Road map!

- Module 1- Demystifying Timeseries Data and Modeling
- Module 2- Setting up Deep Forecasting Environment
- Module 3- Exponential Smoothing
- Module 4- ARIMA models
- Module 5- Machine Learning for Time series Forecasting
- Module 6- Deep Neural Networks
- Module 7- Deep Sequence Modeling (RNN, LSTM)
- Module 8- Prophet (Forecasting at Scale)

Module 1

Demystifying Timeseries Data and Modeling

TS Basics

Modeling

Forecasting

Key topics

1. Timeseries Basics

- What is sequence data?
- Time series tasks
- Trend, Seasonality, Residual
- ACF and PACF
- Stationarity

3. Modeling

- White Noise
- Econometrics
- Machine Learning
- Deep Learning

2. Forecasting strategies

- One-step, Multi-step, Multi-output
- Univariate vs Multivariate
- Benchmarks

4. Can we beat wall street?

- Is stock price Random Walk? (EMH)
- Does Uni-variate Model work?
- Can Multi-variate Model help?
- So What?

Pedram Jahangiry

Module 1- Part 1

1. Timeseries Basics

   - What is sequence data?
   - Time series tasks
   - Trend, Seasonality, Residual
   - ACF and PACF
   - Stationarity

2. Forecasting strategies

   - One-step, Multi-step, Multi-output
   - Univariate vs Multivariate
   - Benchmarks

3. Modeling

   - White Noise
   - Econometrics
   - Machine Learning
   - Deep Learning

4. Can we beat wall street?

   - Is stock price Random Walk? (EMH)
   - Does Uni-variate Model work?
   - Can Multi-variate Models help?
   - So What?

Pedram Jahangiry

What is Sequence Data?

Sequence data refers to any data that has a specific **order** or sequence to it!

Sequence Data

- Time Series Data
  - Regular TS
  - Irregular TS
- Text Data
- Audio Data
- Video Data

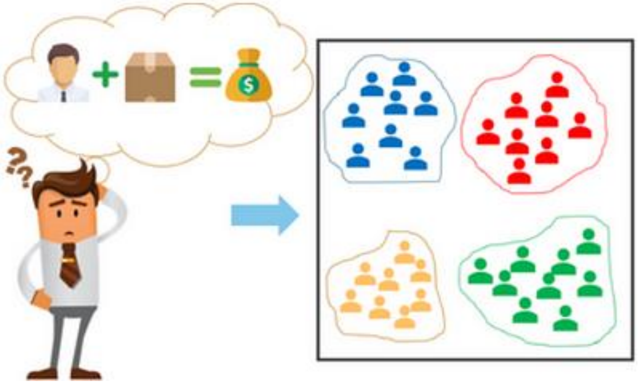
WE ARE HERE

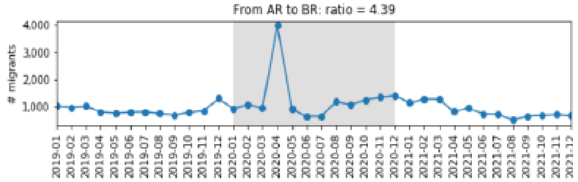
Pedram Jahangiry

Time series Tasks

TS tasks

- Forecasting
  - Qualitative
  - Quantitative (WE ARE HERE)
- Classification
- Clustering
- Anomaly/Event Detection





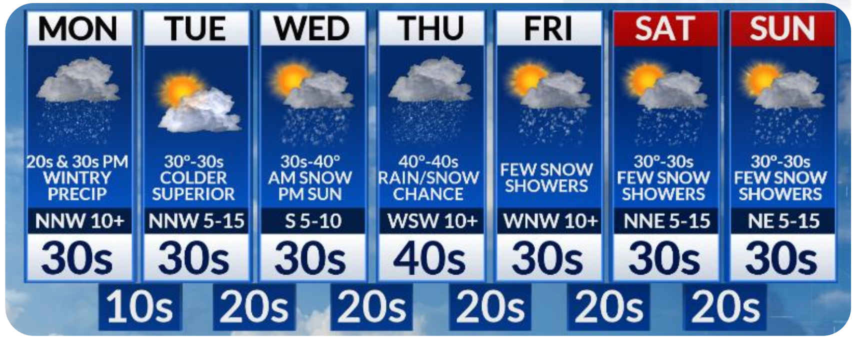
Assumption: **past patterns** will **continue**

Pedram Jahangiry

More Forecasting examples

- Which of the following examples are easier to forecast?

