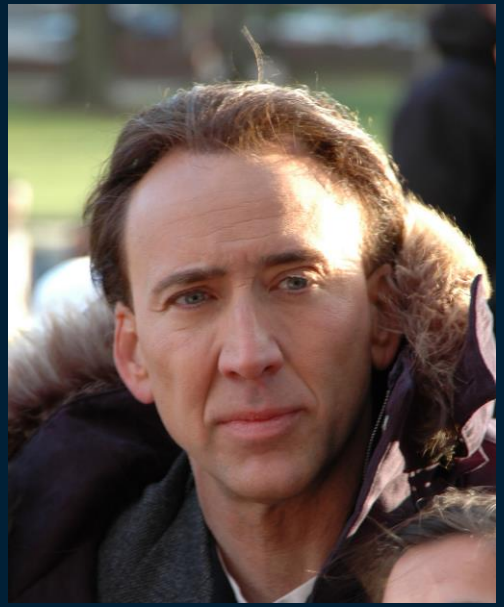
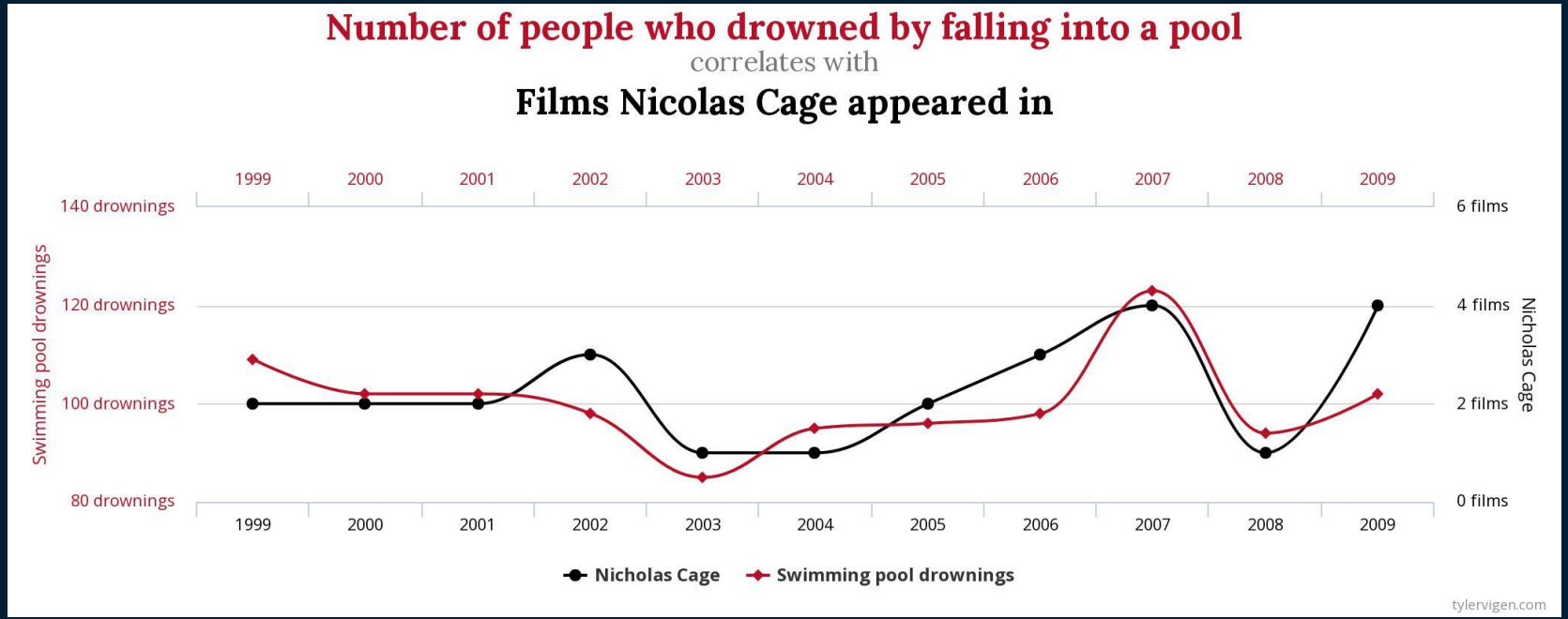
