761
26	24	Dec	1,502029	1,502029	1,451611
27	25	Jan	1,694055	1,694055	1,643637
28	26	Feb	1,752033	1,752033	1,701615
29	27	March	1,819066	1,819066	1,768648
30	28	April	1,70069	1,70069	1,650272
31	29	May	1,909615	1,909615	1,859197
32	30	June	1,790179	1,790179	1,739761
33	31	July	2,19774	3,19774	1,073661
34	32	Aug	2,110298	1,110298	3,944878
35	33	Sept	2,445713	2,445713	2,140146
36	34	Oct	1,994083	1,994083	1,688517
37	35	Nov	1,610876	1,610876	1,30531
38	36	Dec	1,618388	1,618388	1,312822
39					

=(F3+C4)*VLOOKUP(B4;$L$2:$N$13;3;FALSE)

=E3+C4+VLOOKUP(B4;$H$4:$I$15;2;FALSE)

	M	N	O
1	month inc	additive	multiplicative
2	Jan	0	1
3	Feb	0	1
4	March	0	1
5	April	0	1
6	May	0	1
7	June	0	1
8	July	1	0,5
9	Aug	-2	4
10	Sept	1	0,5
11	Oct	0	1
12	Nov	0	1
13	Dec	0	1





File “0304 Time Series Examples….xls” on portal (press F9 to refresh

Methods for Time Series modeling

- Moving average
- Exponential smoothing
  - New Forecast (“level”) = α * Actual + $(1 - \alpha)$ * Old Forecast (“level”)
  - Holt’s model: Smoothing with [additive] trend: New Forecast = New Level + New Trend
    - New Level = α * Actual + $(1 - \alpha)$ * Old Forecast
    - New Trend = β * (New Level – Old Level) + $(1 - \beta)$ * Old Trend
  - Winter’s model: Smoothing with [additive] trend and seasonality
- Multiplicative smoothing methods
- Decompositions: TBATS (trigonometric Fourier transforms)
- Auto-regressive methods
  - ARMA, ARIMA, etc. (ARCH, GARCH, etc. for variance)
- Any of the above with regressors (covariates/features) – “dynamic regressions”

Timeseries modeling in R: your first example

INSEAD The Business School for the World®

Context: a small retailer decides whether to install solar panels on the roof of the store and needs to obtain a 30year forecast of electricity prices:

- Datafile "0304 electric rates data.csv" contains the data for monthly averages of prices over the last 10 years
- R script "0304 time series I - decompositions, ets and tbats.R" contains the code

Goal: predict monthly prices for the next 30 years (360 values). Analytical complications: which model(s) to use?

  - We will first look at the exponential smoothing ("`ets`") and trigonometric decompositions ("`tbats`")

- Coding complications:
  - Neither `ets` nor `tbats` are part of the standard R installation; will need to install a package and call a library

Load the data, define a timeseries

INSEAD
The Business School for the World®

- #install.packages("forecast") #-- do this only once
- #Check the book: https://www.otexts.org/fpp and the blog: http://robjhyndman.com/hyndsight/forecasting/
- library("forecast")
- ElectricPriceData<-read.csv(file.choose(), header=TRUE, sep=",")
- ElectricPrice_ts <- ts(ElectricPriceData$ElectricRate,start=2004, frequency=12) # ts function defines the dataset as timeseries starting Jan 2004 and having seasonality of frequency 12 (monthly)

Decomposition(s)

INSEAD

The Business School for the World®

#plot various decompositions into error/noise, trend and seasonality

fit <- decompose(ElectricPrice_ts, type="multiplicative") #decompose using "classical" method, multiplicativ form