1. Time of sunset this day next month
2. Apple stock price tomorrow
3. Airline ticket demand/price next year
4. New car model sales in the first quarter
5. Monthly rainfall in Utah next winter



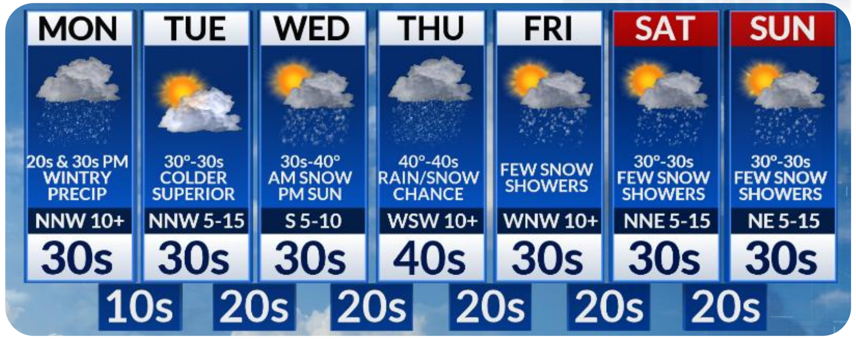
- What impacts Forecastability?
  - Data Availability
  - How similar the future is to the past!
  - Good understanding of the underlying factors

Pedram Jahangiry

More Forecasting examples

- Which of the following examples are easier to forecast?

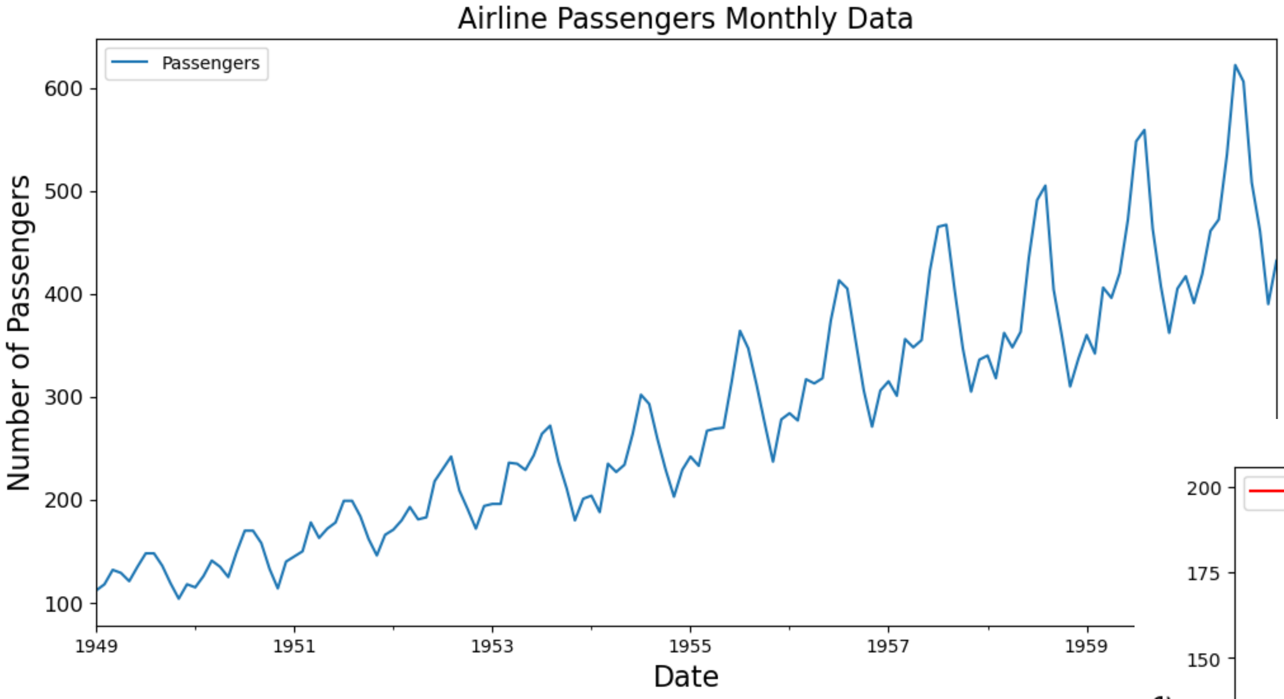
1. Time of sunset this day next month
2. Apple stock price tomorrow
3. Airline ticket demand/price next year
4. New car model sales in the first quarter
5. Monthly rainfall in Utah next winter