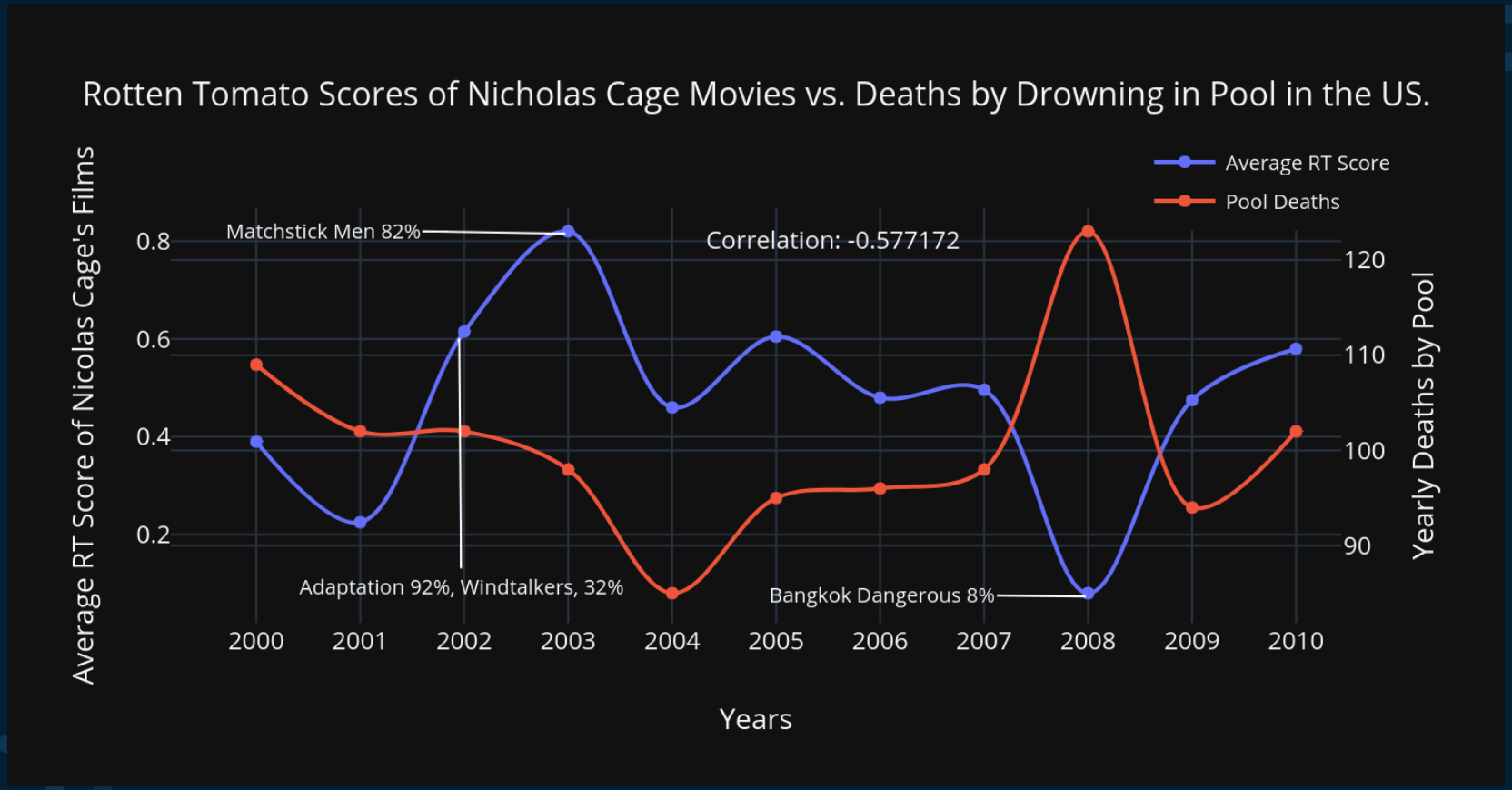
Timeseries Analysis – From ABT to modelling