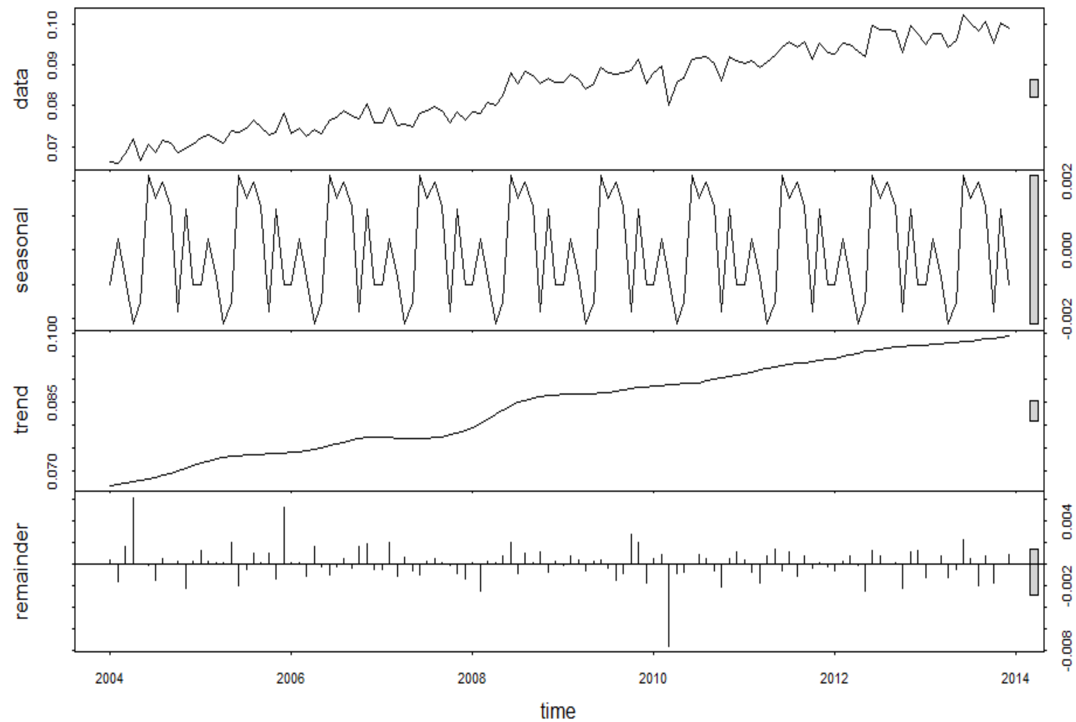
plot(fit)

fit <- decompose(ElectricPrice_ts, type="additive") #decompose using "classical" method, additive form

plot(fit)

fit <- stl(ElectricPrice_ts, t.window=12 s.window="periodic", robust=TRUE) #decompose using STL (Season and trend using Loess)

plot(fit)



Exponential smoothing: `ets` models

`ets` = "error-trend-seasonality"

