STUTERN

Time Series Analysis

SGA07_DATASCI

17th March 2020

Module Overview

- Overview of Time Series
- Core Components of Time Series Data
- Core Concepts: Autocorrelation & Stationary
- Smoothing Techniques: Moving Average
- Forecast of US Electricity Price in R

Time Series (Def.)

- Application
  - Obtain an understanding of the underlying forces and structure that produced the observed data
  - Fit a model and proceed to forecasting, monitoring or even feedback and feedforward control.

> "An ordered sequence of values of a variable at equally spaced time intervals."

Time Factors

- Yearly
- Quarterly
- Monthly
- Weekly
- Daily
- Hourly

- Use cases
  - Financial Data
  - Electricity Data
  - Signal Processing

Core Components

Trend is the increase and decrease in the series over a period of time, it persists over a long period of time

Seasonality is the regular pattern of up and down fluctuations. It is a short-term variation occurring due to seasonal factors

Cyclicity is the medium-term variation caused by circumstances, which repeat in irregular intervals

Irregularity is the variation which occurs due to unpredictable factors and also do not repeat in particular patterns

Core Concepts: Stationary

- Mean is constant with time
- Variance is constant with time
- Covariance is constant with time

> "When all core components are present in a time series data, it is a Non-stationary time series data. i.e the mean and variance of the time series data is non-constant with a clearly defined trend."