by Rune Hjort Nielsen, SAS & Pasi Helenius, SAS





http://tylervigen.com/spurious-correlations



https://towardsdatascience.com/nicholas-cage-pool-saviour-9c13feafff6f

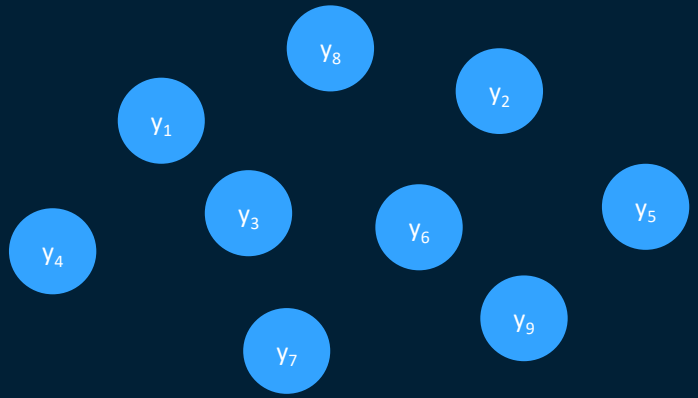
Agenda

Rune Hjorth Nielsen, PhD

Data scientist & AI specialist at SAS Institute

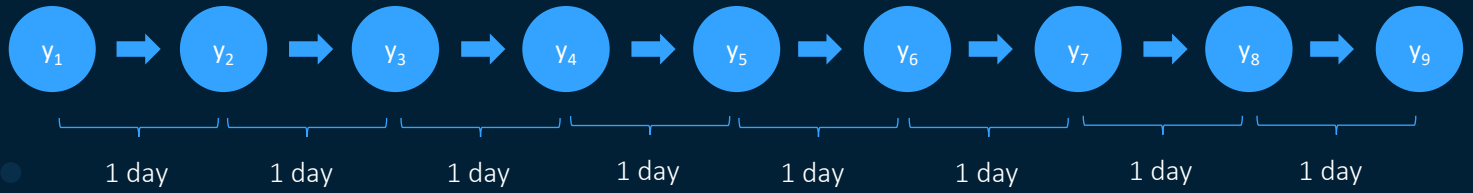
- Time series data
- Unit root warning
- A more flexible approach

Time series data



Cross sectional

Time series



WORK.TRANSACTIONS | Columns: 4 of 4 | Total rows: 300 | Rows 1 to 200

Enter expression

	customer	date	withdrawals	deposits
84	Bill	19NOV07	$461.68	$0.00
85	Bill	19NOV07	$787.63	$0.00
86	Bill	19NOV07	$204.00	$0.00
87	Bill	19NOV07	$890.44	$0.00
88	Bill	20NOV07	$758.27	$0.00
89	Bill	20NOV07	$765.89	$0.00
90	Bill	20NOV07	$968.74	$0.00

```
proc timeseries data=transactions
                out=timeseries;
   by customer;
   id date interval=day accumulate=total;
   var withdrawals deposits;
run;
```

101	Carlos	05NOV07	$938.88	$4,694.41
102	Carlos	05NOV07	$823.85	$0.00
103	Carlos	05NOV07	$588.99	$0.00
104	Carlos	05NOV07	$79.23	$0.00
105	Carlos	05NOV07	$533.07	$0.00
106	Carlos	05NOV07	$284.50	$0.00
107	Carlos	05NOV07	$346.30	$0.00
108	Carlos	05NOV07	$44.00	$0.00
109	Carlos	06NOV07	$109.51	$0.00
110	Carlos	06NOV07	$675.75	$0.00
111	Carlos	06NOV07	$812.59	$2,711.29