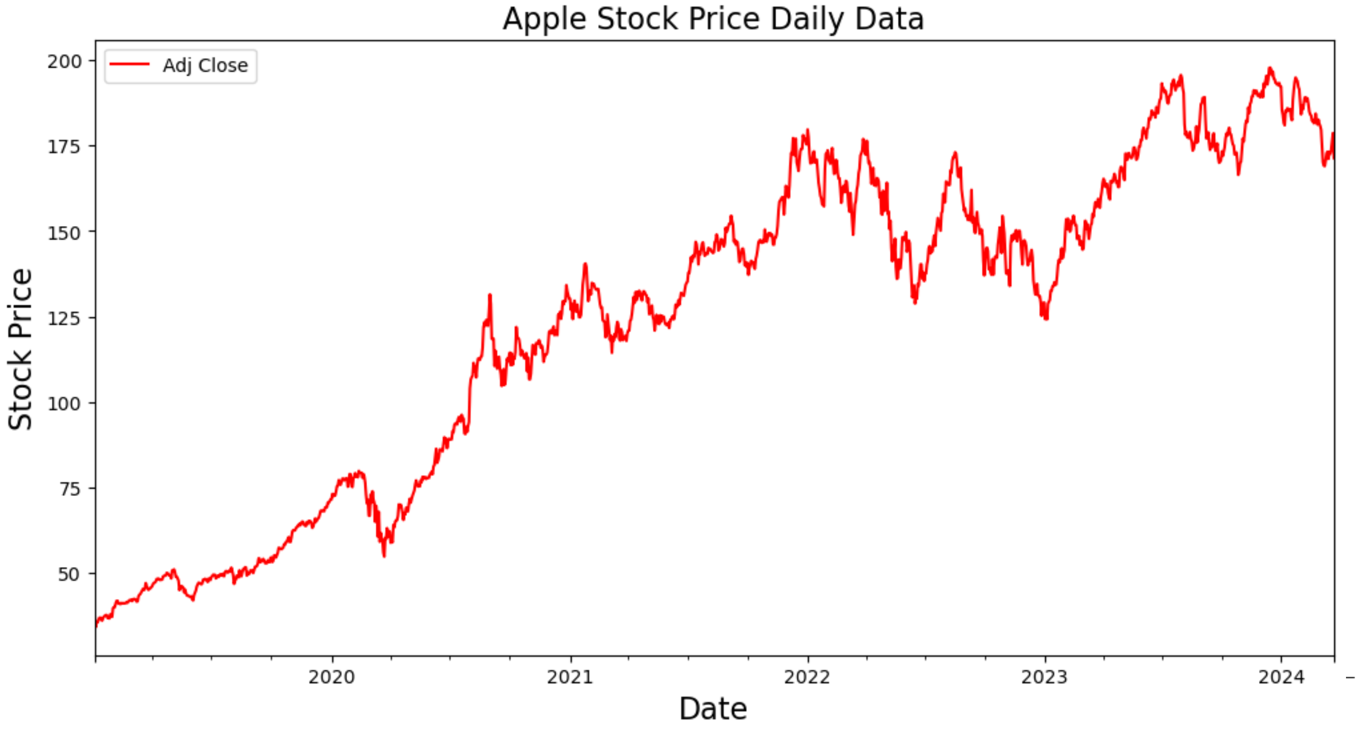


- What impacts Forecastability?
  - Data Availability
  - How similar the future is to the past!
  - Good understanding of the underlying factors

