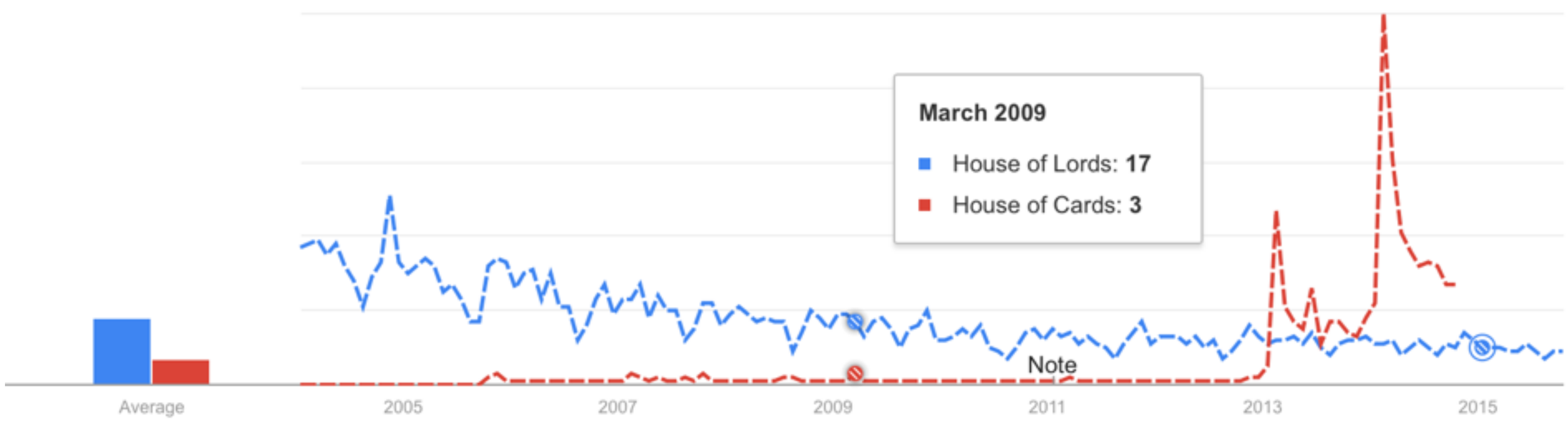
Time-Series Analysis

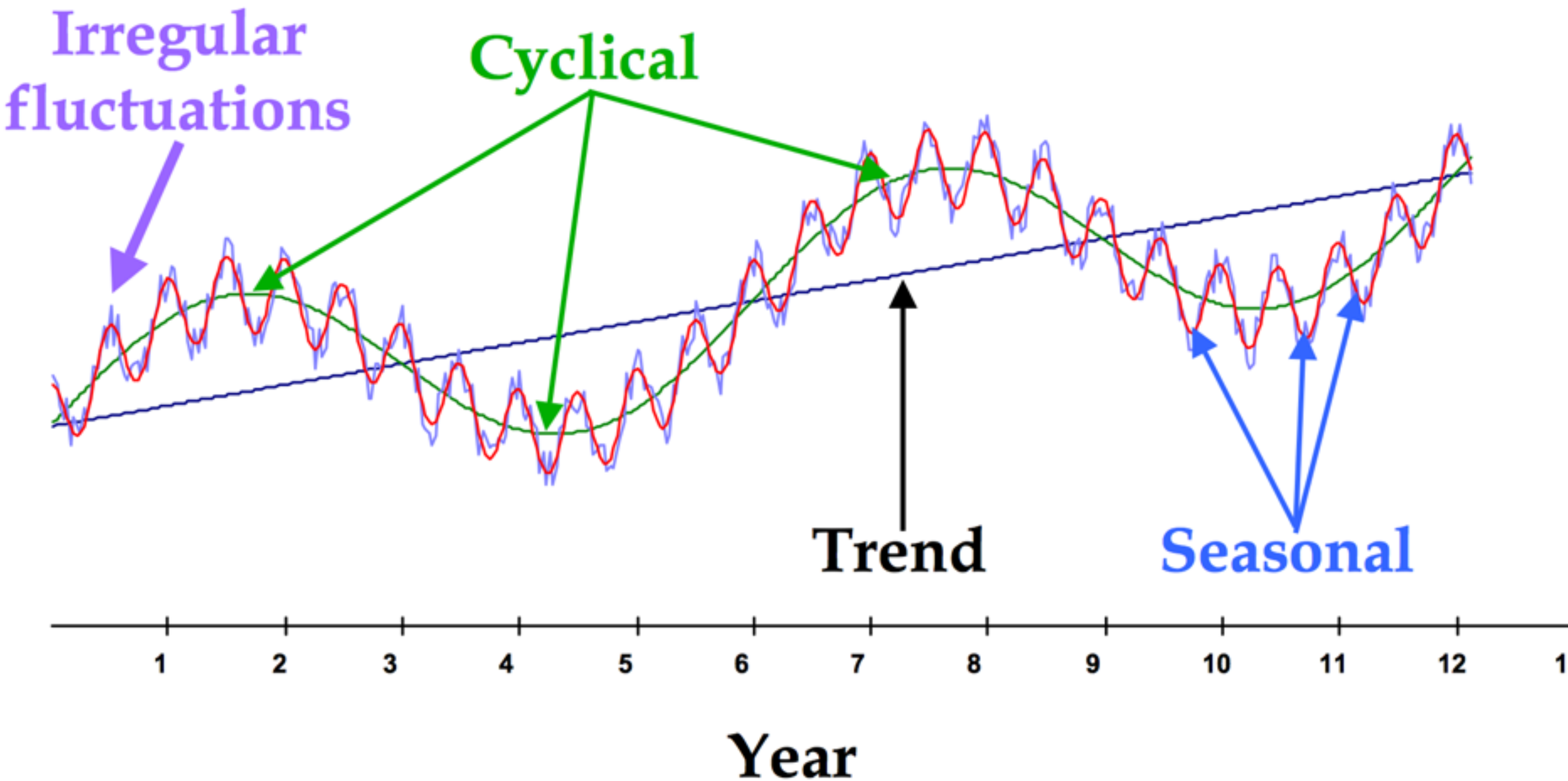
Forecasting, Periodicity-detection

Time Series

- An ordered sequence of values (data points) of variables at equally spaced time intervals



Time Series Components

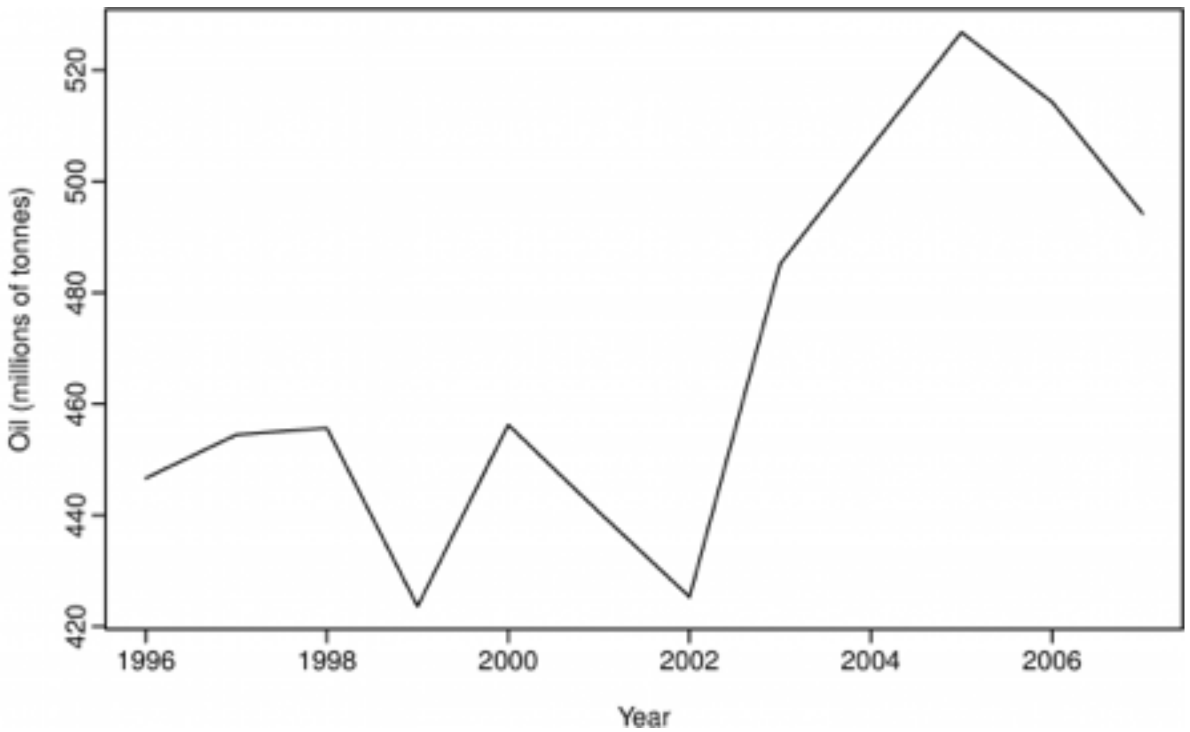


Forecasting

- Forecasting using time-series a classical task to predict the future values based on past observations
- Heavily used in many domains for forecasting:
  - Stock prices
  - Market adoption
  - Weather conditions
  - Sales of articles
  - Environmental carbon footprint

Forecasting

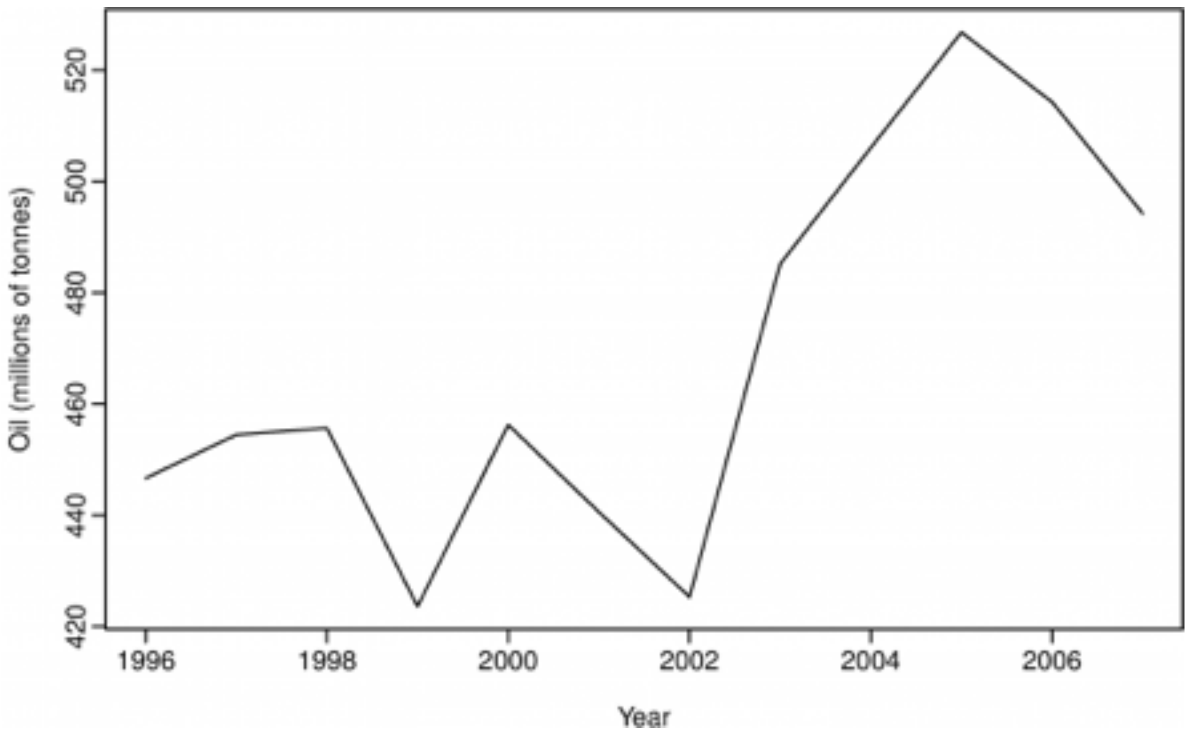
- Given a series of observations $\{y_T\}$ predict forecast the observation at T+h



Given that observations
have been made until time t

Forecasting

- Given a series of observations $\{y_T\}$ predict forecast the observation at T+h

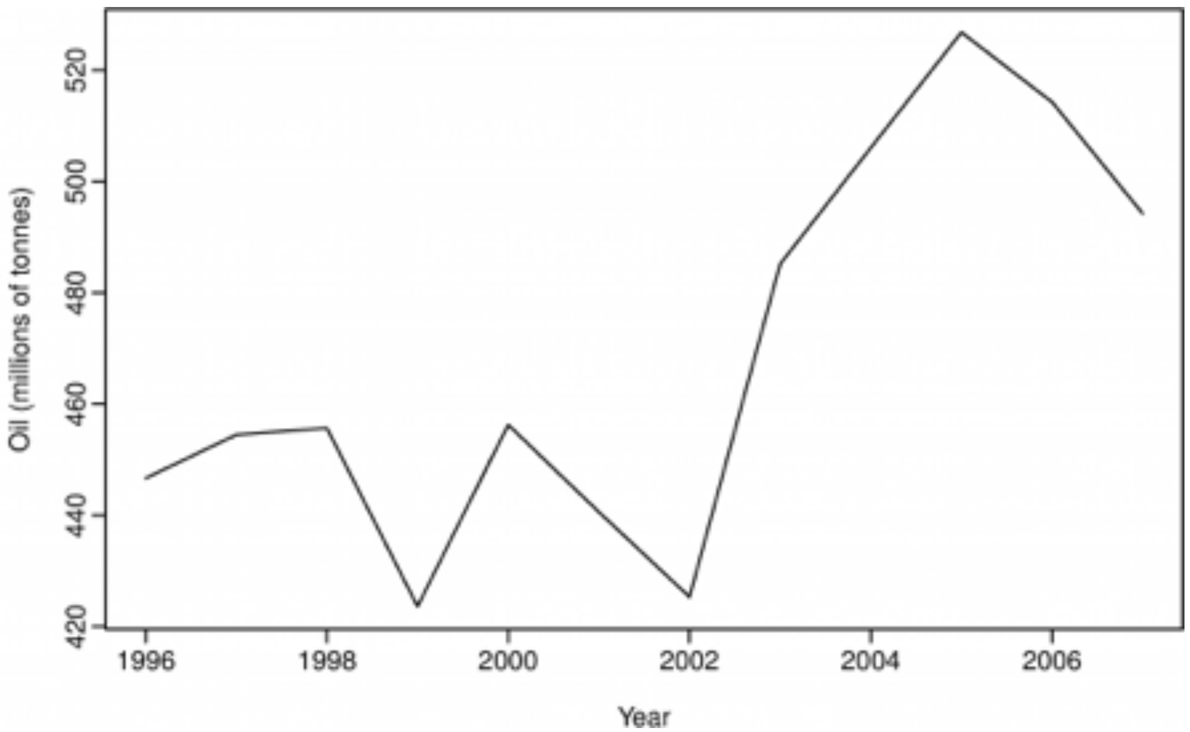


$$\hat{y}_{t+h|t} = y_t$$

Given that observations have been made until time t

Forecasting

- Given a series of observations $\{y_T\}$ predict forecast the observation at T+h