WORK.TIMESERIES | Columns: 4 of 4 | Total rows: 51 | Rows 1 to 51

Enter expression

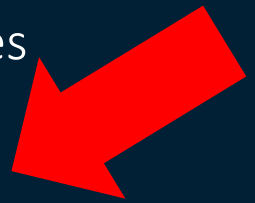
	customer	date	withdrawals	deposits
1	Bill	05NOV2007	$2,230.72	$2,537.91
2	Bill	06NOV2007	$3,080.24	$4,369.05
3	Bill	07NOV2007	$3,061.54	$1,471.90
4	Bill	08NOV2007	$3,548.72	$169,691.41
5	Bill	09NOV2007	$2,908.35	$0.00
6	Bill	10NOV2007	$2,281.90	$0.00
7	Bill	11NOV2007	$4,162.34	$173,012.45
8	Bill	12NOV2007	$2,211.12	$40,922.81
9	Bill	13NOV2007	$2,287.67	$0.00
10	Bill	14NOV2007	$3,232.42	$127,847.00
11	Bill	15NOV2007	$2,636.91	$195,691.39
12	Bill	16NOV2007	$3,854.54	$0.00
13	Bill	17NOV2007	$5,202.28	$42,951.14
14	Bill	18NOV2007	$1,600.47	$27,664.44
15	Bill	19NOV2007	$2,637.30	$0.00
16	Bill	20NOV2007	$3,682.39	$32,960.10
17	Bill	21NOV2007	$2,626.56	$9,547.37
18	Bill	22NOV2007	$1,384.46	$0.00
19	Carlos	05NOV2007	$3,638.82	$4,694.41
20	Carlos	06NOV2007	$3,632.56	$112,855.03
21	Carlos	07NOV2007	$2,159.47	$103,467.96
22	Carlos	08NOV2007	$1,165.49	$123,615.00
23	Carlos	09NOV2007	$2,654.67	$0.00
24	Carlos	10NOV2007	$1,400.61	$62,766.47
25	Carlos	11NOV2007	$3,689.09	$0.00
26	Carlos	12NOV2007	$2,759.42	$75,073.20
27	Carlos	13NOV2007	$3,654.53	$144,668.59
28	Carlos	14NOV2007	$3,811.02	$85,707.47

Why do we need specific time series tools?

Ordinary least squares (OLS)

OLS assumptions

1. Random sample
2. Linear relationship
3. No perfectly correlated independent variables
4. Conditional mean is zero
5. Homoscedasticity and no autocorrelation
6. Normally distributed errors



Time series analysis

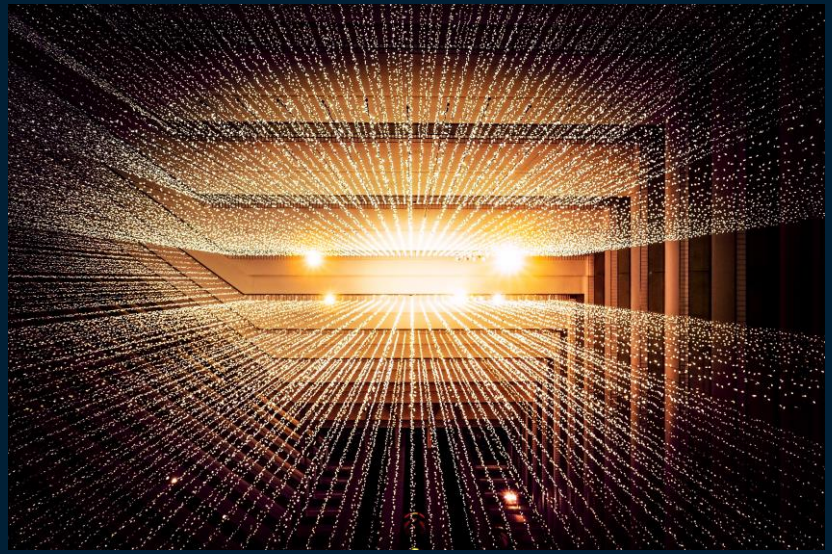
Choice 1

Classical time series analysis and econometrics

- Follows theory closely
- Needs focus on spurious conclusions
- Have solutions to specific complex problems

Time series analysis

Choice 2



Model selection by forecasting abilities

- Follows the data
- Needs stringent focus on data understanding and data ethics
- Can handle very complex data structures

Unit root warning

Spurious regression and unit roots

Autoregression of order 1, AR(1):

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

With unit root $a_1 = 1$:

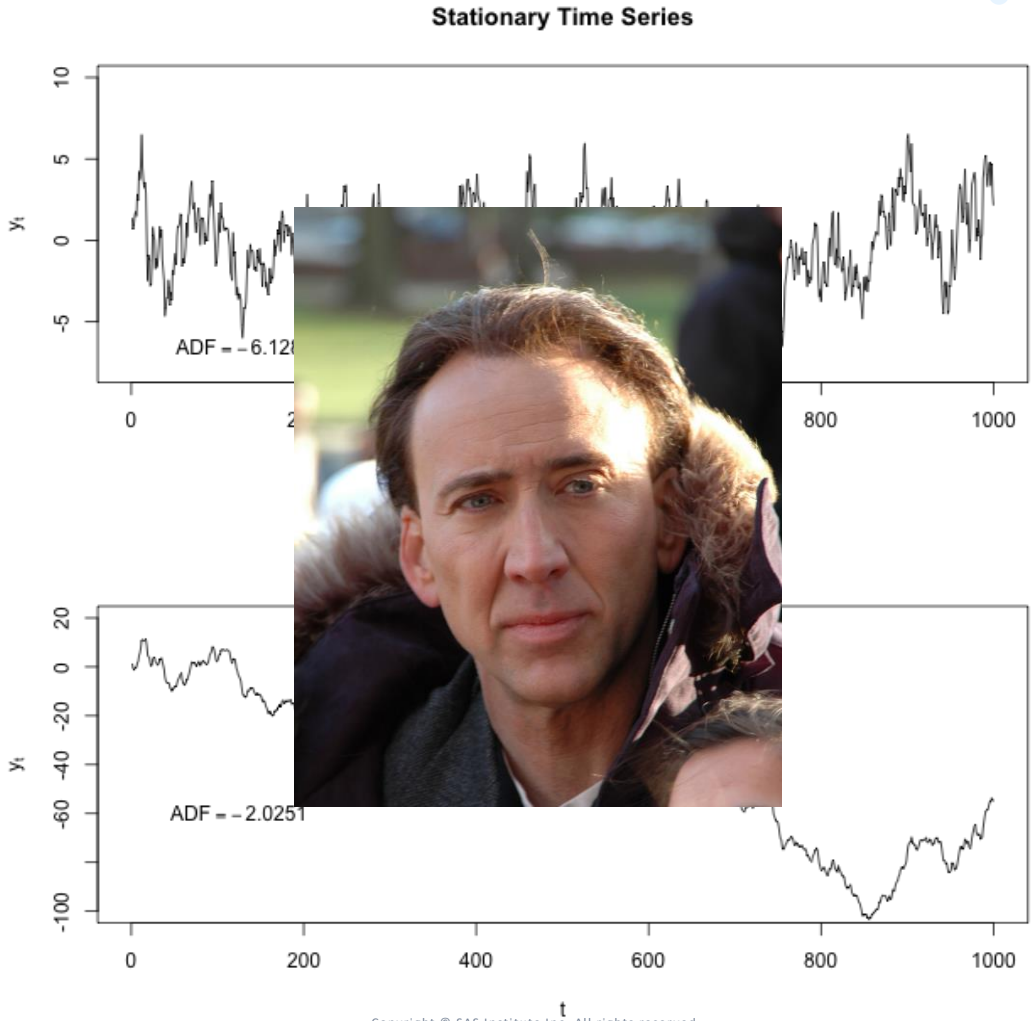
$$y_t = y_{t-1} + \varepsilon_t$$

With repeated substitution:

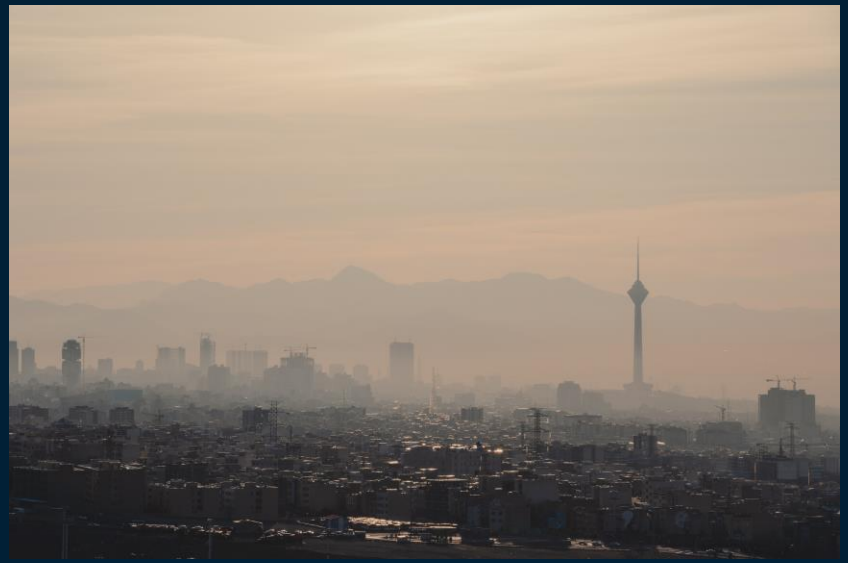
$$y_t = y_0 + \sum_{j=1}^{t} \varepsilon_j$$