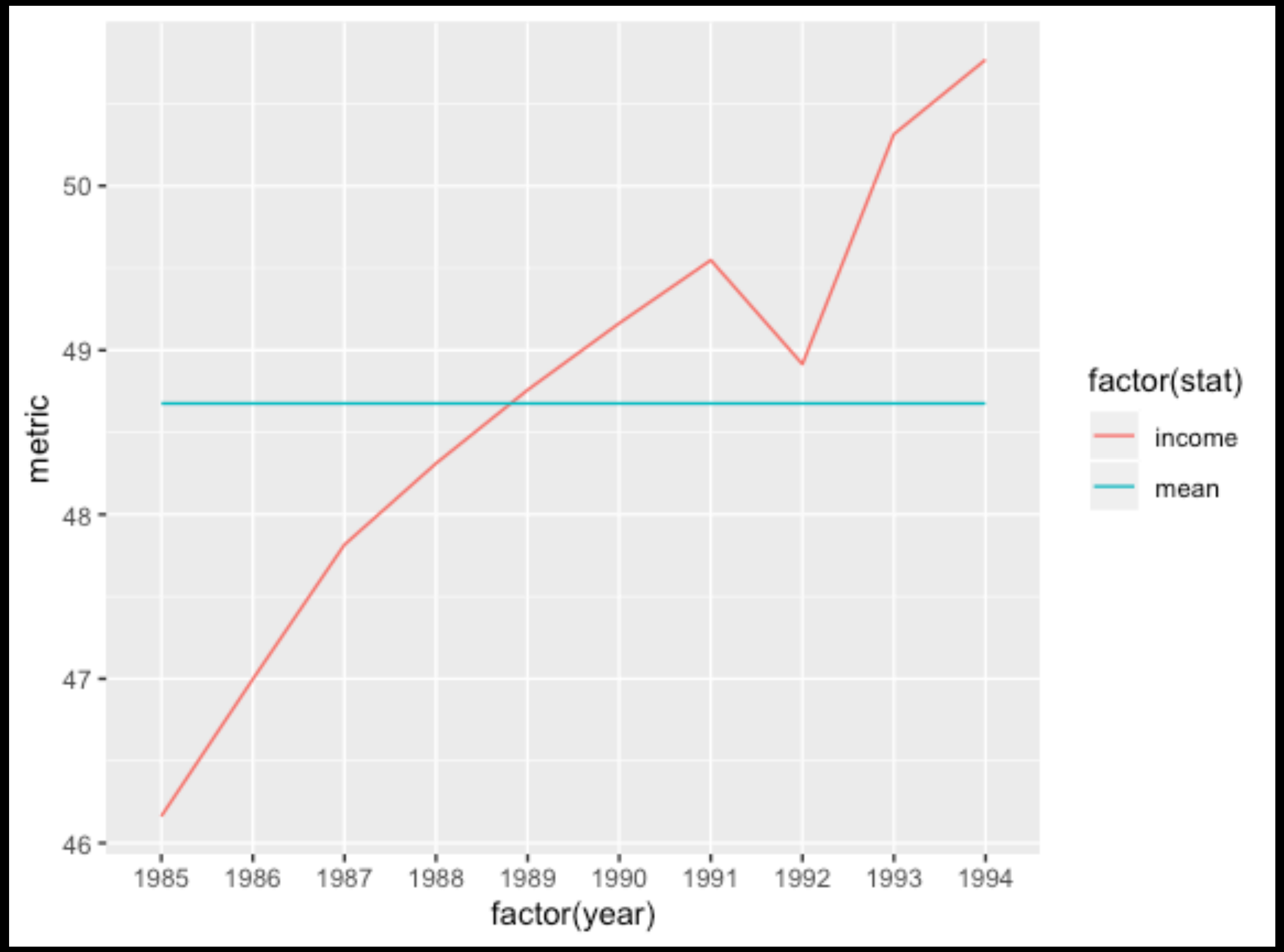
Smoothing Techniques

- Reduce canceling effect due to random variations
- Revels more about the underlying trend, seasonal and cyclic components
- Two smoothing methods