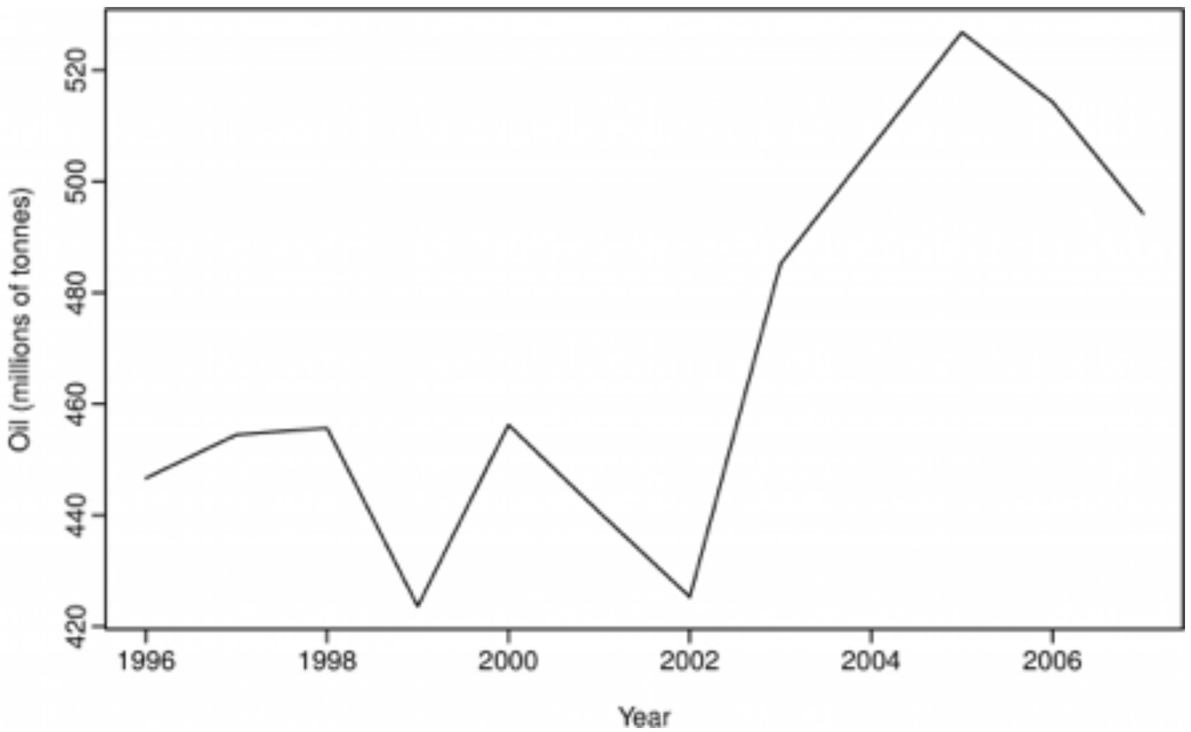
Given that observations have been made until time t

forecast for time t+h

$$\hat{y}_{t+h|t} = y_t$$

Forecasting

- Given a series of observations $\{y_T\}$ predict forecast the observation at T+h



forecast for time t+h

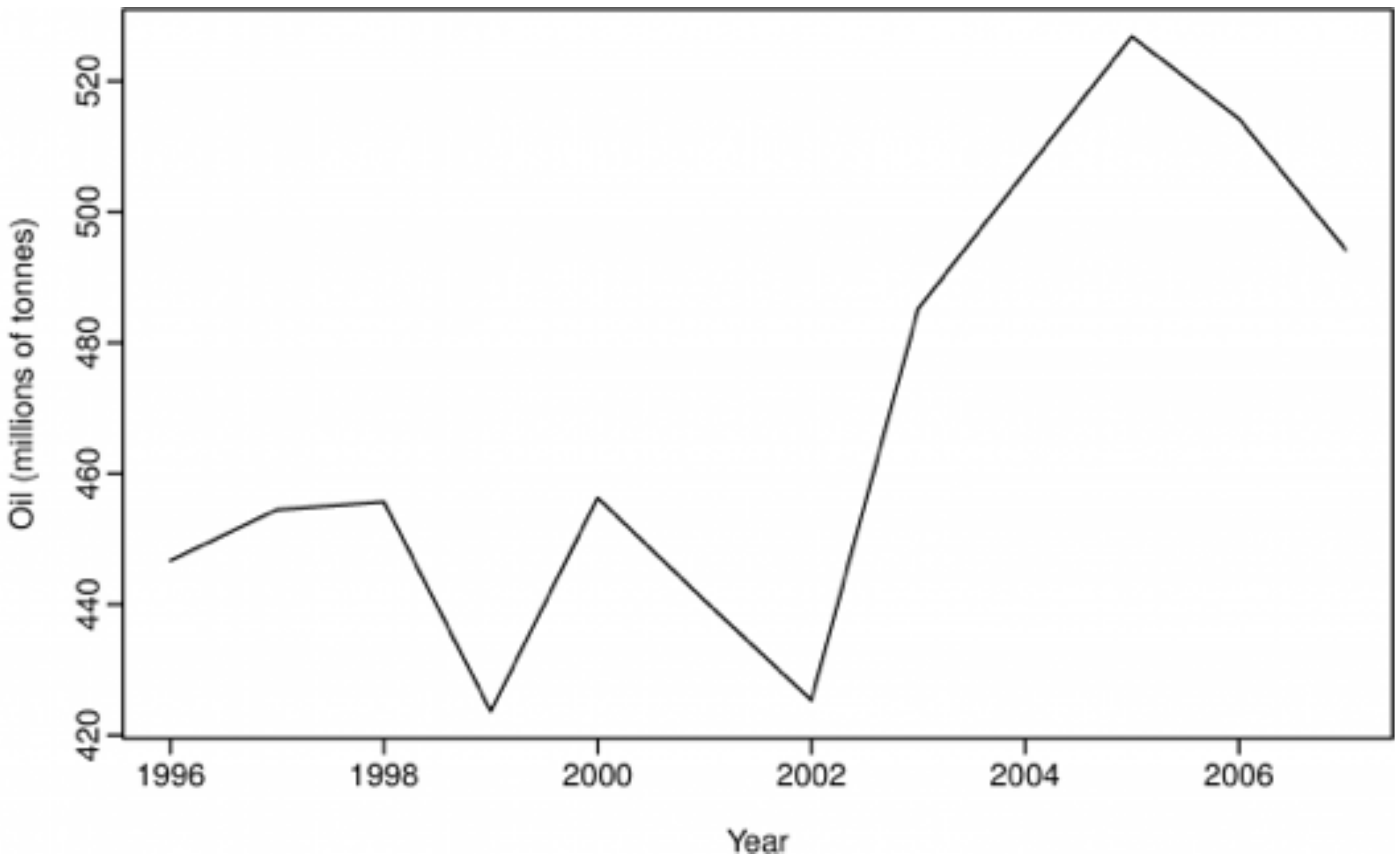
$$\hat{y}_{t+h|t} = y_t$$

observed value at t

Given that observations have been made until time t

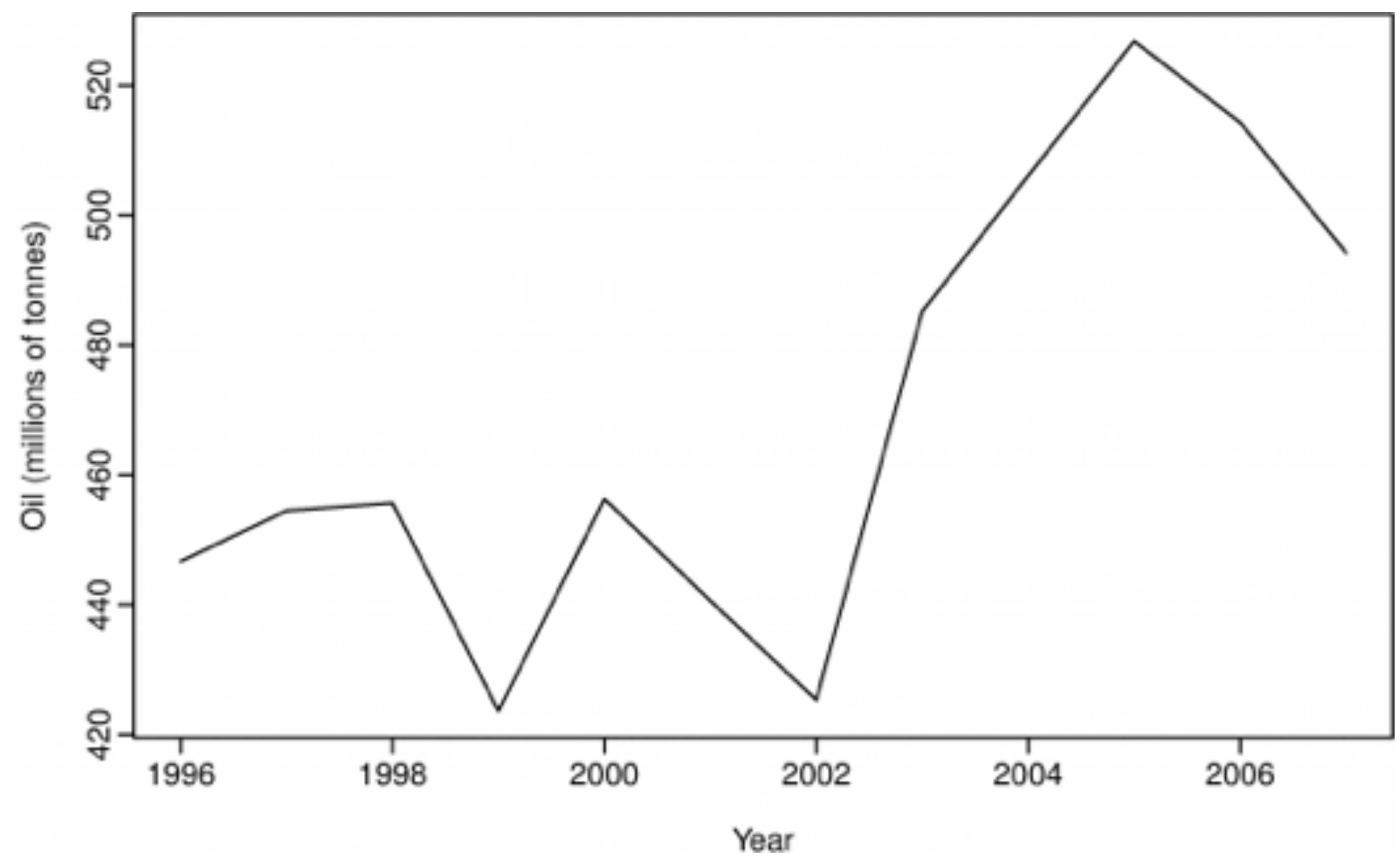
Forecasting

- What is the forecast for the next time point ?