The variance of y_t:

$$Var(y_t) = \sum_{j=1}^{t} \sigma^2 = t\sigma^2$$

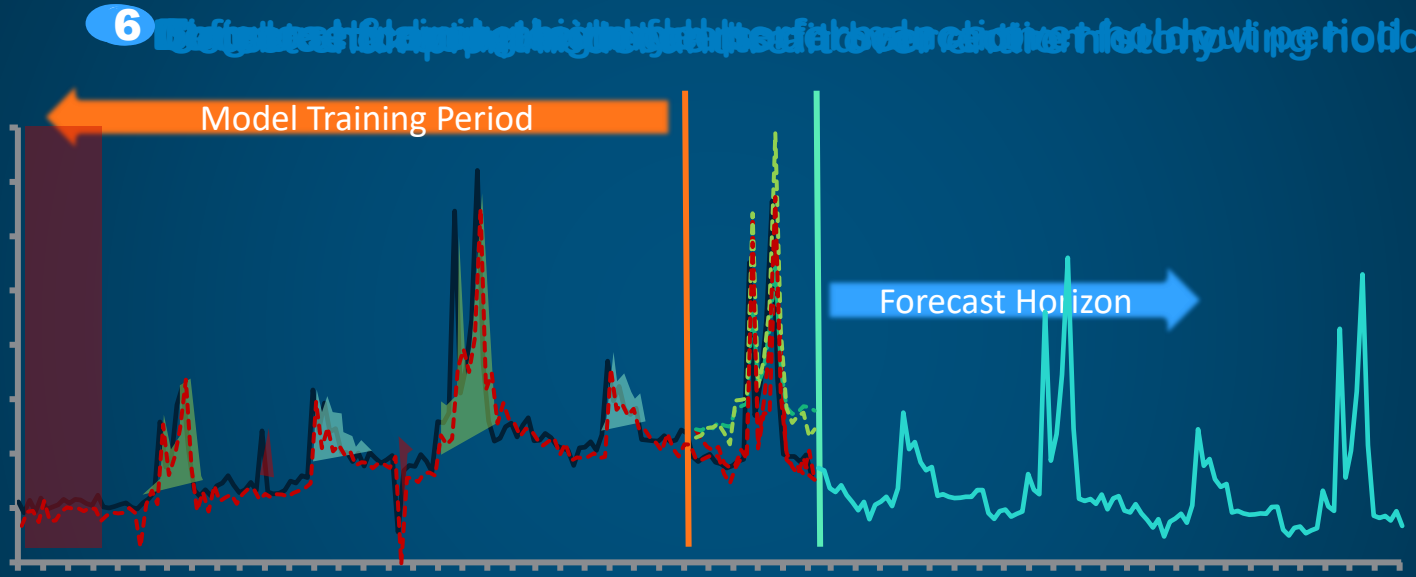


www.en.wikipedia.org/wiki/Stationary_process




Dat has a better idea

Automated Large Scale Time Series Forecasting



Error	
13.64%	Model1: *ARIMA*, f(History, Outliers, Price, Promotions, Inventory, Christmas, Black Friday, Catalog)
11.05%	Model2: *Exponential Smoothing*, f(History, Seasonality)
8.58%	Model3: *UCM*, f(History, Outliers, Price, Christmas, Catalog)

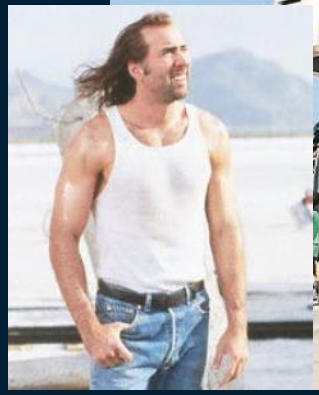
Trend
Seasonality
Outliers

Rune Hjorth Nielsen
Providing insights within data science and AI
for SAS customer advisory

SAS Visual Forecasting

Pasi Helenius

SAS Visual Forecasting



SAS Forecast Server