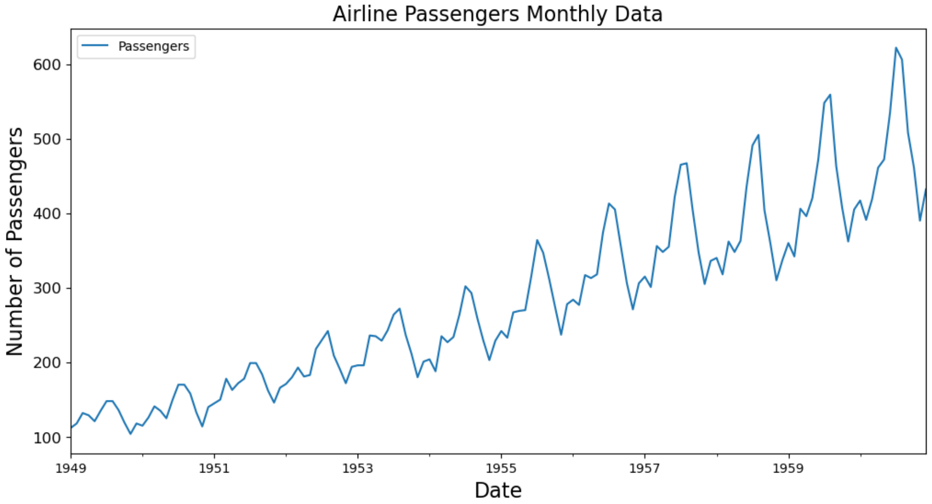
Pedram Jahangiry

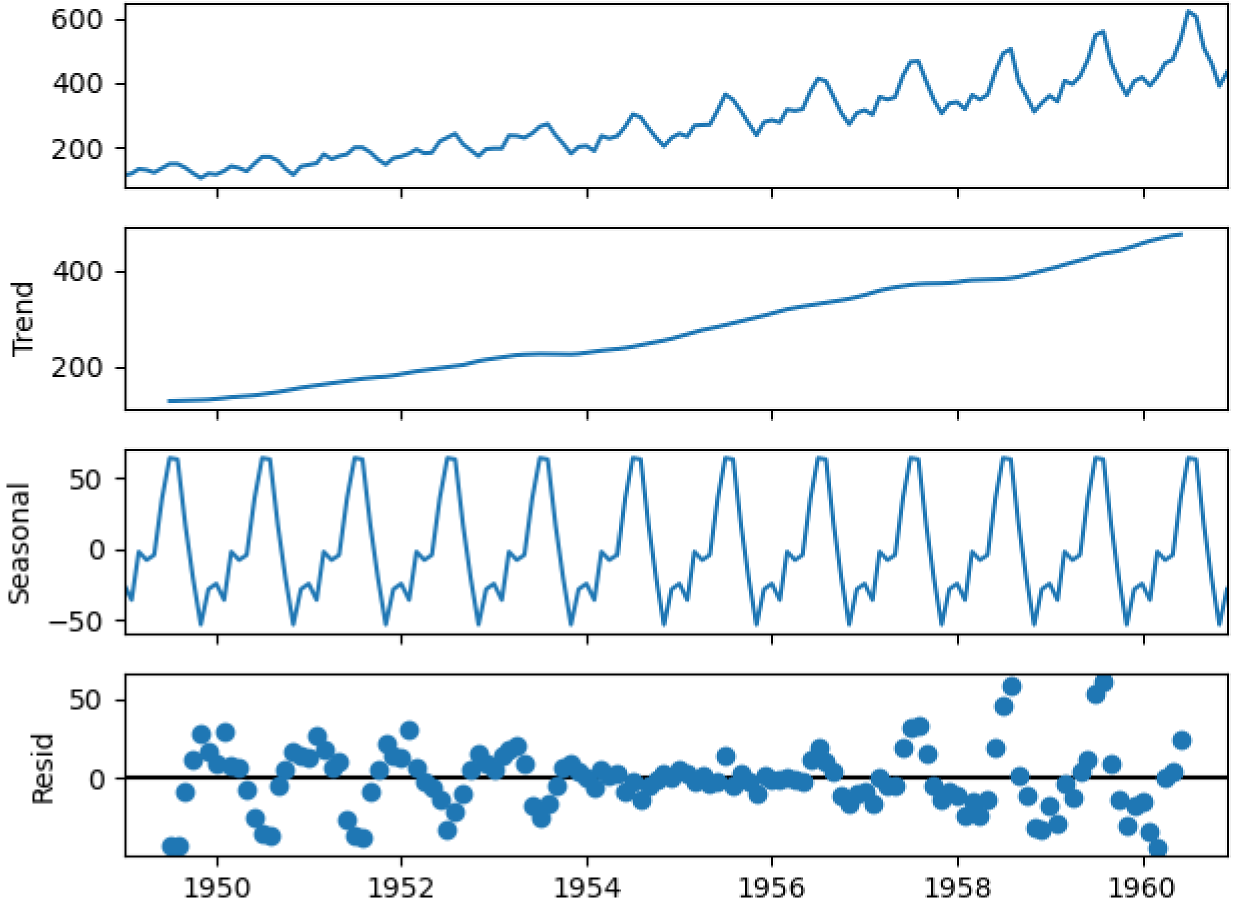
Quantitative Forecasting