Forecasting

- What is the forecast for the next time point ?
  - Simple Average
  - Moving Average
  - Linear Regression
  - Weighted Average



Idea : More weights to recent data

Simple Exponential Smoothing

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots \, ,$$

- Forecast for time t+1 given that t values have been observed
- Exponentially decreasing weights

Simple Exponential Smoothing

smoothing
parameter in [0,1]

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots ,$$

- Forecast for time t+1 given that t values have been observed
- Exponentially decreasing weights

Simple Exponential Smoothing

smoothing
parameter in [0,1]

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots ,$$

- Forecast for time t+1 given that t values have been observed
- Exponentially decreasing weights

Observation	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
y_T	0.2	0.4	0.6	0.8
y_{T-1}	0.16	0.24	0.24	0.16
y_{T-2}	0.128	0.144	0.096	0.032
y_{T-3}	0.1024	0.0864	0.0384	0.0064
y_{T-4}	$(0.2)(0.8)^4$	$(0.4)(0.6)^4$	$(0.6)(0.4)^4$	$(0.8)(0.2)^4$
y_{T-5}	$(0.2)(0.8)^5$	$(0.4)(0.6)^5$	$(0.6)(0.4)^5$	$(0.8)(0.2)^5$

SES: Component Form

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots ,$$

weighted average form

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

Component Form

Interpretation: SES is the linear combination of last observed value and last forecast value

SES: Component Form

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots,$$

weighted average form

forecast for time t+1

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

Component Form

Interpretation: SES is the linear combination of last observed value and last forecast value

SES: Component Form

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots ,$$

weighted average form

forecast for time t+1

forecast for time t

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

Component Form

Interpretation: SES is the linear combination of last observed value and last forecast value

Simple Exponential Smoothing

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$$ component form

$$\ell_{t-1} = \hat{y}_{t|t-1}$$

Simple Exponential Smoothing

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$$ component form

$$\ell_{t-1} = \hat{y}_{t|t-1}$$

is the level (or the smoothed value) of the series at time t

Simple Exponential Smoothing

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$$ component form

$$\ell_{t-1} = \hat{y}_{t|t-1}$$

is the level (or the smoothed value) of the series at time t

$$\begin{aligned}\ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t\end{aligned}$$ error correction form

Simple Exponential Smoothing

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$$ component form

$$\ell_{t-1} = \hat{y}_{t|t-1}$$

is the level (or the smoothed value) of the series at time t

$$\begin{aligned}\ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t\end{aligned}$$ error correction form

Interpretation: SES is sum of last observed value and smoothed error from the last measurement

Simple Exponential Smoothing

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots \,,$$

Weighted Average form

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$$

component form

$$\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1})$$

error correction form

SES gives us the forecast for level

Simple Exponential Smoothing

in action

Query vol. for “hannover”

1	23	NA	23
2	40	23	
3	25	26.4	
4	27	26.12	
5	32	26.296	
6	48	27.437	
7	33	31.549	
8	37	31.840	
9	37	32.872	
10	50	33.697	

since no prior info. exists

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$$ component form

Simple Exponential Smoothing

in action

Query vol. for "hannover"

1	23	NA	23
2	40	23	(.2)(40)+(.8)(23)=26.4
3	25	26.4	(.2)(25)+(.8)(26.4)=26.12
4	27	26.12	(.2)(27)+(.8)(26.12)=26.296
5	32	26.296	(.2)(32)+(.8)(26.296)=27.437
6	48	27.437	(.2)(48)+(.8)(27.437)=31.549
7	33	31.549	(.2)(48)+(.8)(31.549)=31.840
8	37	31.840	(.2)(33)+(.8)(31.840)=32.872
9	37	32.872	(.2)(37)+(.8)(32.872)=33.697
10	50	33.697	(.2)(50)+(.8)(33.697)=36.958

since no prior info. exists

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$$ component form

Choice of Weights

- To estimate the smoothing parameter, forecasts are computed for equal to .1, .2, .3, …, .9 and the sum of squared forecast error is computed for each

- The smallest Mean Square Error (MSE) is chosen for use in producing the future forecasts.

$$\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2$$

error $e_t = \hat{y}_t - y_t$

Goodness of Forecast

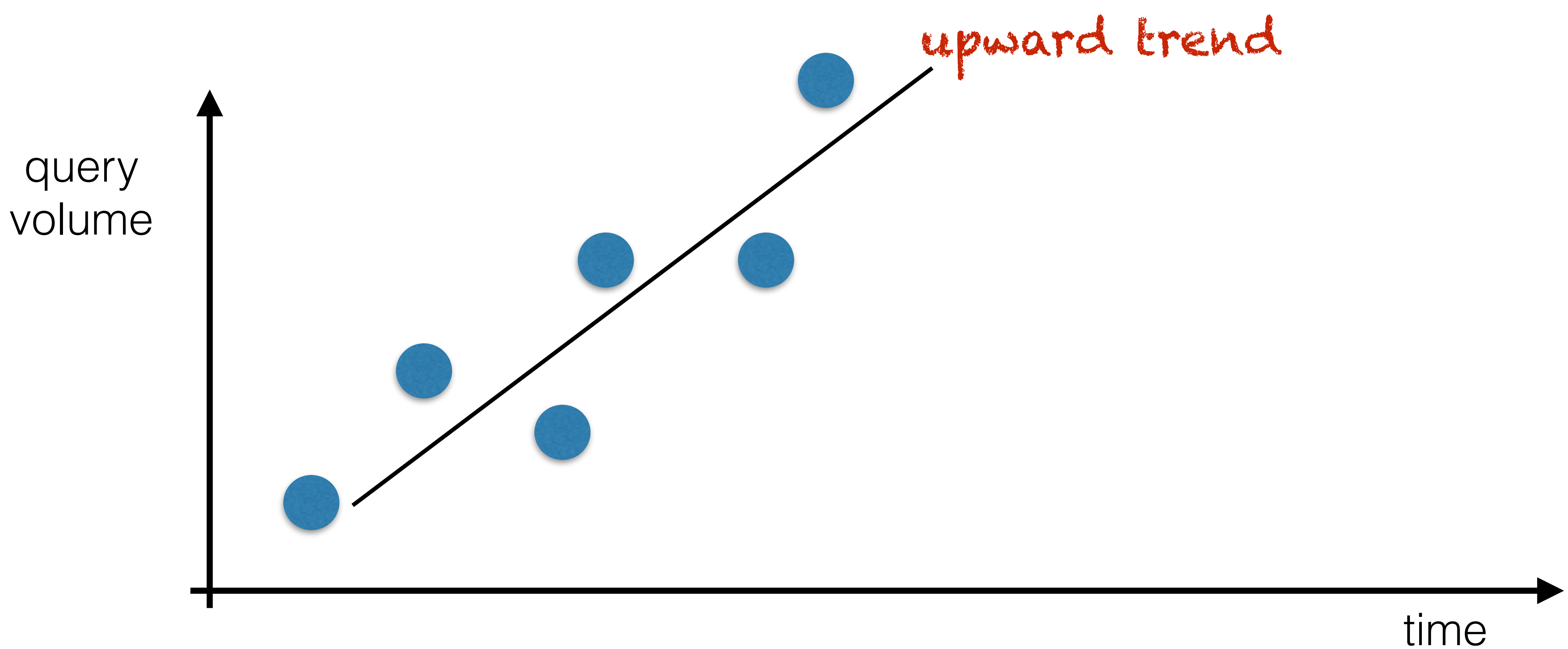
- Root Mean Squared Error (RMSE) : $\sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$

- Mean Absolute Error (MAE): $\frac{1}{n}\sum_{t=1}^{n} |e_t|$

- Mean Absolute Percentage Error (MAPE): $\frac{100\%}{n}\sum_{t=1}^{n} \left|\frac{e_t}{y_t}\right|$

Holt's Model

Double Exponential Smoothing



SES responds slow to trend data

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$$

simple exponential smoothing

Holt's model explicitly models **Trend** alongwith **Level**

Holt's Model

Double Exponential Smoothing

$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} \boxed{+ b_{t-1}})$$

Level Equation

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

Trend Equation

Adds a growth factor (or trend factor) to the smoothing equation as a way of adjusting for the trend.

Holt's Model

Double Exponential Smoothing

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} \boxed{+ b_{t-1}})$$

Level Equation

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

Trend Equation

Adds a growth factor (or trend factor) to the smoothing equation as a way of adjusting for the trend.

$$\hat{y}_{t+h|t} = \ell_t \boxed{+ hb_t}$$

Forecast Equation

Choice of Weights

- To weights α and β can be selected subjectively or by minimizing a measure of forecast error such as RMSE
- Large weights result in more rapid changes in the component. Small weights result in less rapid changes.

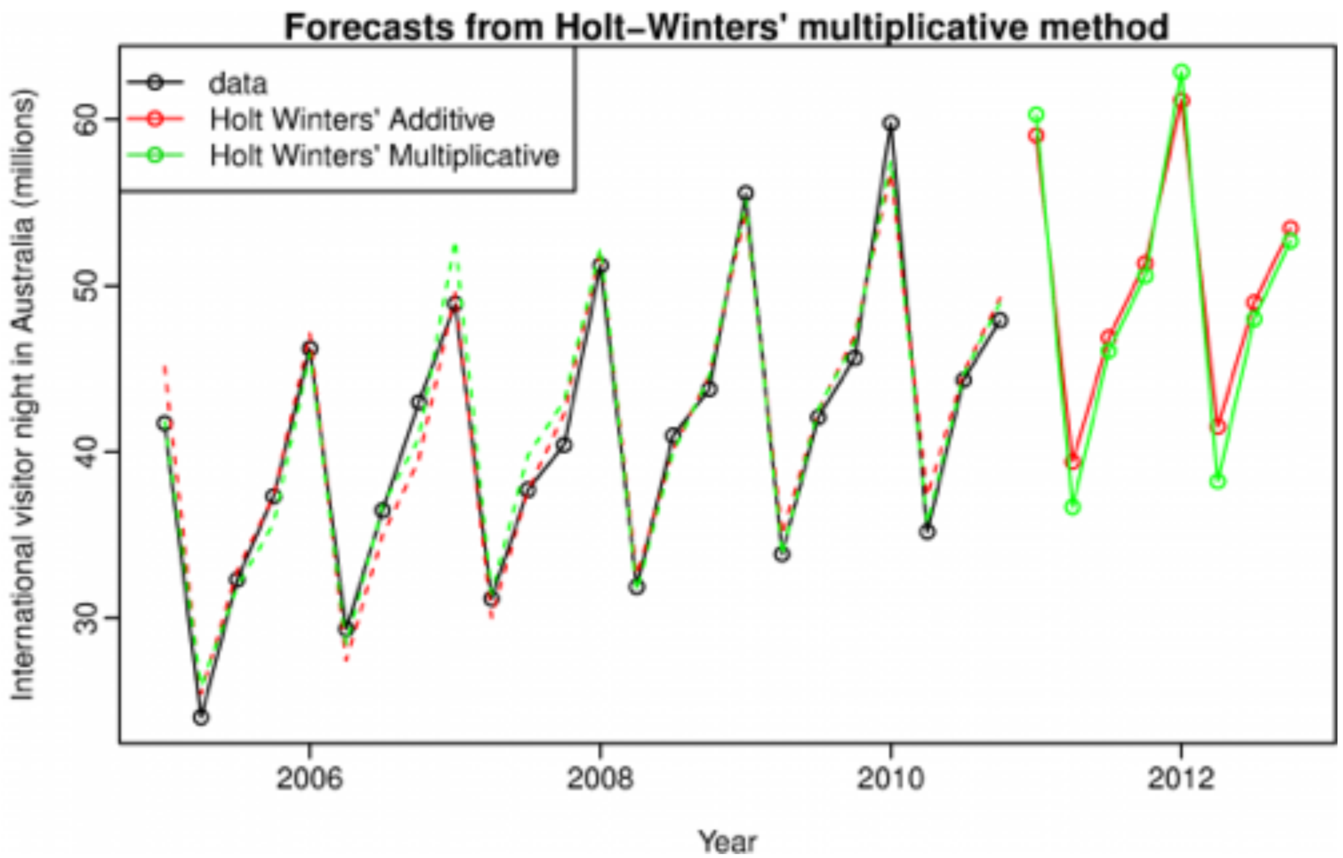
Holt-Winters Model

Triple Exponential Smoothing

- Winter's exponential smoothing model is the second extension of the basic Exponential smoothing model.
- It is used for data that exhibit both trend and seasonality.
- It is a three parameter model that is an extension of Holt's method.
- An additional equation adjusts the model for the seasonal component.

Holt-Winters Model

Triple Exponential Smoothing



$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},$$

Holt-Winters Model

Triple Exponential Smoothing

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},$$

Seasonality Equation

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

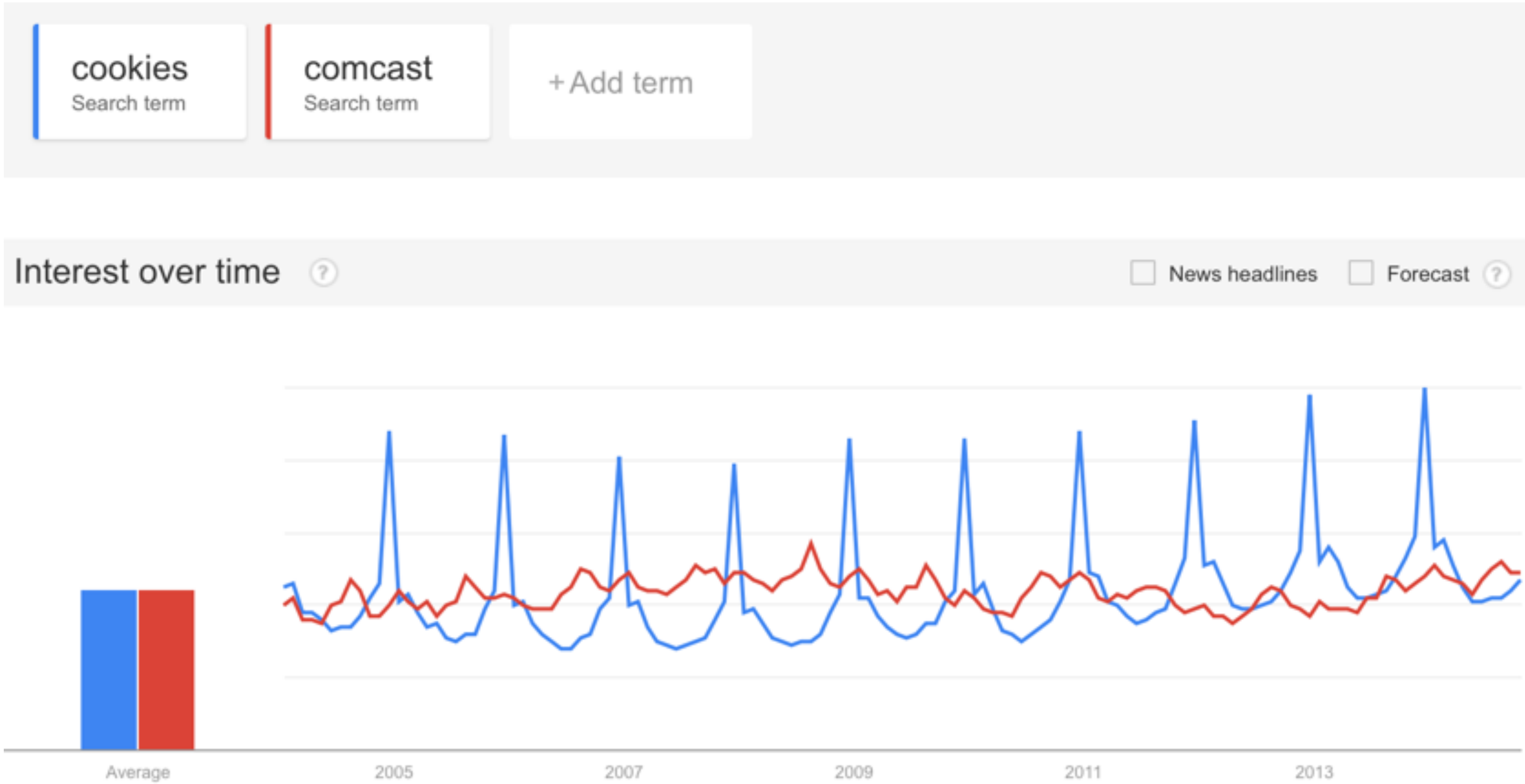
Trend Equation

$$\ell_t = \alpha(y_t \boxed{- s_{t-m}}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