error ={A,M}

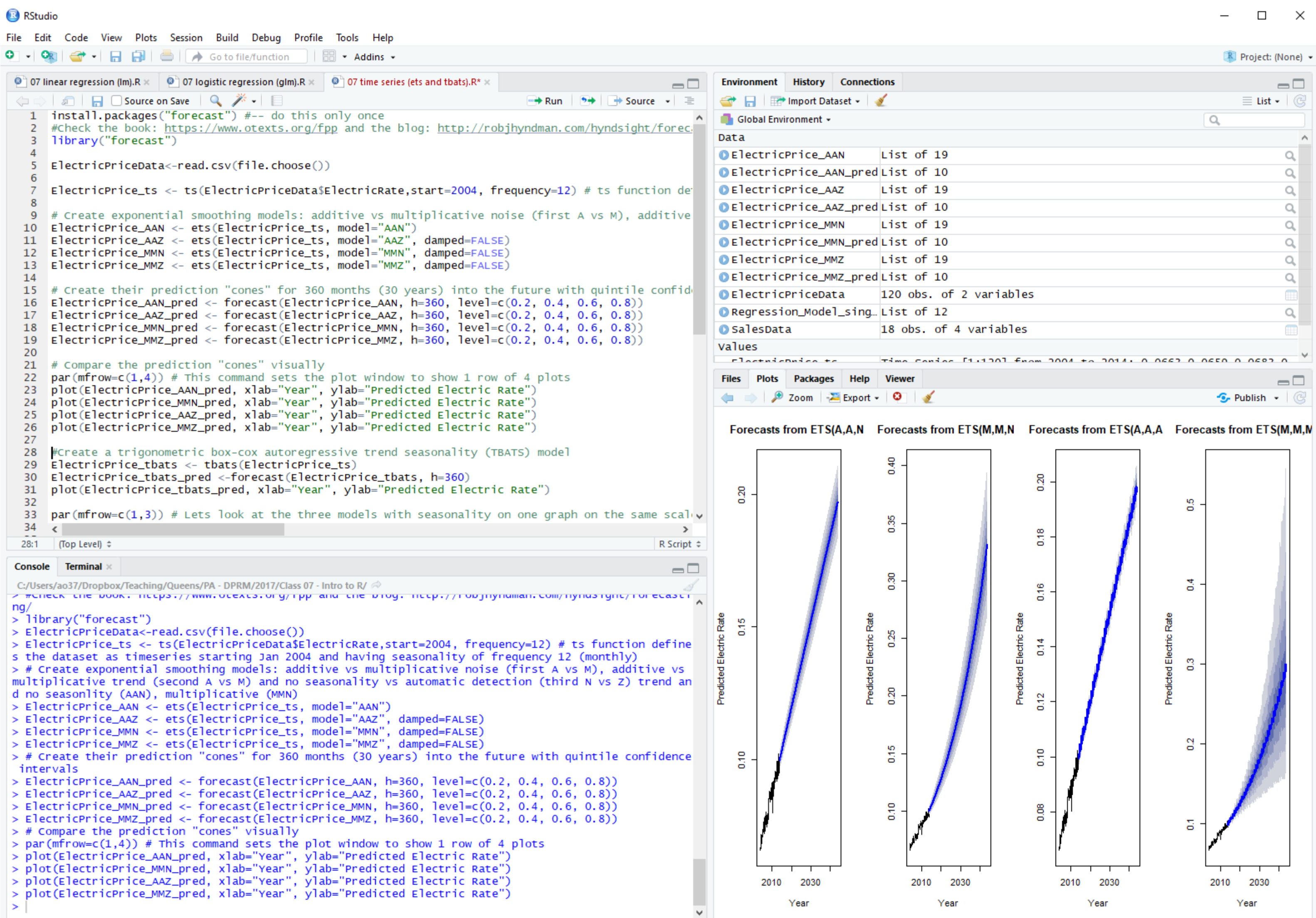
trend ={N,A,Ad,M,Md}

seasonality ={N,A,M}

		Seasonal Component	
Trend Component	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)
M (Multiplicative)	(M,N)	(M,A)	(M,M)
M_d (Multiplicative damped)	(M_d,N)	(M_d,A)	(M_d,M)

Super-powerful `ets` models in R are VERY easy

INSEAD
The Business School for the World®



File with code: “0304 time series I - decompositions, ets and tbats.R” on portal

`ets` models: taxonomy

INSEAD

The Business School for the World®

Trend	Seasonal N	Seasonal A	Seasonal M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t / \ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
A$_d$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$
M	$\hat{y}_{t+h\|t} = \ell_t b_t^h$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1}b_{t-1})) + (1-\gamma)s_{t-m}$
M$_d$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h}$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^{\phi}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t / (\ell_{t-1}b_{t-1}^{\phi})) + (1-\gamma)s_{t-m}$

`ets` models: reading output

> ElectricPrice_AAZ

ETS(A,A,A) What does AAA mean? **A**dditive noise, **A**dditive trend, **A**dditive seasonalit

Call: ets(y = ElectricPrice_ts, model = "AAZ", damped = FALSE)

Smoothing parameters:

alpha = 0.0583 What does a small alpha mean? Stable series, not much reaction to the noise

beta = 5e-04 If beta~0, does this mean no trend? No, beta=0 means a stable trend (little change)

gamma = 0.0311 What does small gamma mean? Stable seasonal indicies

Initial states:

l = 0.0678

b = 3e-04

s=-0.0013 0.0015 -0.0011 0.0012 0.0025 9e-04 0.0022 -0.0019 -0.0016 -0.0016 3e-04 -0.001

(N,N)	=	simple exponential smoothing
(A,N)	=	Holts linear method
(M,N)	=	Exponential trend method
(A_d,N)	=	additive damped trend method
(M_d,N)	=	multiplicative damped trend method
(A,A)	=	additive Holt-Winters method
(A,M)	=	multiplicative Holt-Winters method
(A_d,M)	=	Holt-Winters damped method

Method	Initial values
(N,N)	$\ell_0 = y_1$
(A,N) (A_d,N)	$\ell_0 = y_1, b_0 = y_2 - y_1$
(M,N) (M_d,N)	$\ell_0 = y_1, b_0 = y_2/y_1$
(A,A) (A_d,A)	$\ell_0 = \frac{1}{m}(y_1 + \cdots + y_m)$ $b_0 = \frac{1}{m}\left[\frac{y_{m+1}-y_1}{m} + \cdots + \frac{y_{m+m}-y_m}{m}\right]$ $s_0 = y_m - \ell_0,\ s_{-1} = y_{m-1} - \ell_0,\ \ldots,\ s_{-m+1} = y_1 - \ell_0$
(A,M) (A_d,M)	$\ell_0 = \frac{1}{m}(y_1 + \cdots + y_m)$ $b_0 = \frac{1}{m}\left[\frac{y_{m+1}-y_1}{m} + \cdots + \frac{y_{m+m}-y_m}{m}\right]$ $s_0 = y_m/\ell_0,\ s_{-1} = y_{m-1}/\ell_0,\ \ldots,\ s_{-m+1} = y_1/\ell_0$

Trigonometric seasonality: sin() and cos() "waves"

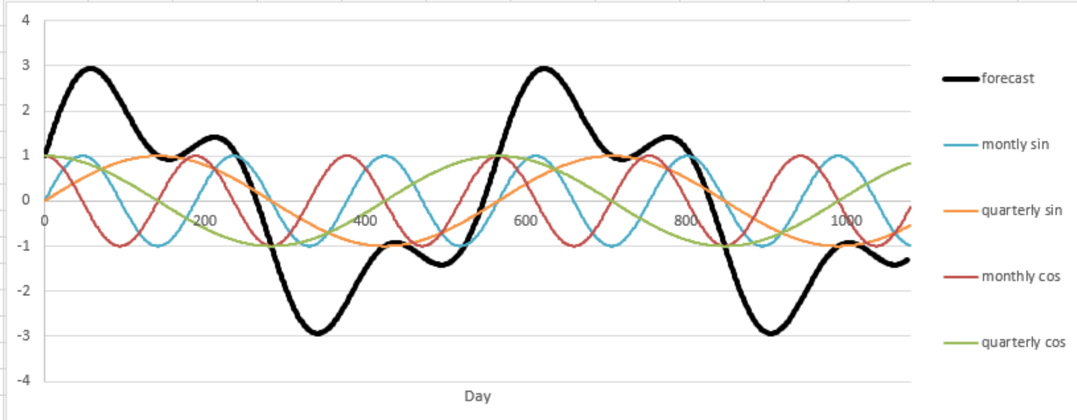
INSEAD
The Business School for the World®