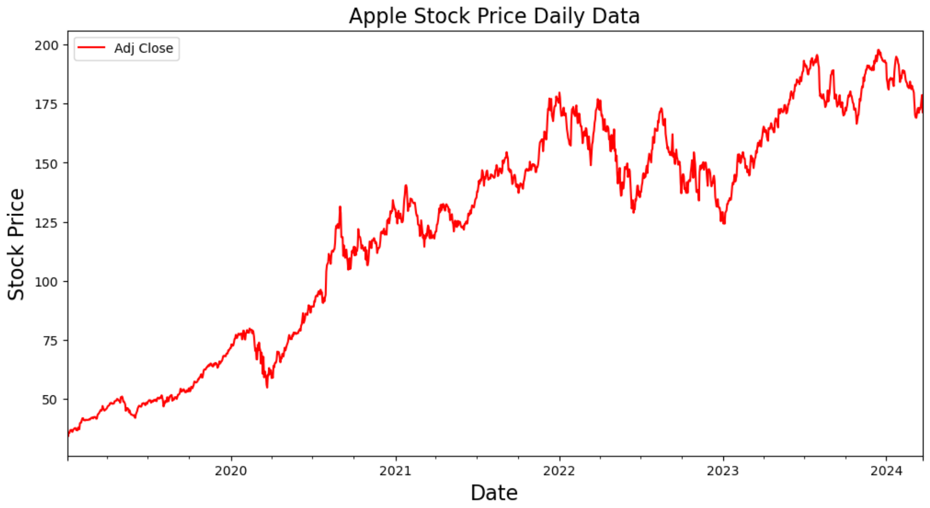


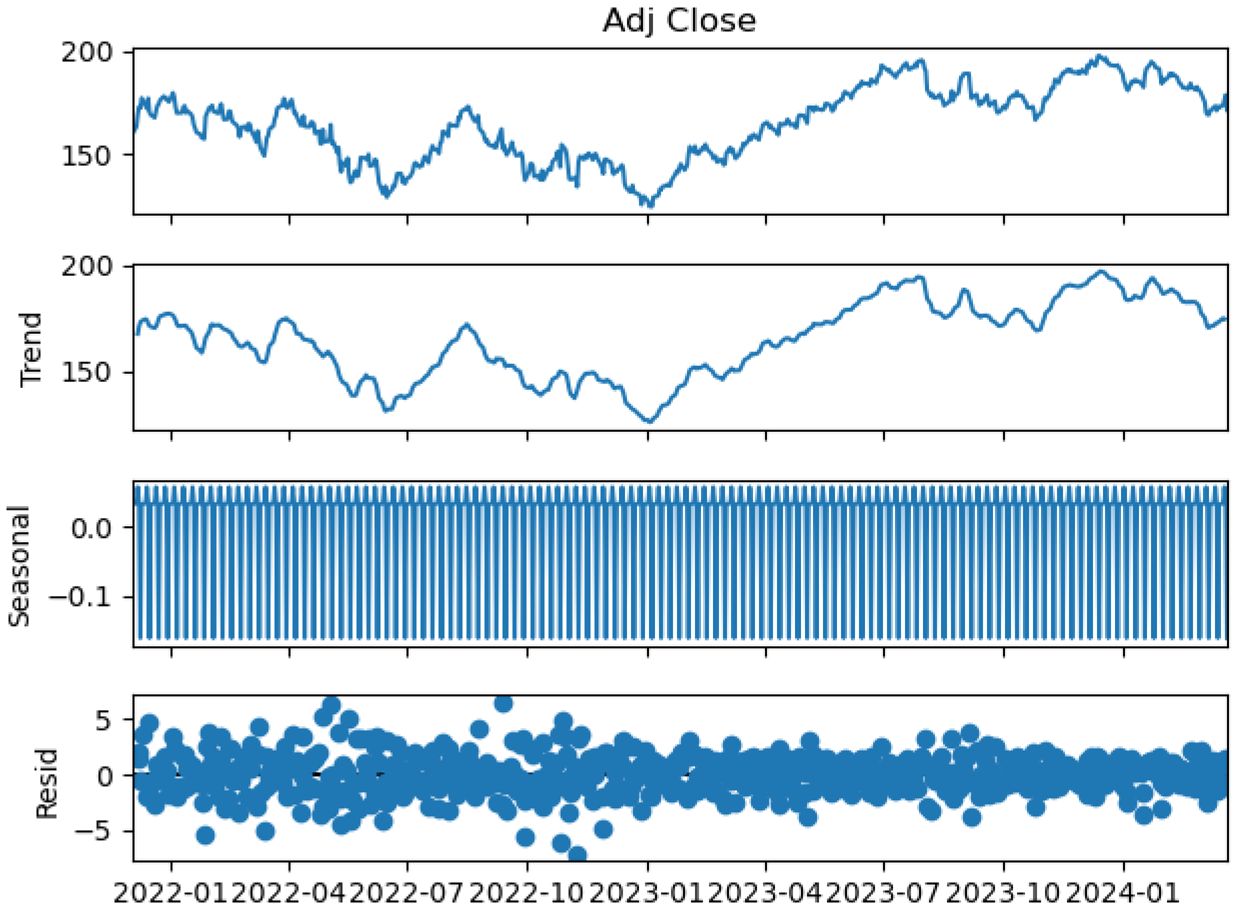
TS Decomposition: Trend, Seasonality, Residual