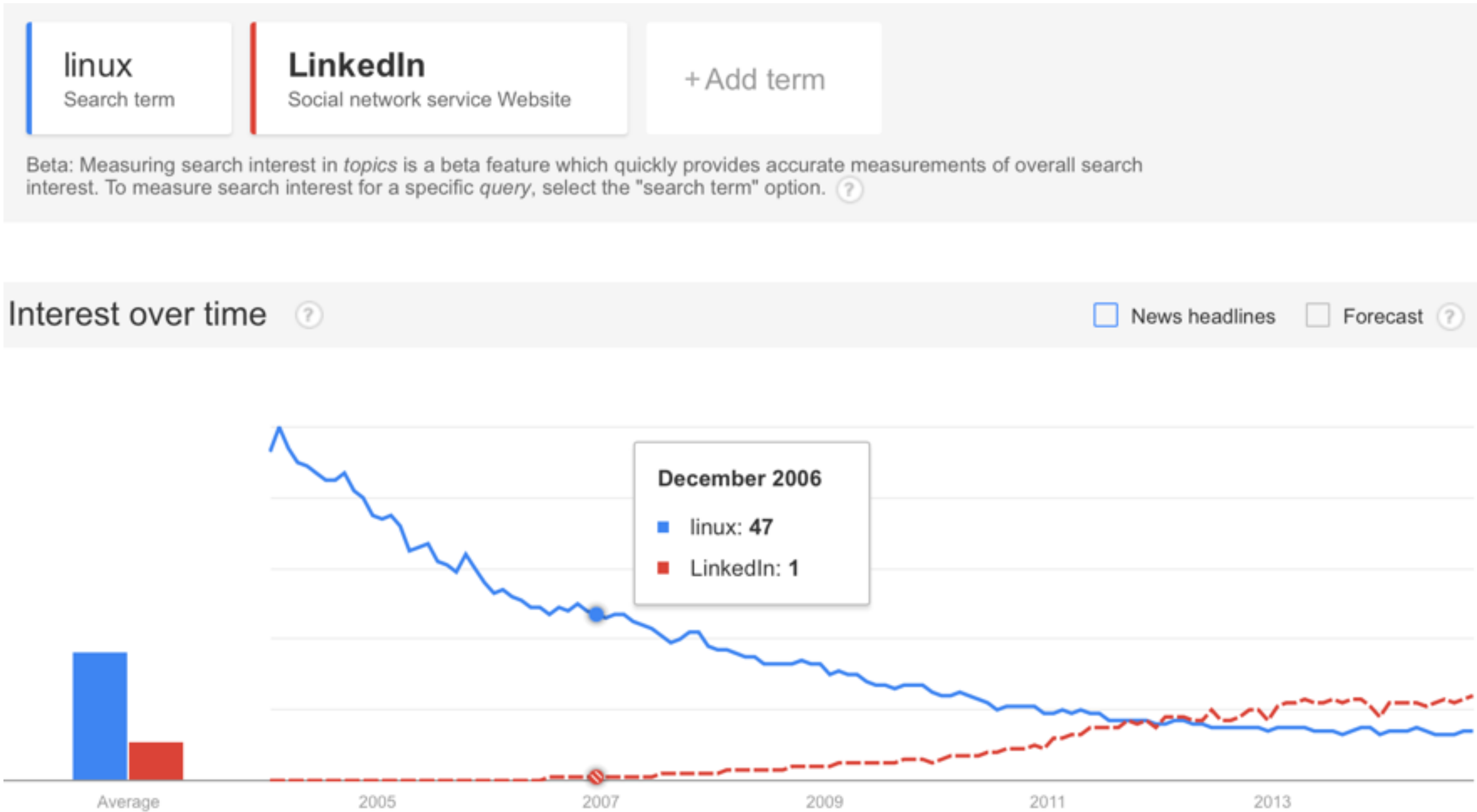
Level Equation

$$\hat{y}_{t+1|t} = \ell_t + hb_t \boxed{+ s_{t-m+h_m^+}}$$

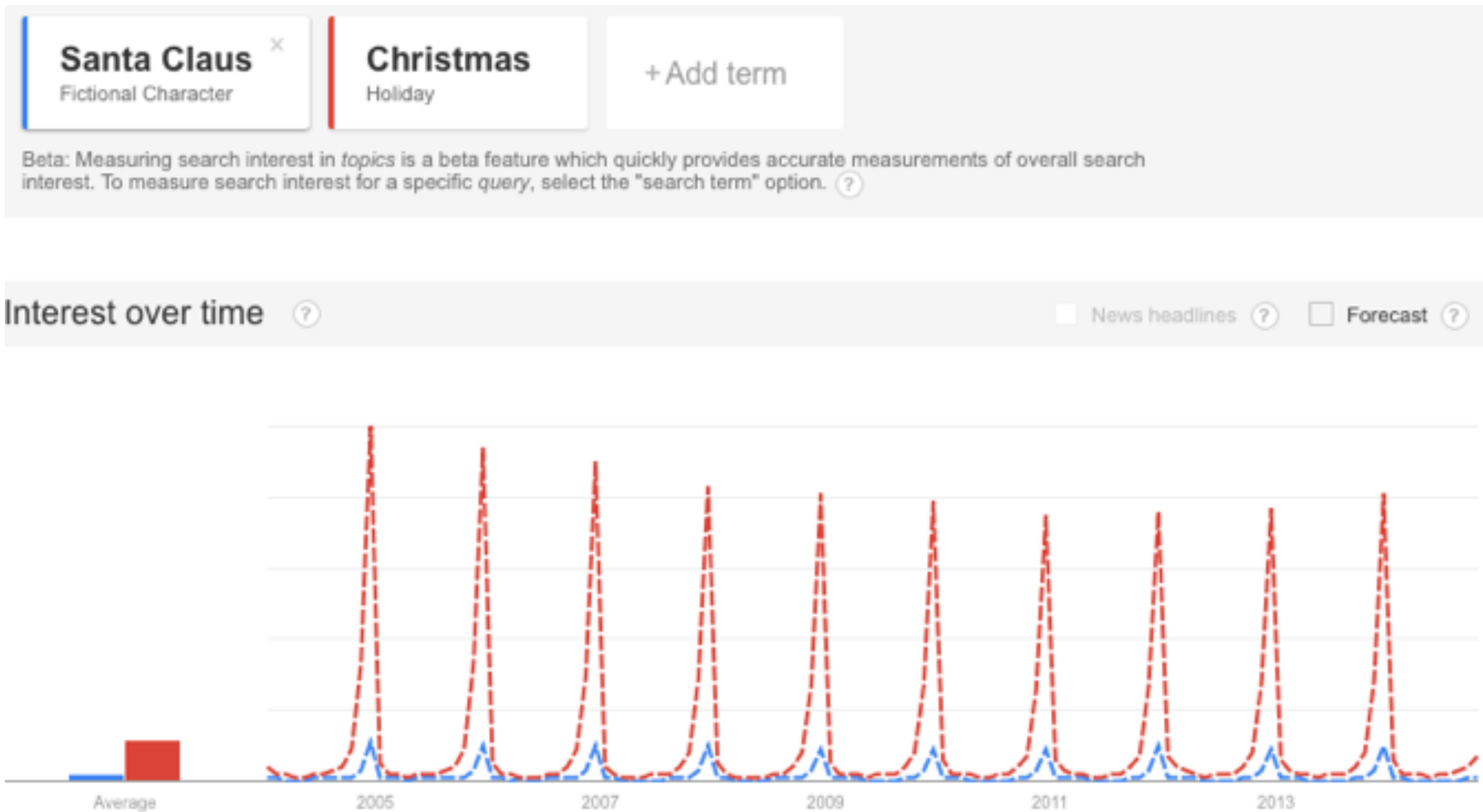
Which smoothing model to use ?



Which smoothing model to use ?



Periodicity Detection



- How does one identify periodic values

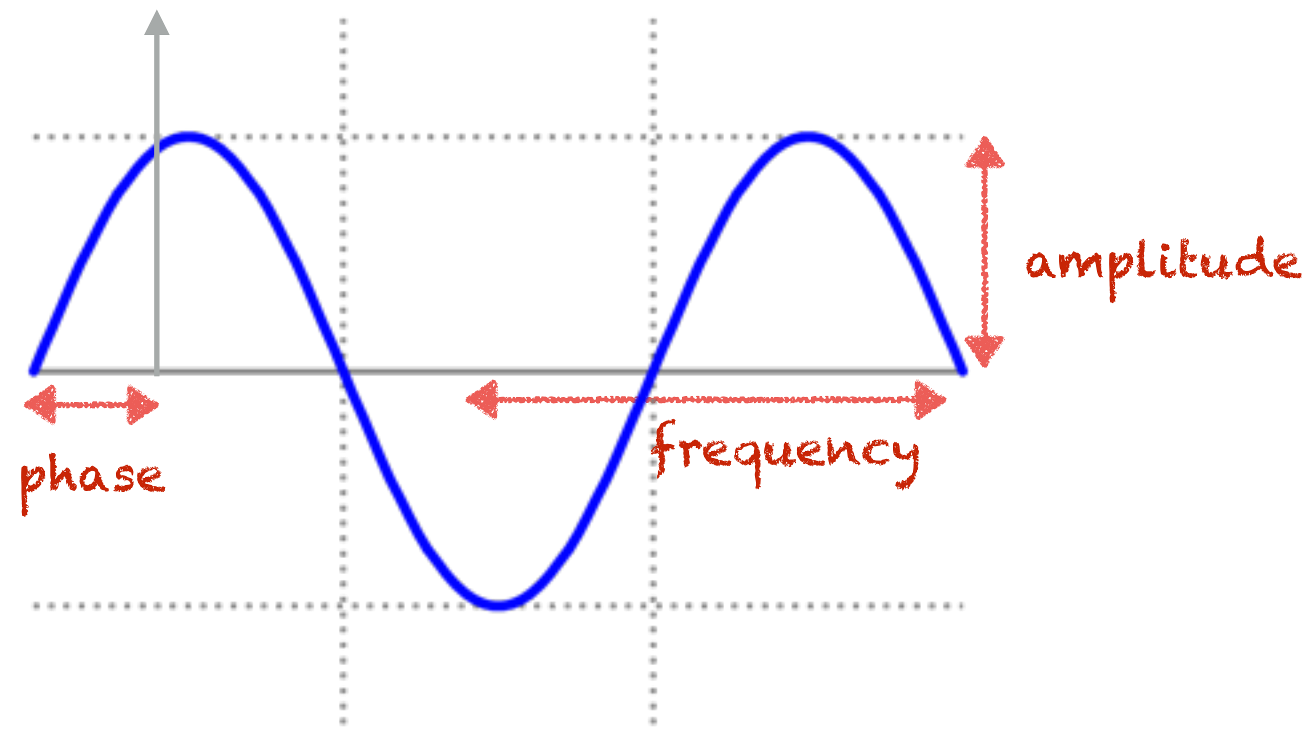
Periodicity Detection

- Time-series is in the time domain
- Method1 (DFT): Identify the underlying periodic patterns by transforming into the frequency domain
- Method 2 (Autocorrelation) Correlate the signal with itself

Find dominant frequencies

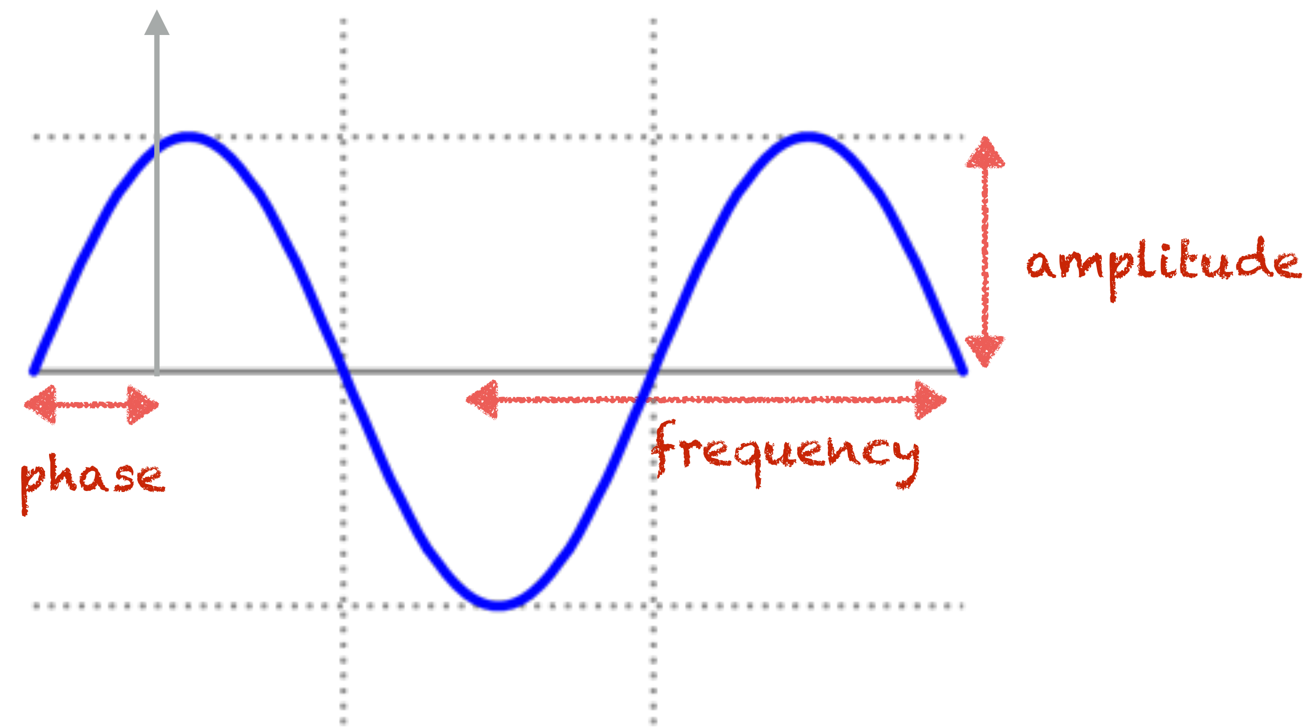
Fourier Transform

- A signal has an amplitude (strength), frequency (periodicity) and phase (offset)