=A4/30

=SIN(B4

=SUMPRODUCT($D$2:$G$2;D4:G4)

	A	B	C	D	E	F	G	H
1					weights			
2		Indicies		1	0	2	1	
3	Day	Monthly	Quarterly	montly sin	monthly cos	quarterly sin	quarterly cos	forecast
4	0	0	0	0	1	0	1	1
5	1	0,033333	0,011111	0,033327161	0,999444496	0,011110882	0,999938272	1,055487
6	2	0,066667	0,022222	0,066617295	0,997778601	0,022220393	0,999753097	1,110811
7	3	0,1	0,033333	0,099833417	0,995004165	0,033327161	0,999444496	1,165932
8	4	0,133333	0,044444	0,132938623	0,991124272	0,044429814	0,999012508	1,220811
9	5	0,166667	0,055556	0,165896133	0,986143232	0,055526982	0,998457187	1,275407
10	6	0,2	0,066667	0,198669331	0,980066578	0,066617295	0,997778601	1,329683
11	7	0,233333	0,077778	0,231221806	0,972901062	0,077699384	0,996976833	1,383597
12	8	0,266667	0,088889	0,263517391	0,964654645	0,08877188	0,996051983	1,437113
13	9	0,3	0,1	0,295520207	0,955336489	0,099833417	0,995004165	1,490191
14	10	0,333333	0,111111	0,327194697	0,944956946	0,110882629	0,993833509	1,542793
15	11	0,366667	0,122222	0,358505671	0,933527549	0,121918151	0,992540158	1,594882
16	12	0,4	0,133333	0,389418342	0,921060994	0,132938623	0,991124272	1,64642
17	13	0,433333	0,144444	0,419898367	0,907571133	0,143942682	0,989586027	1,69737
18	14	0,466667	0,155556	0,449911881	0,893072953	0,154928971	0,987925612	1,747695
19	15	0,5	0,166667	0,479425539	0,877582562	0,165896133	0,986143232	1,797361



- By adding more "waves" of different periodicities (monthly, quarterly, semi-annually, et.c), changing their weights, and adding "shifted waves" (month$_{t-1}$ wave, etc.) you can create <u>very elaborate</u> seasonal patterns – "TBATS" model

File "0304 Time Series Examples….xls" on portal

TBATS model

(trigonometric box-cox autoregressive trend seasonality)

INSEAD
The Business School for the World®

```
> ElectricPrice_tbats
TBATS(1, {0,0}, 1, {<12,5>})

Call: tbats(y = ElectricPrice_ts)

Parameters
  Alpha: 0.3119223
  Beta: 0.001059521
  Damping Parameter: 1
  Gamma-1 Values: -2.603098e-05
  Gamma-2 Values: 6.753875e-05

Seed States:
              [,1]
 [1,]  6.745943e-02
 [2,]  1.987776e-04
 [3,] -1.073112e-03
 [4,]  1.006799e-03
 [5,] -3.903458e-04
 [6,] -7.505053e-04
 [7,]  2.052910e-04
 [8,] -7.061269e-04
 [9,]  4.476049e-05
[10,]  1.308425e-04
[11,]  4.571218e-05
[12,]  9.594969e-04

Sigma: 0.001754992
AIC: -916.3708
>
>
>
>
>
>
>
>
>
>
```

Understanding TBATS output

"1":

Box-cox transformation (removing outliners)

1=nothing removed

"{0,0}":

Autoregressive moving average:

{p,q} = numbers of values in ARMA

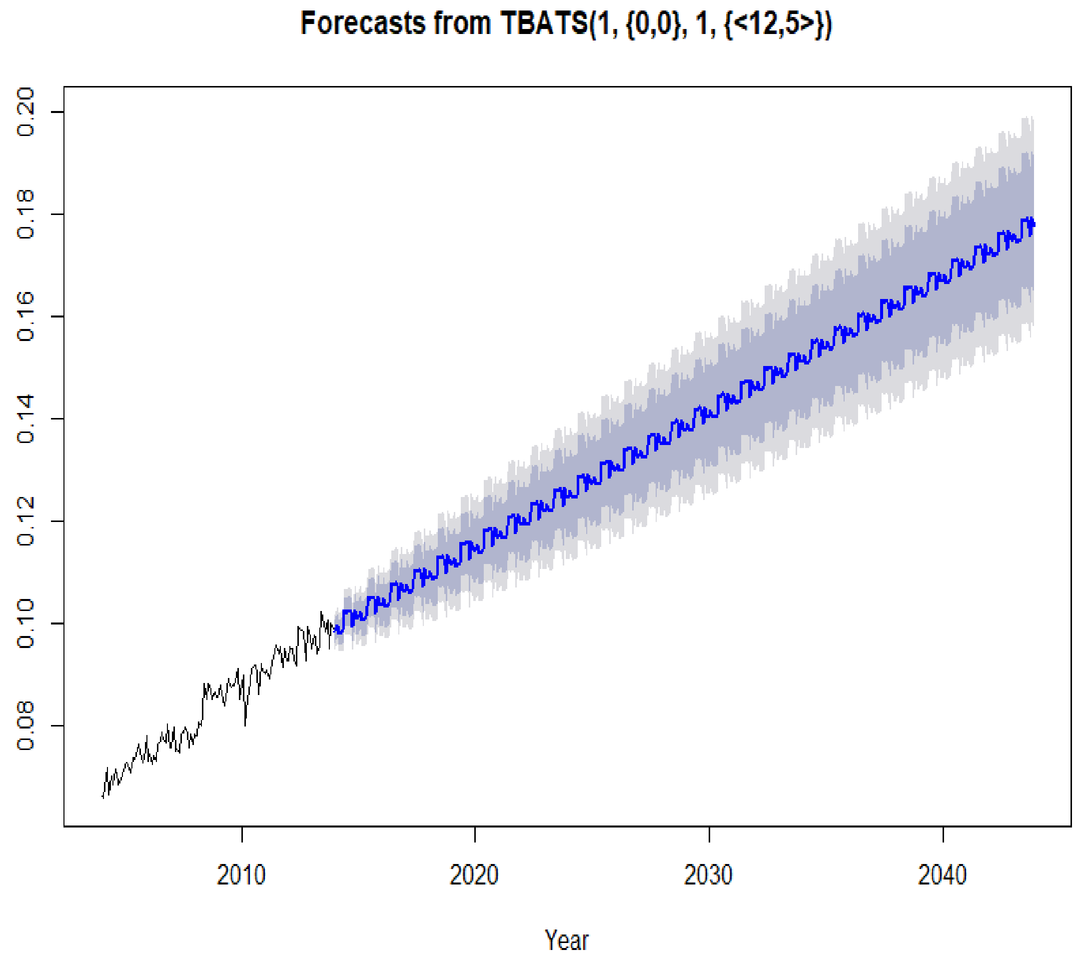
"1":

Dampening parameter (1=not dampened)

"(12,5)":

Number of seasonal periods,

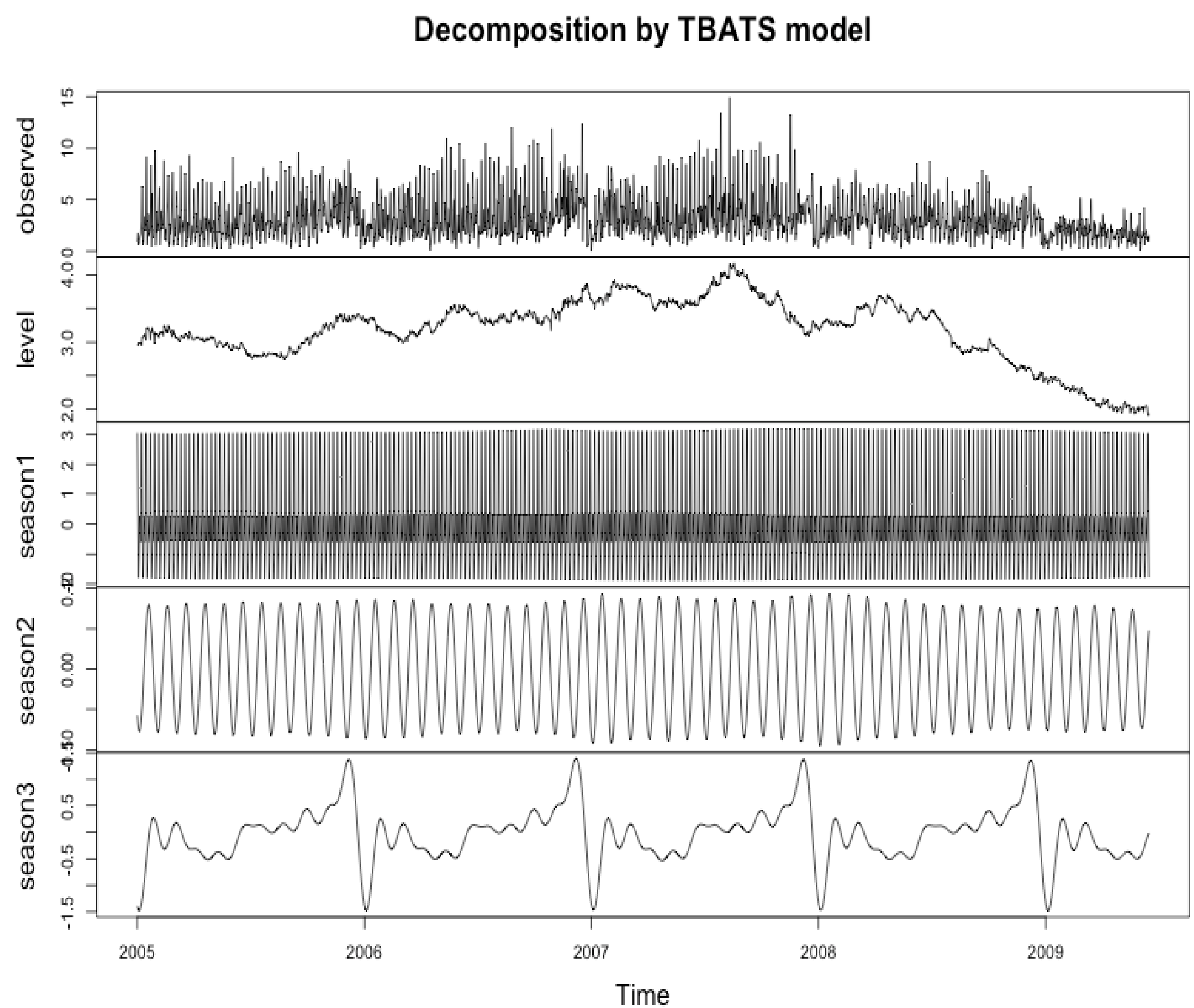
Number of trigonometric ("Fourier") terms for each seasonality