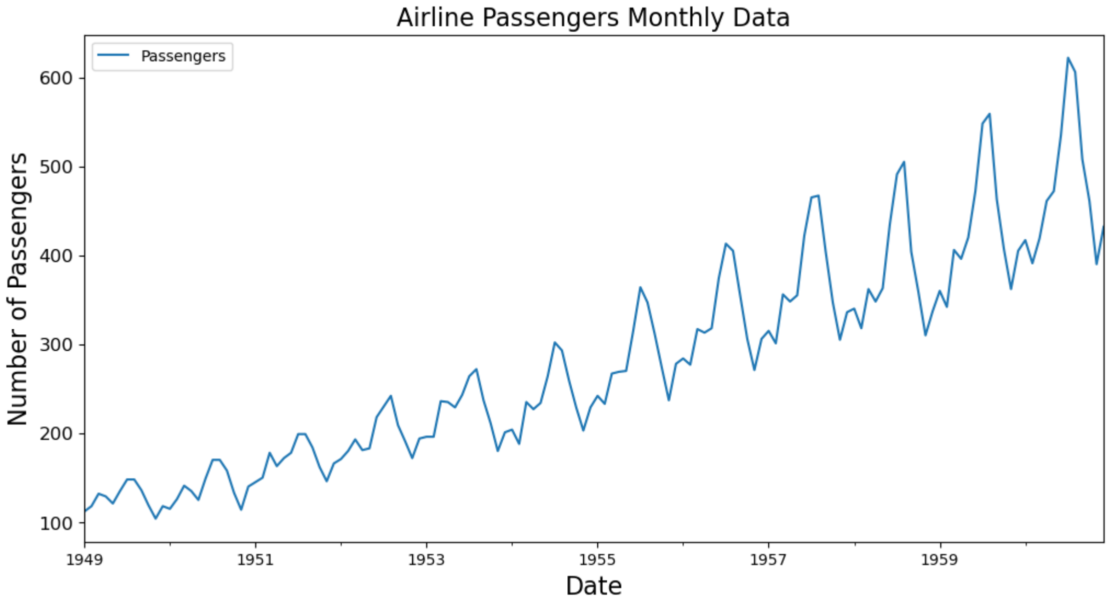
TS Decomposition: Trend, Seasonality, Residual