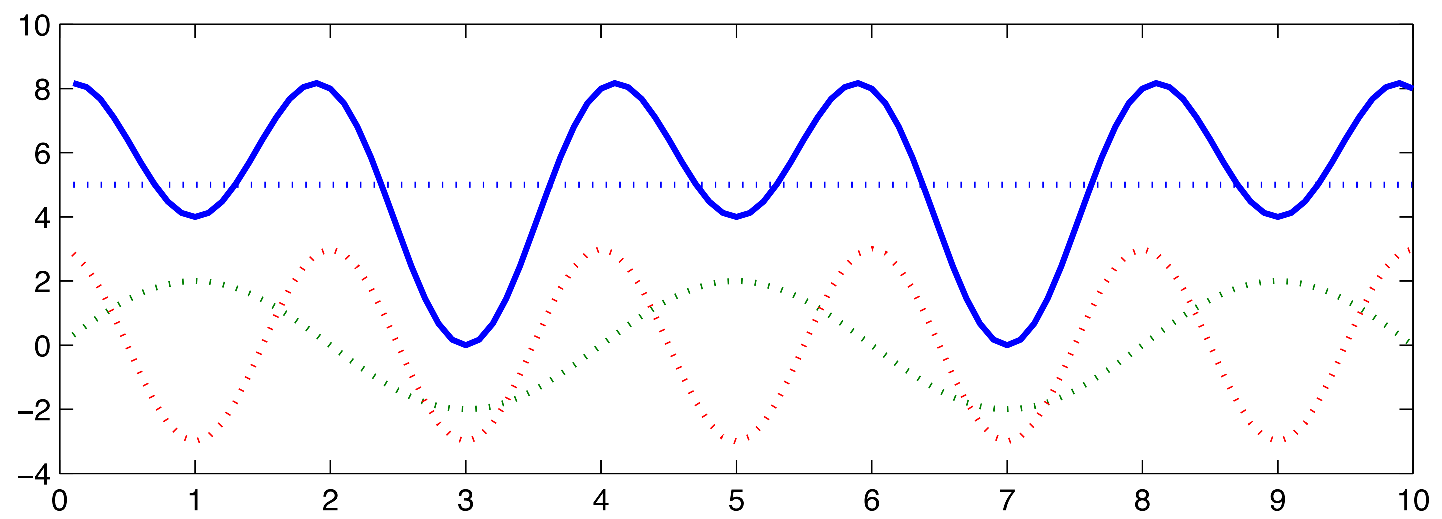


Fourier Transform

- A signal has an amplitude (strength), frequency (periodicity) and phase (offset)

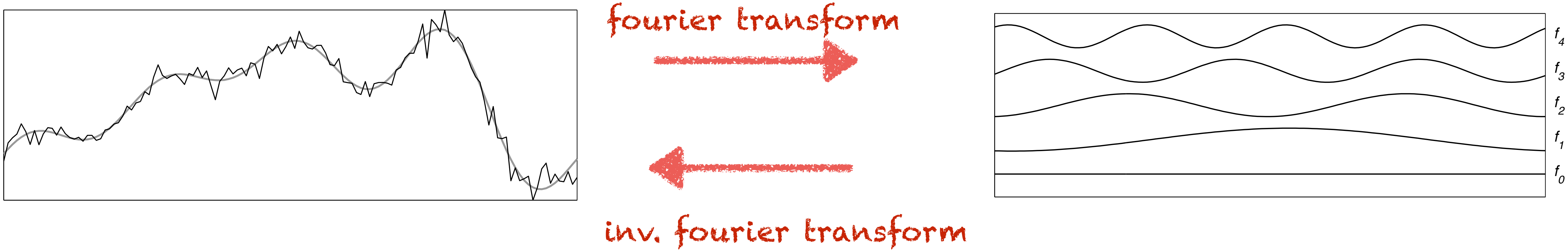


- Fourier Transform converts a signal from the time domain to the frequency domain



Discrete Fourier Transform (DFT)

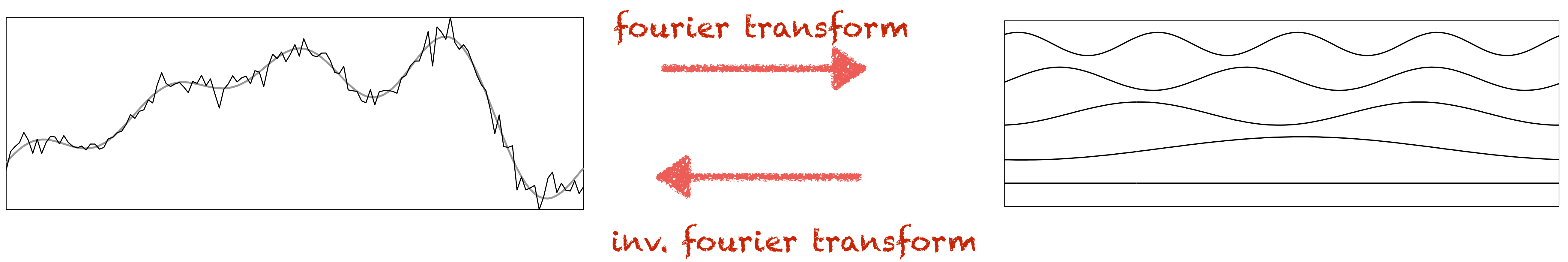
- A Fourier analysis is a method for expressing a function as a sum of periodic components, and for recovering the function from those components.
- When both the function and its Fourier transform are replaced with discretized counterparts, it is called the discrete Fourier transform (DFT).



Advantages of DFT apart from periodicity detection ?

Discrete Fourier Transform (DFT)

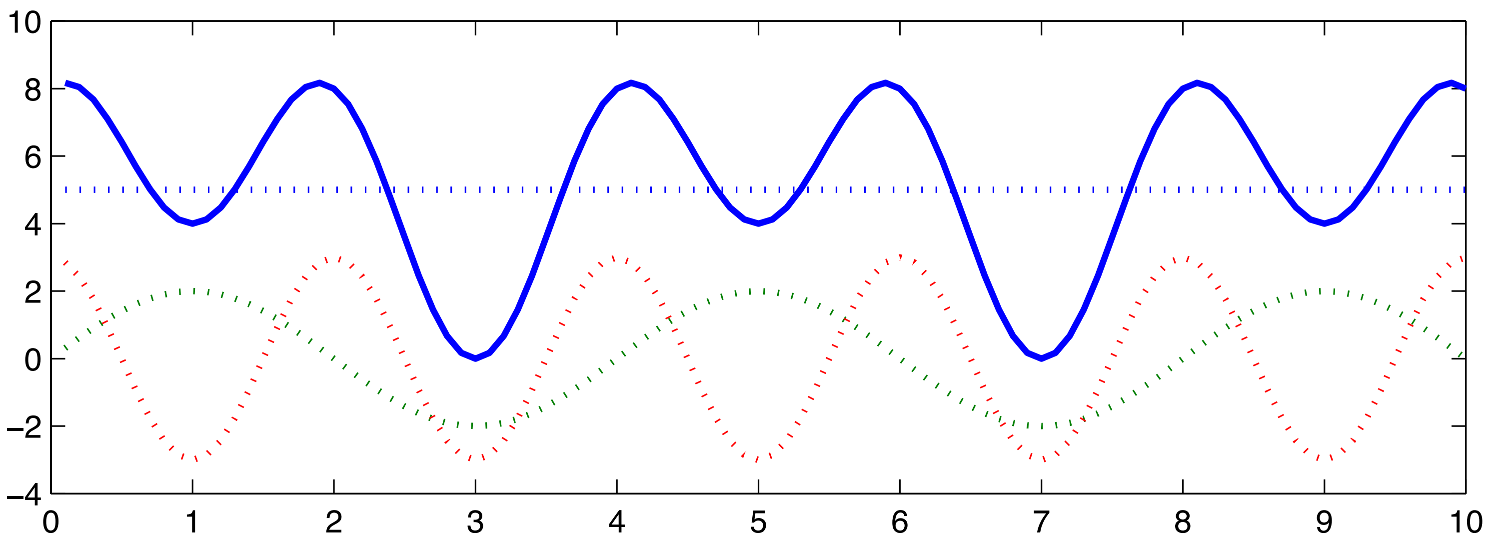
- A Fourier analysis is a method for expressing a function as a sum of periodic components, and for recovering the function from those components.
- When both the function and its Fourier transform are replaced with discretized counterparts, it is called the discrete Fourier transform (DFT).



Advantages of DFT apart from periodicity detection ?

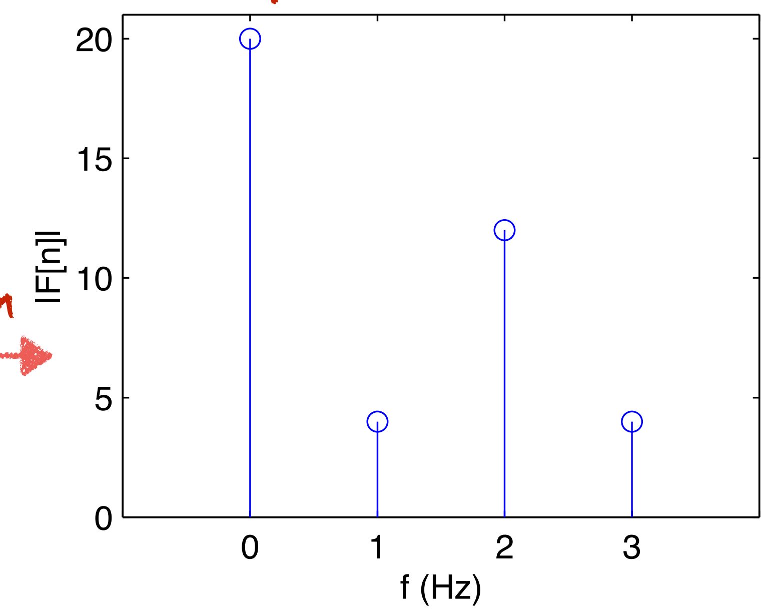
denoising, compression

Discrete Fourier Transform (DFT)



fourier transform

periodogram



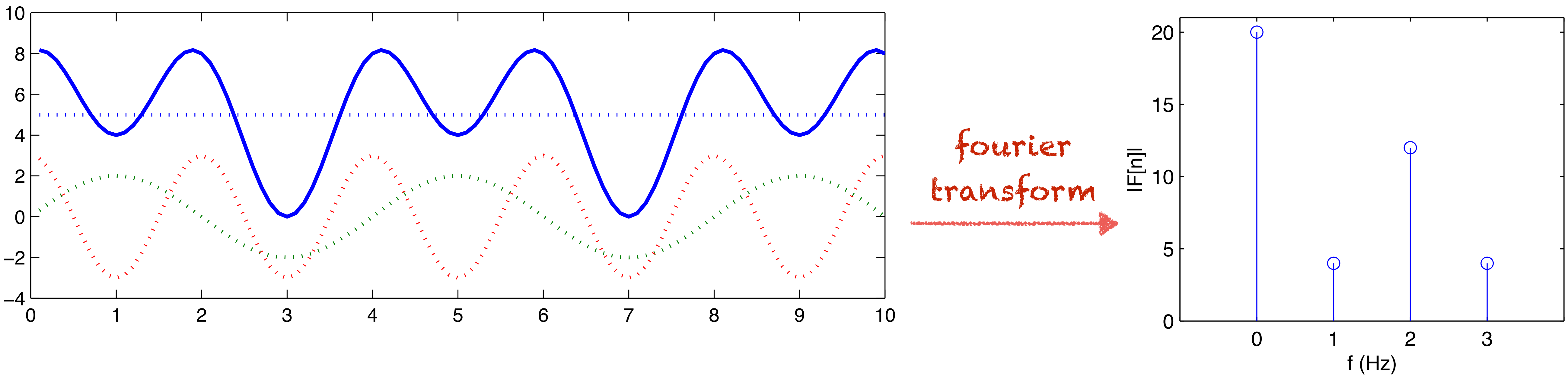
fourier coefficients

sinusoid

$$X(f_{k/N}) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n)\ e^{\frac{-j2\pi kn}{N}}$$