  - Moving average method
  - Exponential smoothing Method

STUTERN

Recap on Mean Squared Errors

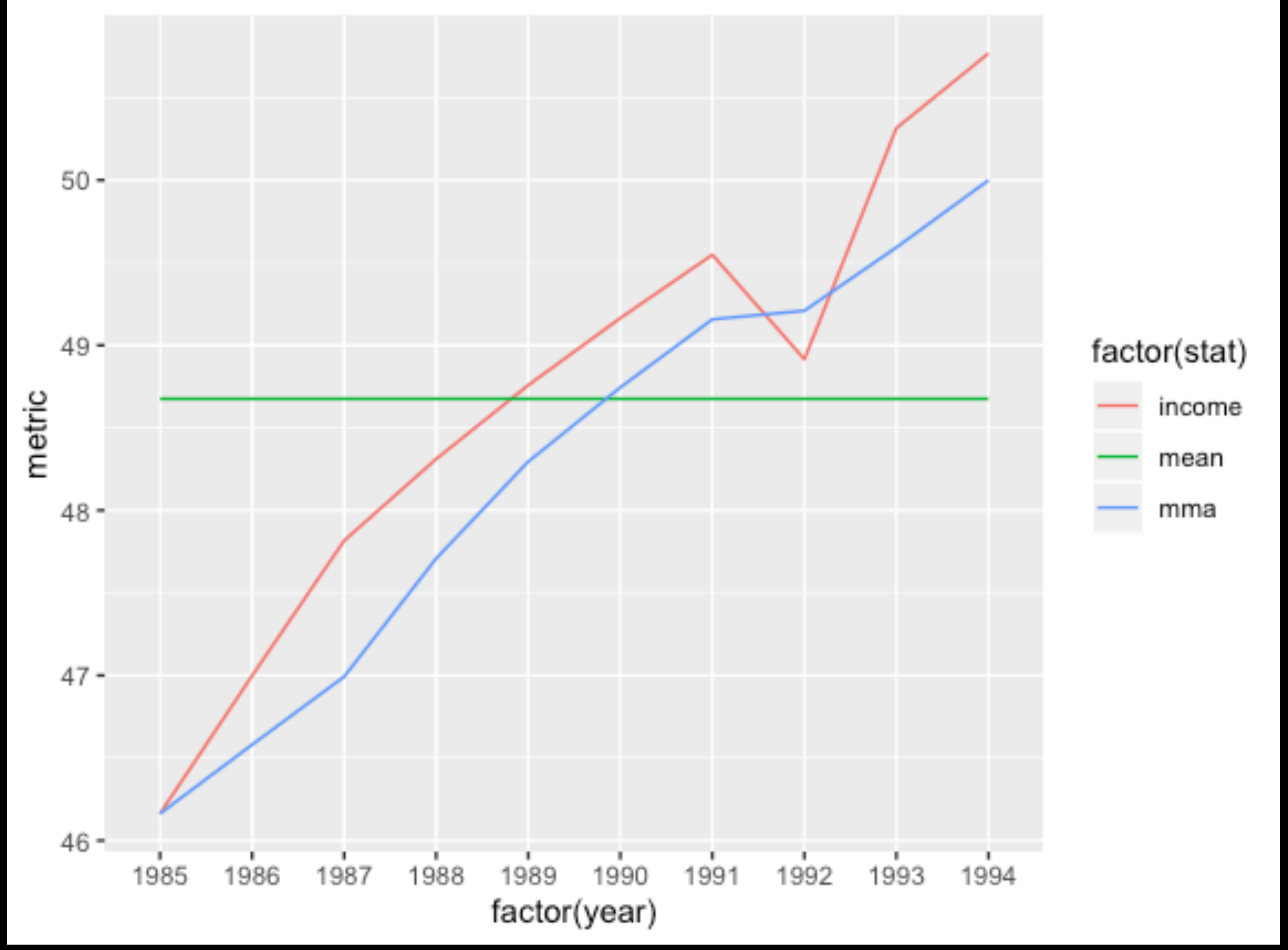
- The next table gives the income before taxes of a PC manufacturer between 1985 and 1994.