Multiple seasonalities: beyond this course ☹

INSEAD

The Business School for the World®



ARIMA: "differencing", moving average and auto-regression

INSEAD
The Business School for the World®

- "differencing" – creating new variables by taking differences between consecutive observations (y_t-y_{t-1})
- Auto-regression: $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t,$
  - y_t-=a+b*y_{t-1}+$error_t$
- Moving average: $y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q},$
  - y_t-=a+b*$error_{t-1}$+$error_t$
- ARIMA: "auto-regressive integrated moving average"
  - ARIMA(p,d,q): $y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t,$
    - p = num of auto-regressive terms
    - d = "order" of first difference
    - q = num of moving average terms

White noise	ARIMA(0,0,0)
Random walk	ARIMA(0,1,0) with no constant
Random walk with drift	ARIMA(0,1,0) with a constant
Autoregression	ARIMA(*p*,0,0)
Moving average	ARIMA(0,0,*q*)

- Auto-Correlation Function (ACF), ACF plot

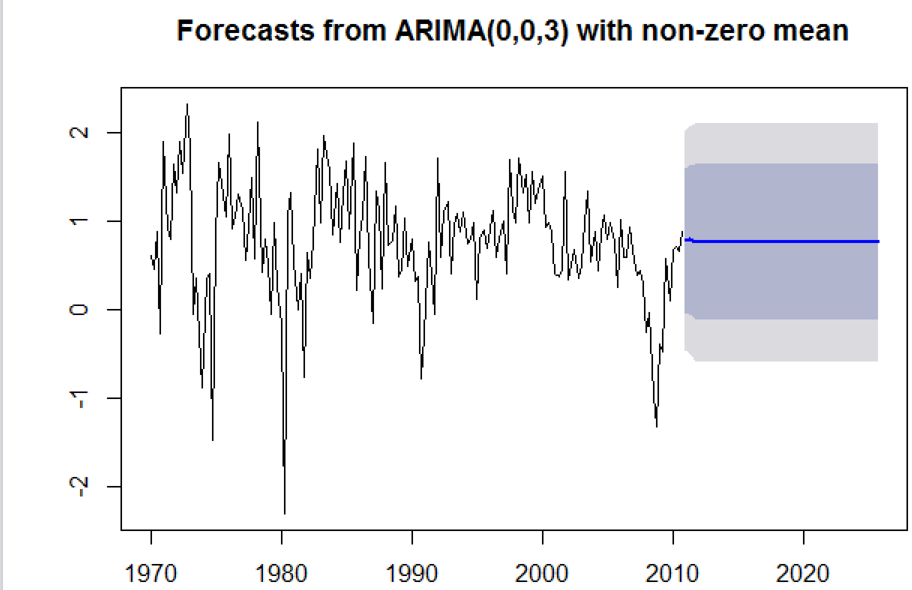
ARIMA example in R: US consumption data (from FPP)