- The fourier coefficients encode both the amplitude and phase

Power Spectral Density (PSD) Estimation

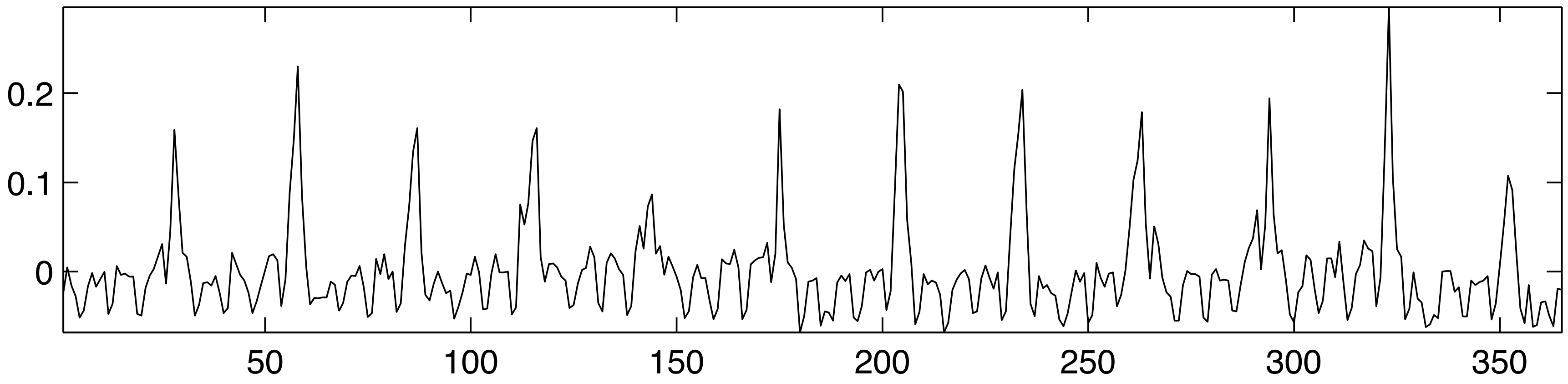


- To find out the dominant frequency we need to find the power at each frequency

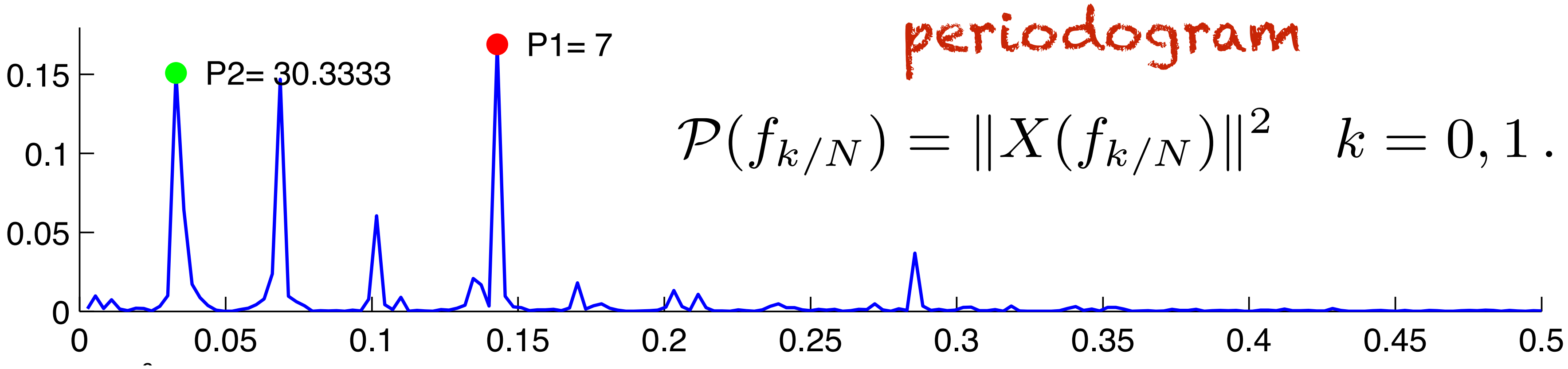
- **Periodogram** encodes the strength at a given frequency

$$\mathcal{P}(f_{k/N}) = \|X(f_{k/N})\|^2 \quad k = 0, 1 \ldots \lceil \tfrac{N-1}{2} \rceil$$

PSD estimation using Periodogram



time series data
or signal

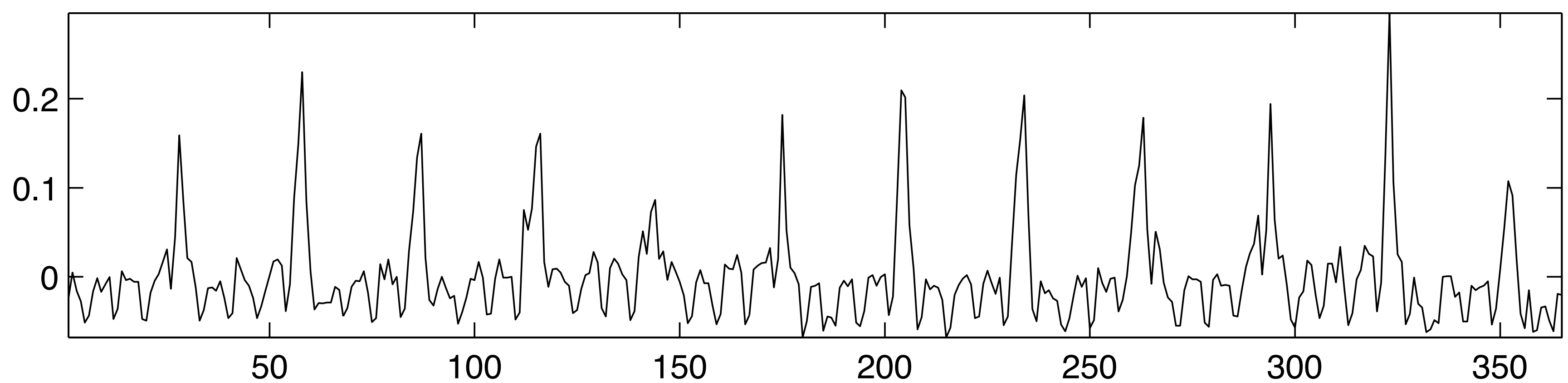


periodogram

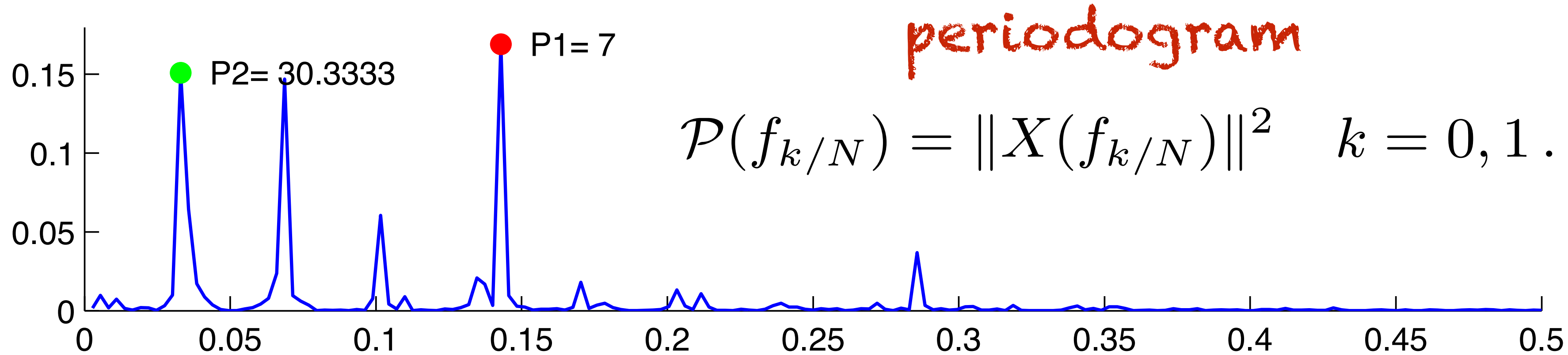
$$\mathcal{P}(f_{k/N}) = \|X(f_{k/N})\|^2 \quad k = 0, 1 \ldots \lceil \frac{N-1}{2} \rceil$$

- To find the dominant frequencies choose the top-k dominant frequencies

Disadvantages of the Periodogram



time series data or signal



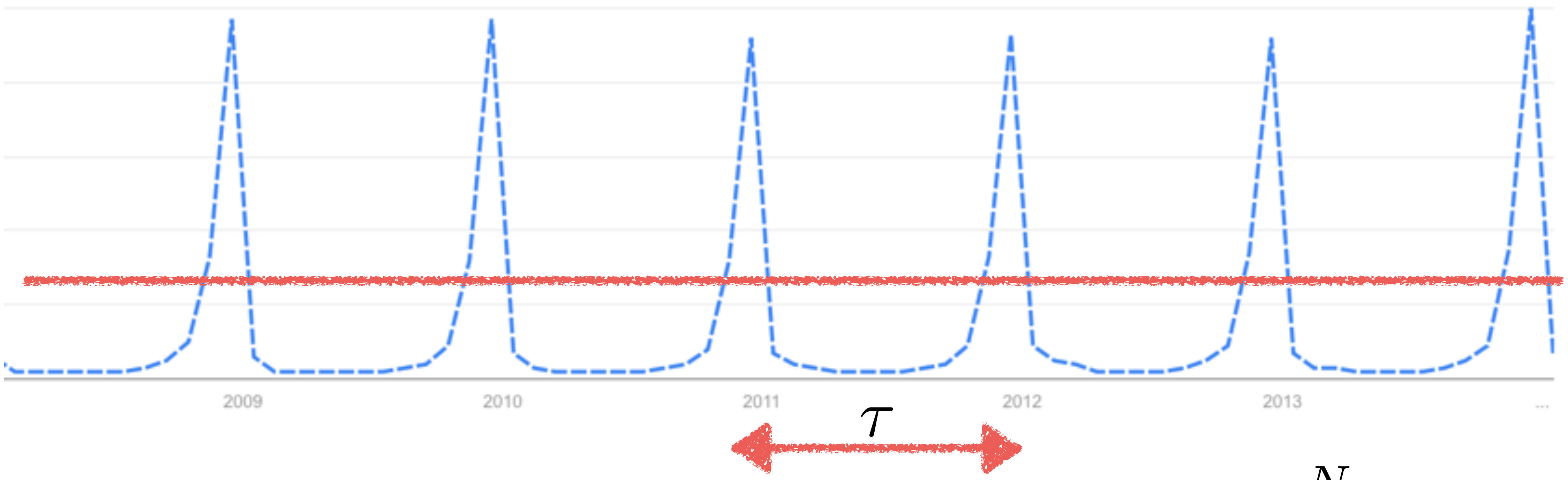
periodogram

$$\mathcal{P}(f_{k/N}) = \|X(f_{k/N})\|^2 \quad k = 0, 1 \ldots \lceil \tfrac{N-1}{2} \rceil$$

- Good only for short and medium periodicities
- Spectral leakage - frequencies not integral multiples of the DFT bin spread over other bins — false alarms