Years	$ (millions)	Error	SSE
1985	46.163	-2.5126	6.31315876
1986	46.998	-1.6776	2.81434176
1987	47.816	-0.8596	0.73891216
1988	48.311	-0.3646	0.13293316
1989	48.758	0.0824	0.00678976
1990	49.164	0.4884	0.23853456
1991	49.548	0.8724	0.76108176
1992	48.915	0.2394	0.05731236
1993	50.315	1.6394	2.68763236
1994	50.768	2.0924	4.37813776
1995	?		
	48.6756		1.81288344

Moving Average

- Moving average as a smoothing process is continued by advancing one period and calculating the next average of three numbers, dropping the first number.



Years	$ (millions)	MA (3)	Error	SSE
1985	46.163			
1986	46.998	46.992	0.006	0.000
1987	47.816	47.71	0.11	0.012
1988	48.311	48.30	0.02	0.000
1989	48.758	48.74	0.01	0.000
1990	49.164	49.16	0.01	0.000
1991	49.548	49.21	0.34	0.115
1992	48.915	49.59	-0.68	0.459
1993	50.315	50.00	0.32	0.100
1994	50.768			
1995	?			
	48.6756	48.712		0.086

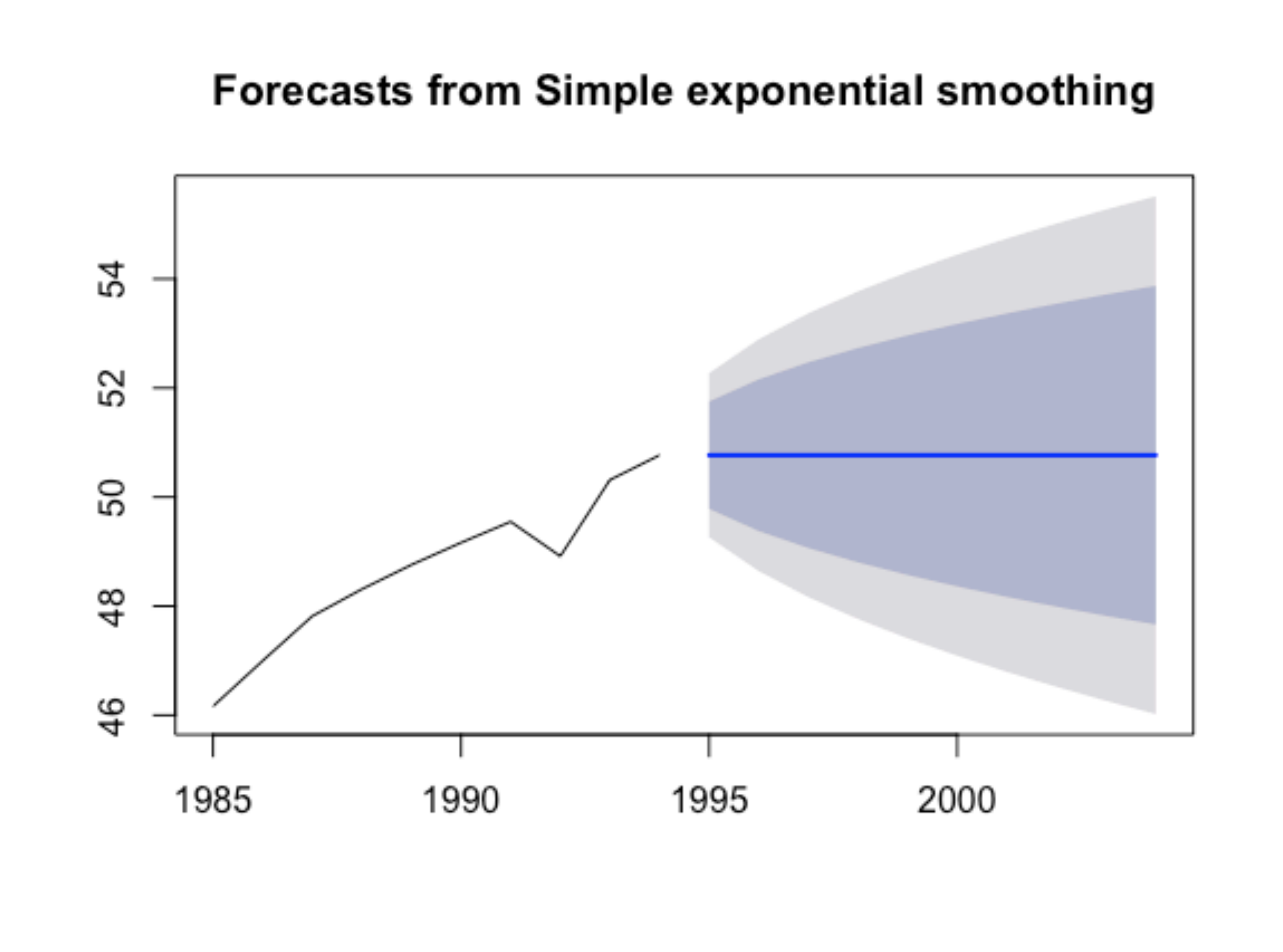
Exponential Smoothing

- This method is suitable for forecasting data with no clear trend or seasonal pattern.
- Forecasts are calculated using weighted averages, where the weights decrease exponentially as observations come from further in the past — the smallest weights are associated with the oldest observations
- The process has to start somewhere, so we let the first fitted value at time 1 be denoted by ℓ_0 (which we will have to estimate)

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \ldots$$

where $0 \le \alpha \le 1$

Exponential Smoothing



Years	$ (millions)	ES	Error	SSE
1985	46.163			
1986	46.998	46.16	0.83	0.70
1987	47.816	47.00	0.82	0.67
1988	48.311	47.82	0.50	0.25
1989	48.758	48.31	0.45	0.20
1990	49.164	48.76	0.41	0.16
1991	49.548	49.16	0.38	0.15
1992	48.915	49.55	-0.63	0.40
1993	50.315	48.92	1.40	1.96
1994	50.768	50.31	0.45	0.21
	48.6756	0.9999		0.52

ARIMA Model

- ARIMA : Auto Regressive Integrated Moving Average
- Factors
  - Number of autoregressive terms (AR)
  - How many non-seasonal differences are needed to achieve stationarity (I)
  - Number of lagged forecast errors in the prediction equation (MA)
- Assumes that the time series data is stationary (i.e trend and seasonality have been removed)

Core Concepts: Autocorrelation

- Tells us how correlated points are with each other, based on how many steps that are separates them.
- Used to determine how past and future data points are related in a time series. It's value n range from -1 to 1

"Autocorrelation is the similarity between observations as a function of the time lag between them."

Core Concepts: Partial Autocorrelation

- Gives the partial correlation of time series with its own lagged values
- It's value n range from -1 to 1

> Partial Autocorrelation is the degree of association between two variables while adjusting the effect of one or more additional variable.

Practice Lab

Build a time series forecast model in using R

Use the following Instructions:

- Get your data in R
- Explore the data
- Build a forecast model
- Validate the model

Recap/Summary

At the end of this Module, you should understand;

- Overview of Time Series
- Core Components of Time Series
- Core Concepts: Autocorrelation & Stationary
- Smoothing Techniques: Moving Average
- Forecast of US Electricity Price in R

Suggested Material

- https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775
- http://www.stat.columbia.edu/~rdavis/lectures/Session6.pdf
- https://otexts.com/fpp2/
- https://www.youtube.com/watch?v=gj4L2isnOf8
- https://www.youtube.com/watch?v=Y5T3ZEMZZKs