INSEAD
The Business School for the World®

```
> fit <- auto.arima(usconsumption[,1],seasonal=FALSE) #automatically
 fits the ARIMA model (auto-regressive integrated moving average)
> fit
Series: usconsumption[, 1]
ARIMA(0,0,3) with non-zero mean

Coefficients:
         ma1     ma2     ma3    mean
      0.2542  0.2260  0.2695  0.7562
s.e.  0.0767  0.0779  0.0692  0.0844

sigma^2 estimated as 0.3953:  log likelihood=-154.73
AIC=319.46   AICc=319.84   BIC=334.96
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
```



File with code: "0304 time series II - ARIMAs and dynamic regressions" on portal

ARIMA and seasonality example: a10 dataset (from FPP)

- $\text{ARIMA}\underbrace{(p,d,q)}_{\text{non-seasonal part}}\ \underbrace{(P,D,Q)}_{\text{seasonal part}}{}_{m}$ ← number of periods in a per season

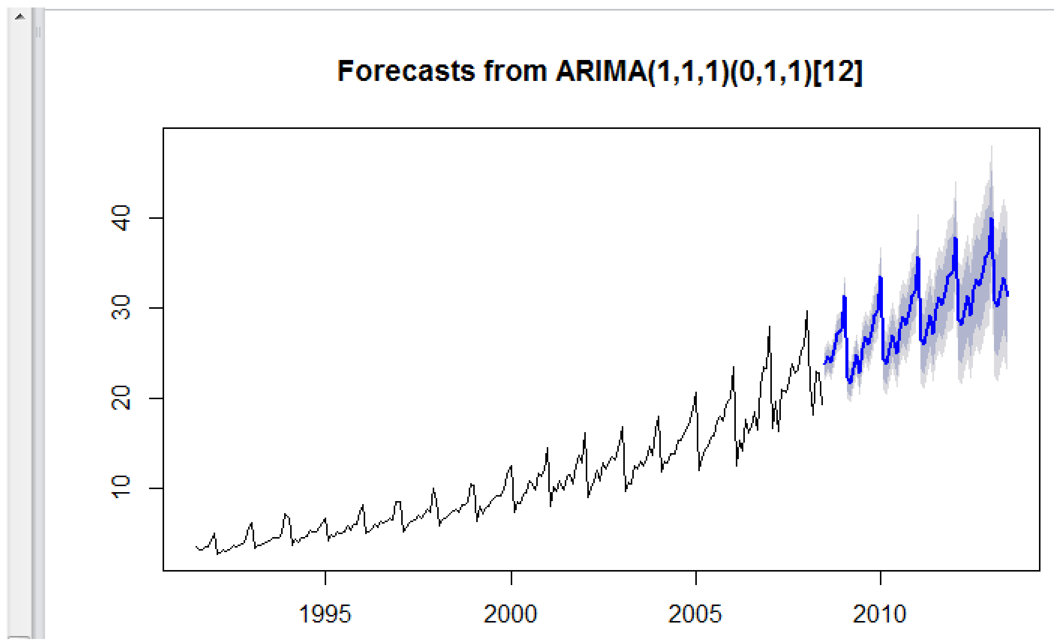
```
>
> fit <- auto.arima(a10,seasonal=FALSE)
> fit
Series: a10
ARIMA(1,1,1) with drift

Coefficients:
        ar1      ma1   drift
      0.314  -0.9164  0.0965
s.e.  0.075   0.0262  0.0171

sigma^2 estimated as 3.626:  log likelihood=-417.92
AIC=843.84   AICc=844.04   BIC=857.09
> fit <- auto.arima(a10,seasonal=TRUE)
> fit
Series: a10
ARIMA(1,1,1)(0,1,1)[12]

Coefficients:
          ar1      ma1     sma1
      -0.2504  -0.6674  -0.4725
s.e.   0.1007   0.0870   0.0641

sigma^2 estimated as 0.8756:  log likelihood=-258.82
AIC=525.63   AICc=525.85   BIC=538.64
> plot(forecast(fit,60)) #60 stands for 60 months = 5 years
>
>
```



File with code: “0304 time series II - ARIMAs and dynamic regressions” on portal

Dynamic Regression: timeseries with regressors/covariates

INSEAD
The Business School for the World®

- In classes 0102 (Sarah's case) we considered predicting Y variable (price) by knowing some other variables (weight, color, etc.).
  - Those variables are sometimes called covariates, or features
- Today (so far) we considered predicting Y variable (eclectic price, consumption, etc.) by knowing how Y itself evolved in the past – time series analyses
- Can the two approaches be combined?
  - Of course! This has two names:
    - "dynamic regression" or
    - "time series with regressors/covariates/features"