Autocorrelation

- Correlate the time series with itself

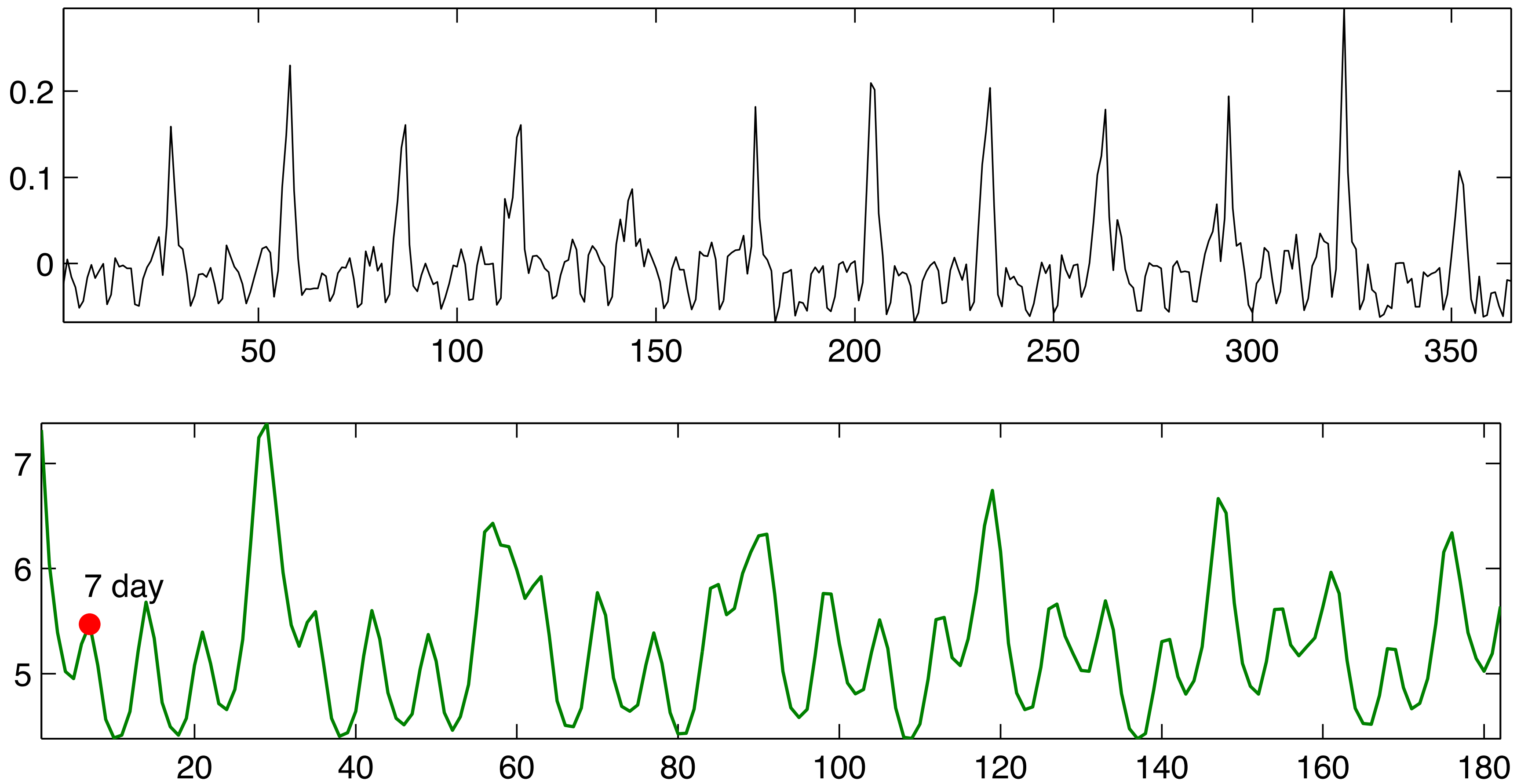


$$ACF(\tau) = \frac{\sum_{i=1}^{N} Y_i \cdot Y_{i+\tau}}{N}$$

- Peaks get amplified

- Fine-grained periodicity detector

Autocorrelation



time series data
or signal

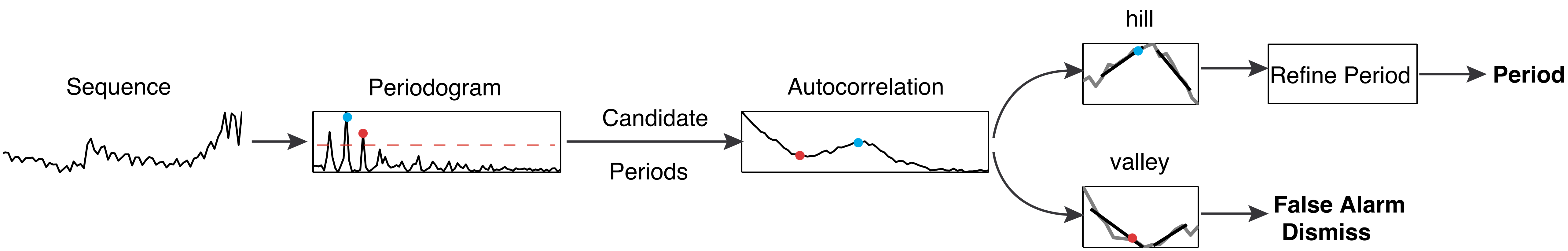
Auto-correlation

$$ACF(\tau) = \frac{\sum_{i=1}^{N} Y_i \cdot Y_{i+\tau}}{N}$$

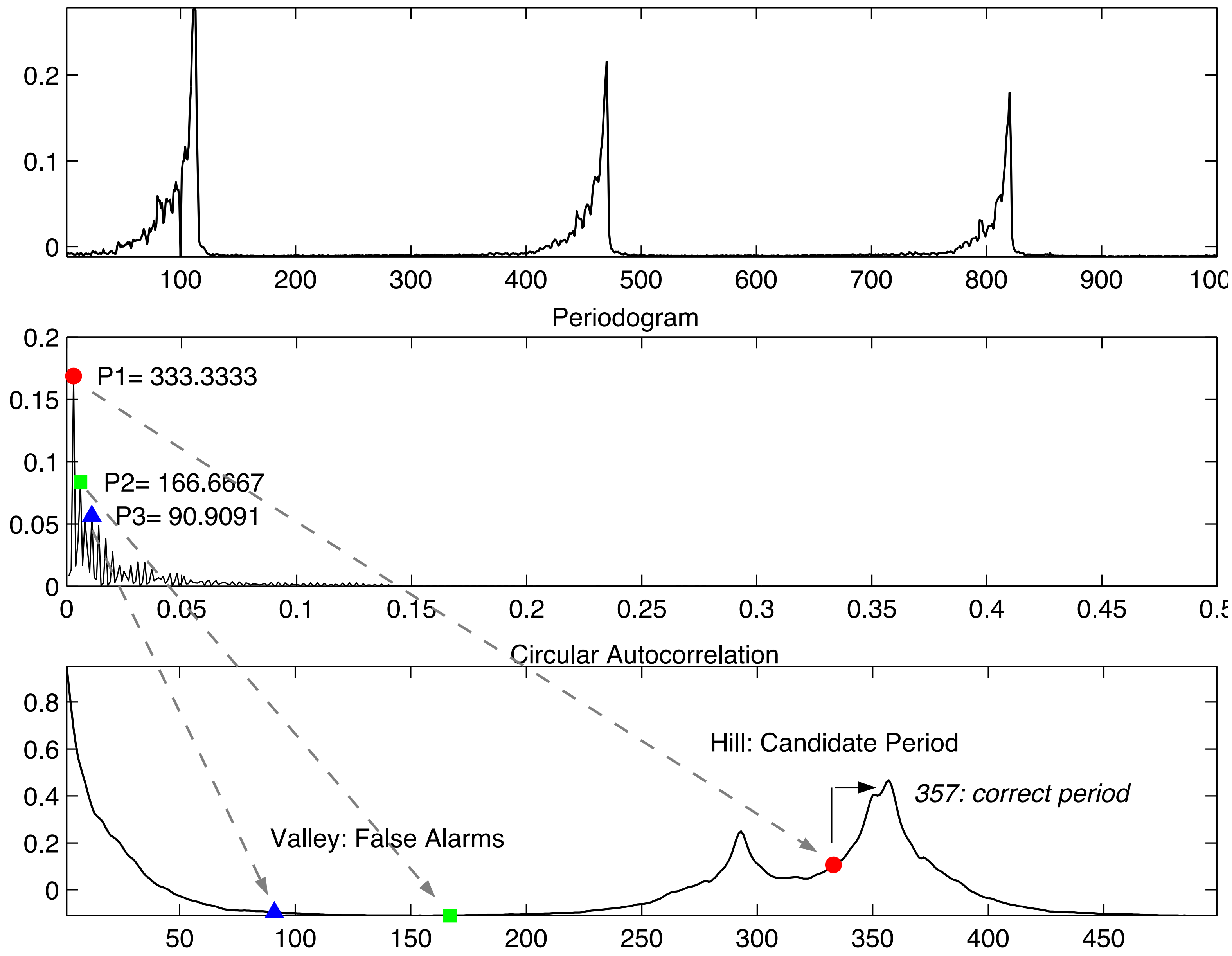
- To determine dominant period significance threshold needs to be specified
- Multiples of the same period are also peaks — needs post processing

Auto-Period

- **Auto-correlation :** Good for large periods but difficult to automatically determine periods
- **Periodogram :** Easy to threshold but accurate for short periods
- **Idea:** Get candidate periods from Periodogram and validate false alarms using Auto-correlation



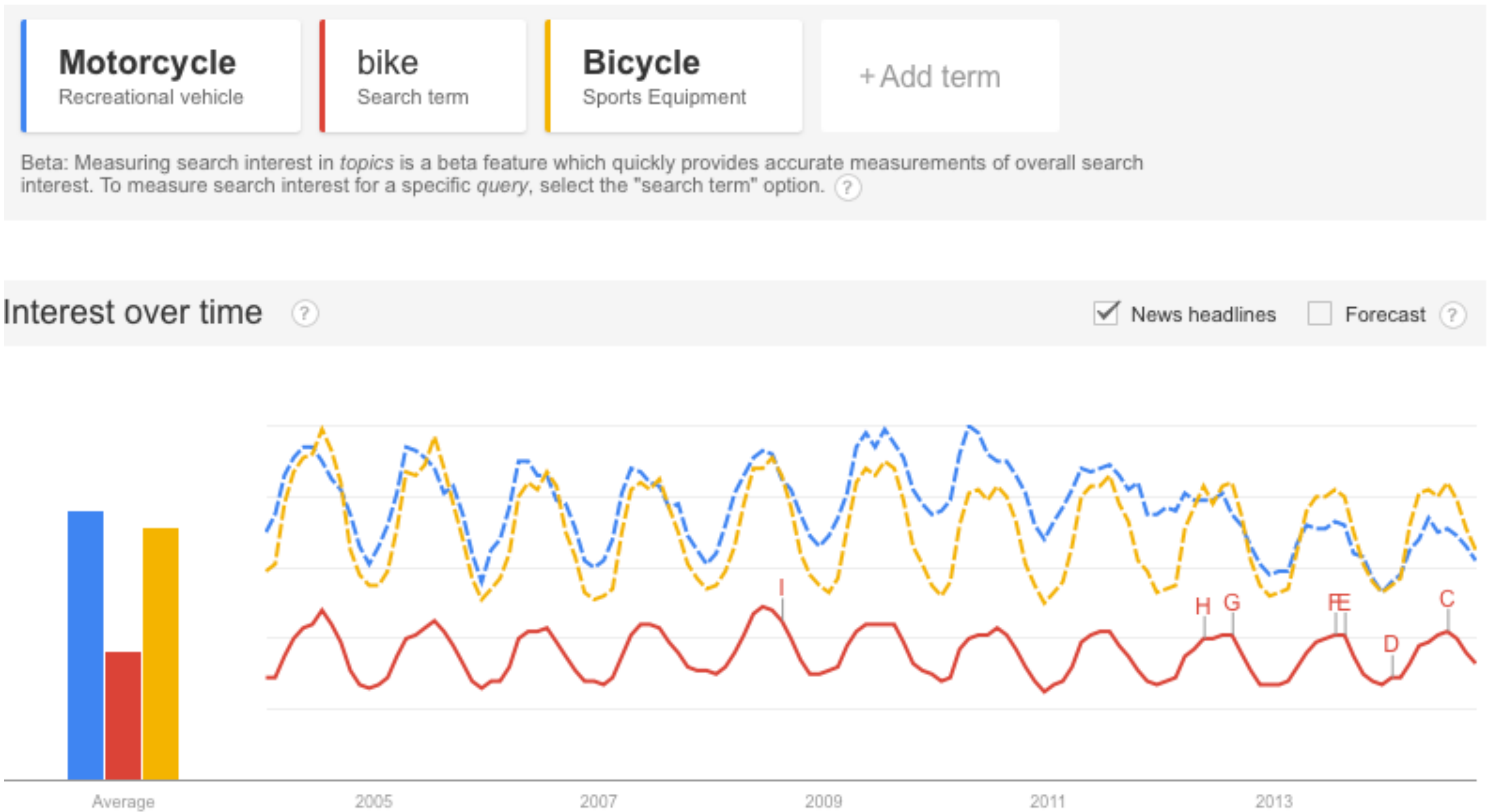
Auto-Period



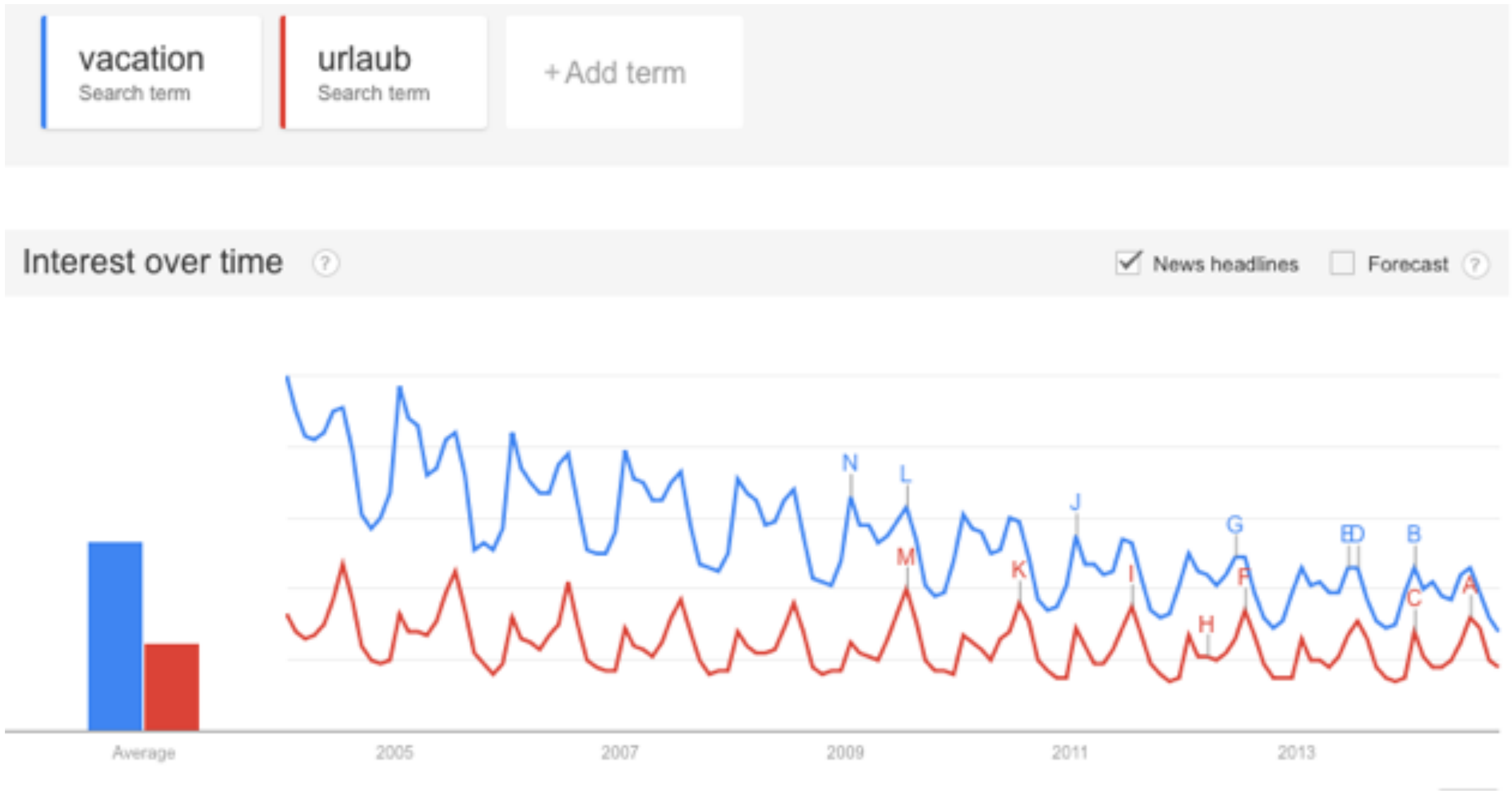
Periodogram
P1= 333.3333
P2= 166.6667
P3= 90.9091
Circular Autocorrelation
Hill: Candidate Period
357: correct period
Valley: False Alarms

Matching Time Series

- Similar time series suggest similar things



Matching Time Series



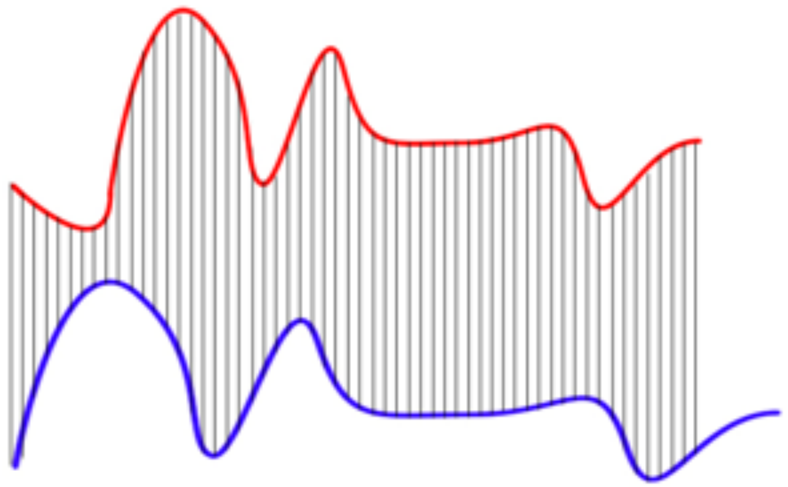
- Correlating time series used for clustering, classification, anomaly detection, speech recognition etc.

Matching Time Series

What measure would you use to match two time series ?

$$d = \sum_t |y_t - x_t|$$

Euclidean Distance



Euclidean Matching

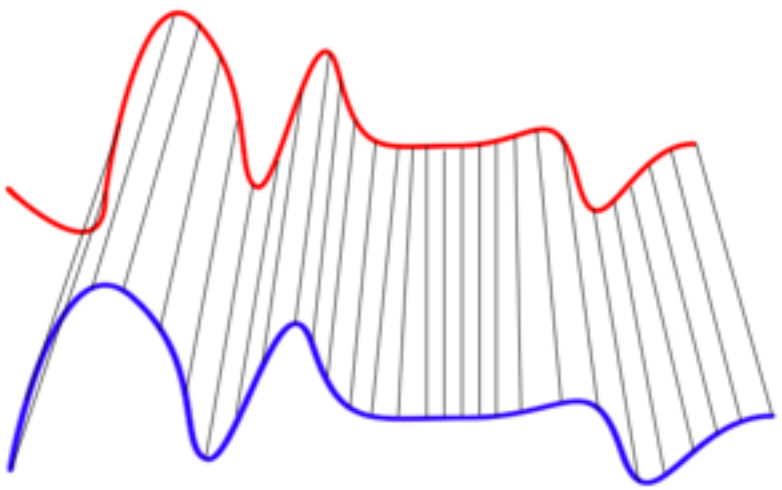
Why is Euclidean matching not good enough ?

Dynamic Time Warping

Time series might be shifted

Time series might be compressed at some point in time

Random noise at some points



Dynamic Time Warping Matching

Dynamic time warping measures the distance between two sequences under certain restrictions.

Not a metric. Triangle inequality doesn't hold

Detour - Edit Distance

- Edit distance measures how many steps it takes to convert a string to another based on restrictions
- Restrictions define cost function — insertion, deletion, replacement

	f	o	x
f	0	1	2
a	1	2	3
x	2	3	2

insertions and deletions

	f	o	x
f	0	1	2
a	1	1	2
x	2	2	1

insertions, deletions and replacements

Edit Distance

	f	o	x
f	0	1	2
a	1	2	3
x	2	3	2

insertions and deletions

	f	o	x
f	0	1	2
a	1	1	2
x	2	2	1

insertions, deletions and replacements

$$d_{ij} = \begin{cases} d_{i-1,j-1} & a_j = b_i \\ \min \begin{cases} d_{i-1,j} + w_{\mathrm{del}}(b_i) \longrightarrow 1 \\ d_{i,j-1} + w_{\mathrm{ins}}(a_j) \longrightarrow 1 \\ d_{i-1,j-1} + w_{\mathrm{sub}}(a_j, b_i) \longrightarrow 1 \text{ or } 2 \end{cases} & a_j \neq b_i \end{cases}, \quad for\ 1 \leq i \leq m, 1 \leq j \leq n.$$

Edit Distance

	f	o	x
f	0	1	2
a	1	2	3
x	2	3	2

insertions and deletions

	f	o	x
f	0	1	2
a	1	1	2
x	2	2	1

insertions, deletions and replacements

$$d_{ij} = \begin{cases} d_{i-1,j-1} & a_j = b_i \\ \min \begin{cases} d_{i-1,j} + w_{\mathrm{del}}(b_i) \longrightarrow 1 \\ d_{i,j-1} + w_{\mathrm{ins}}(a_j) \longrightarrow 1 \\ d_{i-1,j-1} + w_{\mathrm{sub}}(a_j, b_i) \longrightarrow 1 \text{ or } 2 \end{cases} & a_j \neq b_i \end{cases}, \quad for\ 1 \leq i \leq m, 1 \leq j \leq n.$$

Edit Distance

	f	o	x
f	0	1	2
a	1	2	3
x	2	3	2

insertions and deletions

	f	o	x
f	0	1	2
a	1	1	2
x	2	2	1

insertions, deletions and replacements

$$d_{ij} = \begin{cases} d_{i-1,j-1} & a_j = b_i \\ \min \begin{cases} d_{i-1,j} + w_{\text{del}}(b_i) \longrightarrow 1 \\ d_{i,j-1} + w_{\text{ins}}(a_j) \longrightarrow 1 \\ d_{i-1,j-1} + w_{\text{sub}}(a_j, b_i) \longrightarrow 1 \text{ or } 2 \end{cases} & a_j \neq b_i \end{cases}, \quad for\ 1 \leq i \leq m, 1 \leq j \leq n.$$

Edit Distance

	f	o	x
f	0	1	2
a	1	2	3
x	2	3	2

insertions and deletions

	f	o	x
f	0	1	2
a	1	1	2
x	2	2	1

insertions, deletions and replacements

$$d_{ij} = \begin{cases} d_{i-1,j-1} & a_j = b_i \\ \min \begin{cases} d_{i-1,j} + w_{\text{del}}(b_i) \longrightarrow 1 \\ d_{i,j-1} + w_{\text{ins}}(a_j) \longrightarrow 1 \\ d_{i-1,j-1} + w_{\text{sub}}(a_j, b_i) \longrightarrow 1 \text{ or } 2 \end{cases} & a_j \neq b_i \end{cases}, \quad for\ 1 \leq i \leq m, 1 \leq j \leq n.$$

Edit Distance

	f	o	x
f	0	1	2
a	1	2	3
x	2	3	2

insertions and deletions

	f	o	x
f	0	1	2
a	1	1	2
x	2	2	1

insertions, deletions and replacements

$$d_{ij} = \begin{cases} d_{i-1,j-1} & a_j = b_i \\ \min \begin{cases} d_{i-1,j} + w_{\mathrm{del}}(b_i) \longrightarrow 1 \\ d_{i,j-1} + w_{\mathrm{ins}}(a_j) \longrightarrow 1 \\ d_{i-1,j-1} + w_{\mathrm{sub}}(a_j, b_i) \longrightarrow 1 \text{ or } 2 \end{cases} & a_j \neq b_i \end{cases}, \quad for\ 1 \leq i \leq m, 1 \leq j \leq n.$$

Edit Distance

	f	o	x
f	0	1	2
a	1	2	3
x	2	3	2

insertions and deletions

	f	o	x
f	0	1	2
a	1	1	2
x	2	2	1

insertions, deletions and replacements

$$d_{ij} = \begin{cases} d_{i-1,j-1} & a_j = b_i \\ \min \begin{cases} d_{i-1,j} + w_{\text{del}}(b_i) \longrightarrow 1 \\ d_{i,j-1} + w_{\text{ins}}(a_j) \longrightarrow 1 \\ d_{i-1,j-1} + w_{\text{sub}}(a_j, b_i) \longrightarrow 1 \text{ or } 2 \end{cases} & a_j \neq b_i \end{cases}, \quad for\ 1 \leq i \leq m, 1 \leq j \leq n.$$

Edit Distance

	f	o	x
f	0	1	2
a	1	2	3
x	2	3	2

insertions and deletions

	f	o	x
f	0	1	2
a	1	1	2
x	2	2	1

insertions, deletions and replacements

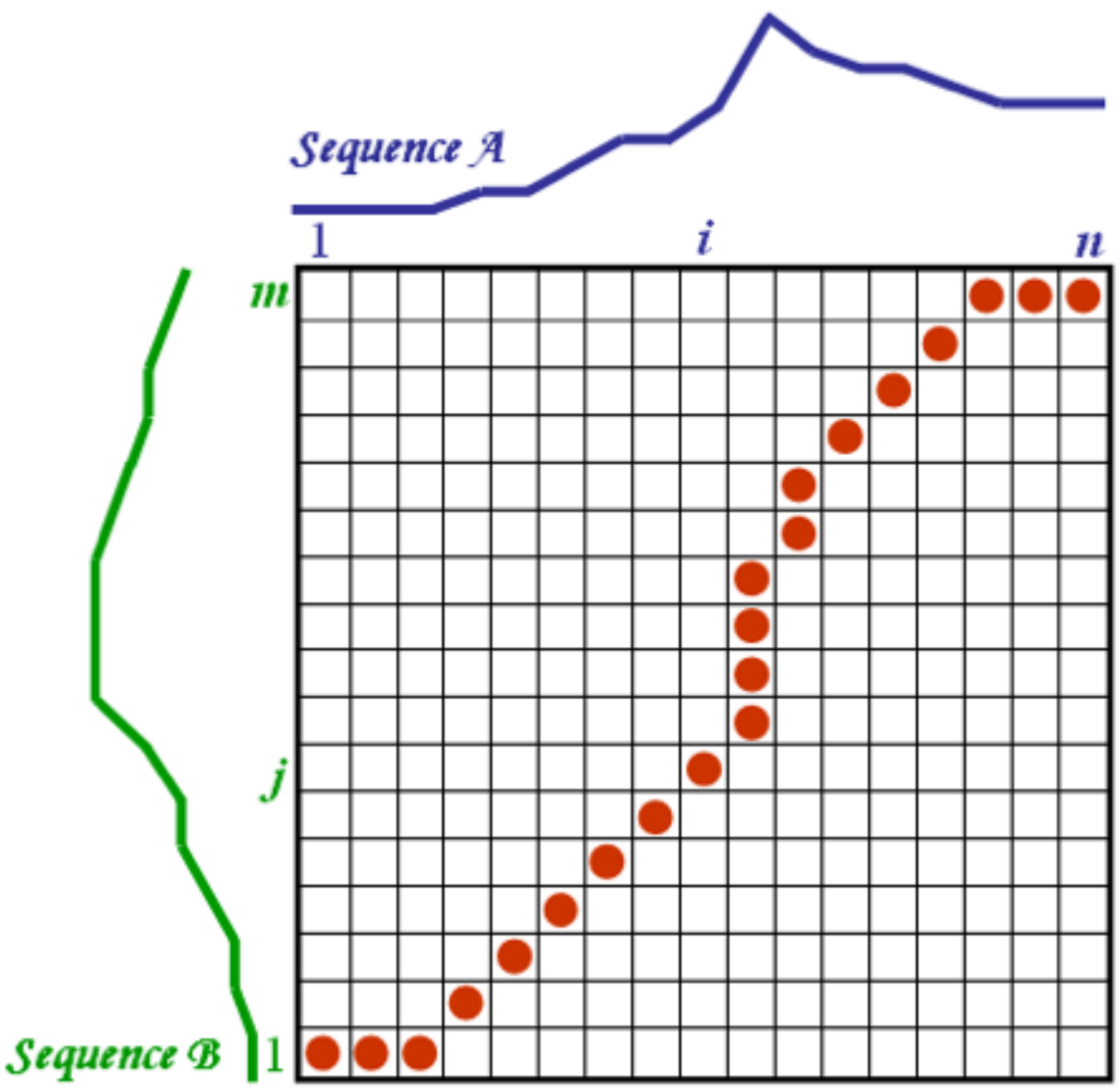
$$d_{ij} = \begin{cases} d_{i-1,j-1} & a_j = b_i \\ \min \begin{cases} d_{i-1,j} + w_{\mathrm{del}}(b_i) \longrightarrow \text{1} \\ d_{i,j-1} + w_{\mathrm{ins}}(a_j) \longrightarrow \text{1} \\ d_{i-1,j-1} + w_{\mathrm{sub}}(a_j, b_i) \longrightarrow \text{1 or 2} \end{cases} & a_j \neq b_i \end{cases}, \quad for\ 1 \leq i \leq m, 1 \leq j \leq n.$$

Dynamic Time Warping

- DTW aligns two sequences of feature vectors by warping the time axis iteratively until an optimal match (according to a suitable metrics) between the two sequences is found.

$$x_1, x_2, \ldots x_n \qquad\qquad y_1, y_2, \ldots y_n$$

$$d(i,j) = c(i,j) + min \begin{cases} d(i-1,j), \\ d(i-1,j-1), \\ d(i,j-1) \end{cases}$$



Summary

- Forecasting of time-series and their applications in IR applications
- Components of a Time series - Level, Trend and Seasonality
- Exponential Smoothing Methods vs Averaging Methods
- Level oriented Methods - Simple Exponential Smoothing
- Adding Trend - Double Exponential Smoothing
- Adding Seasonality - Triple Exponential Smoothing

Summary

- Periodicity of Events
  - Auto-correlation, Periodograms and their combinations
- Burst Detection Techniques and elastic detection
- Matching of time series
  - Euclidean matching
  - Dynamic Time Warping

References

- **Forecasting: principles and practice.** (Chapter 7)
  - https://www.otexts.org/fpp/
- **Introduction to Time Series and Forecasting.**
  - http://www.masys.url.tw/Download/2002-Brockwell-Introduction%20Time%20Series%20and%20Forecasting.pdf
- **“On Periodicity Detection and Structural Periodic Similarity”, 2005.**
  - Michail Vlachos, Philip Yu, Vittorio Castelli
- **Burst Detection in Hierarchical Streams.**
  - Jon Kleinberg
- **Everything you know about Dynamic Time Warping is wrong**
  - Eamonn Keogh