Dynamic Regression example: insurance quotes and advertising

The Business School for the World®

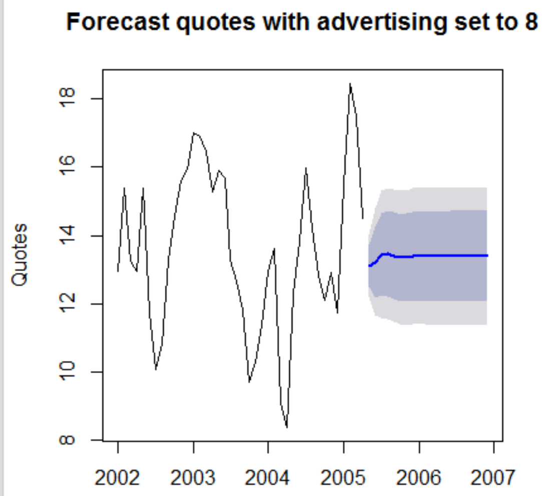
- **Context**: we have data on the number of new insurance quotes and advertising in the previous period. Naturally expect two kinds of dependencies:
  - "direct effect": more advertising today = more new quotes today
  - "carryover effects":
    - more advertising yesterday = more new quotes today
    - more new quotes yesterday = more new quotes today

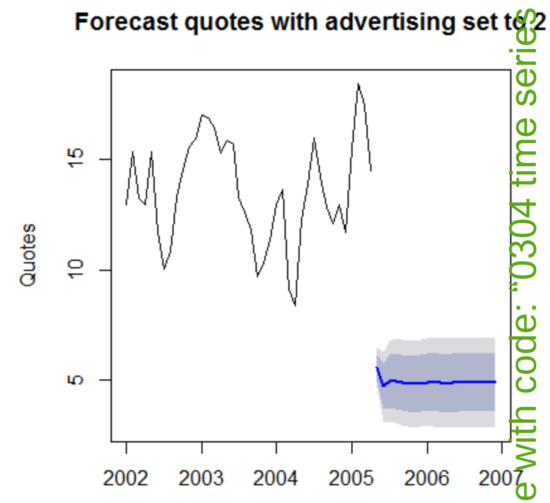
```
>
> Advert <- cbind(insurance[,2],
+                 c(NA,insurance[1:39,2]),
+                 c(NA,NA,insurance[1:38,2]),
+                 c(NA,NA,NA,insurance[1:37,2]))
> colnames(Advert) <- paste("AdLag",0:3,sep="")
>
>
> fit <- auto.arima(insurance[,1], xreg=Advert[,1:2], d=0)
> fit
Series: insurance[, 1]
Regression with ARIMA(3,0,0) errors

Coefficients:
         ar1      ar2     ar3  intercept  AdLag0  AdLag1
      1.4117  -0.9317  0.3591     2.0393  1.2564  0.1625
s.e.  0.1698   0.2545  0.1592     0.9931  0.0667  0.0591

sigma^2 estimated as 0.2165:  log likelihood=-23.89
AIC=61.78   AICc=65.28   BIC=73.6
>
>
> par(mfrow=c(1,2))
> fc8 <- forecast(fit, xreg=cbind(rep(8,20),c(Advert[40,1],rep(8,19))), h=20)
> plot(fc8, main="Forecast quotes with advertising set to 8", ylab="Quotes")
> fc2 <- forecast(fit, xreg=cbind(rep(2,20),c(Advert[40,1],rep(2,19))), h=20)
> plot(fc2, main="Forecast quotes with advertising set to 2", ylab="Quotes")
```

custom (written "by you") R function to create variables with lags 2,3,4





File with code: "0304 time series II - ARIMAs and dynamic regressions" on portal

Summary of Sessions 3-4

INSEAD
The Business School for the World®

- On many occasions data are indexed by time – timeseries data
- Such data requires special analytical tools, which explicitly account for the fact that prediction errors increase over time
- We discussed concepts and implementations of four families of models:
  - Exponential smoothing (`ets`)
  - Trigonometric decompositions (`tbats`)
  - Auto-regressive moving averages (ARIMA)
  - Dynamic regressions (we saw an example based on ARIMA, but the concept applies to any method)
- As with regression: through coding and R, very powerful time series analytics can be implemented in minutes
- Many resources online: e.g., FPP book https://www.otexts.org/fpp and blog: http://robjhyndman.com/hyndsight/forecasting/

Next…

INSEAD
The Business School for the World®

- Group Assignment 1: Yahoo's acquisition of Tumblr
  - Due 8:25am of AB Sessions 5-6 morning (next Tue, Jan 30), upload to INSEAD portal

- Module II of the course: predicting events / "classification"
  - Sessions 5-6: metrics for classification and two main methods: logistic regression and CART
  - Tutorial 2 (mid-term R help, specifically on classifiation)
  - Assignment 2 (predicting credit defaults)
  - Session 7: credit defaults case discussion + other methods: random forest, xgboost, regularizations (LASSO/Ridge)

INSEAD

The Business School for the World®

Europe | Asia | Middle East