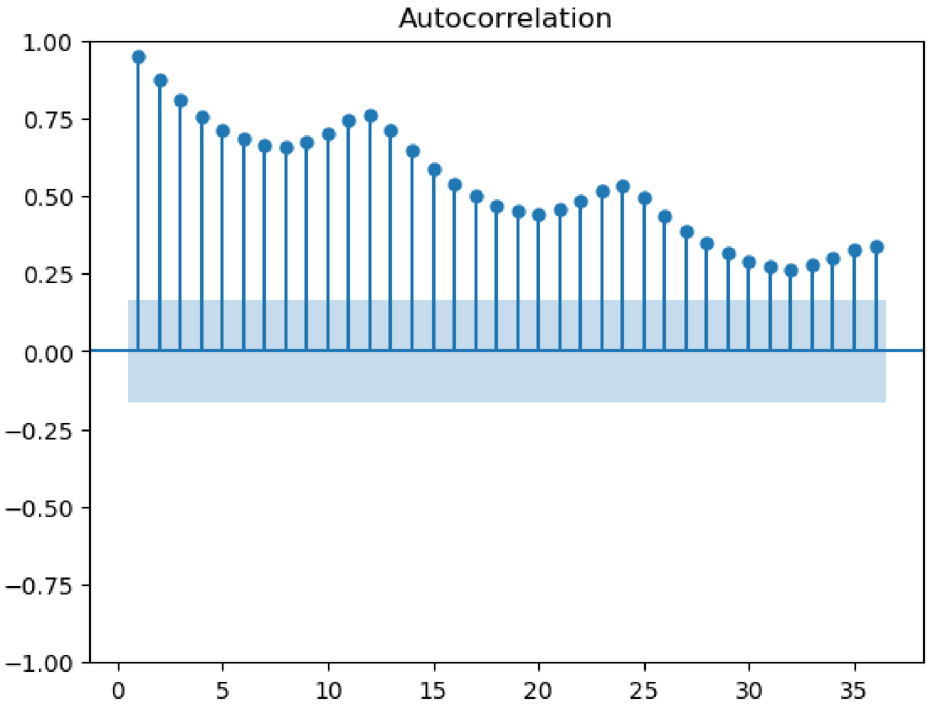




ACF: Autocorrelation Function

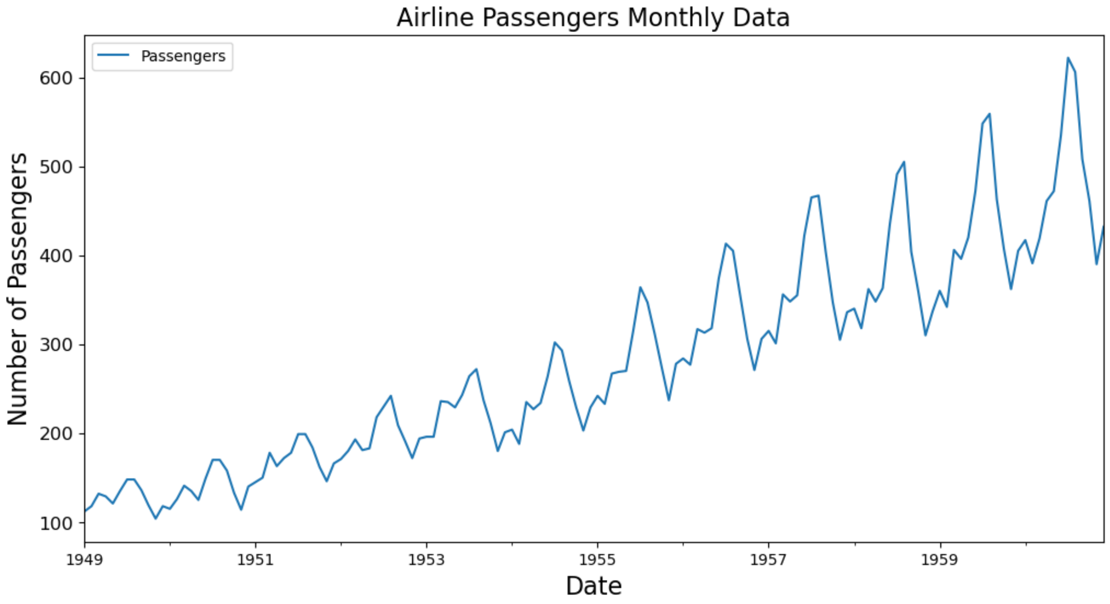
- ACF is a statistical tool that can be used to measure the autocorrelation of a time series.
- Autocorrelation (Serial correlation) is a measure of the correlation between a time series and a lagged version of itself.
- ACF shows the degree to which the past values are predictive of future values.

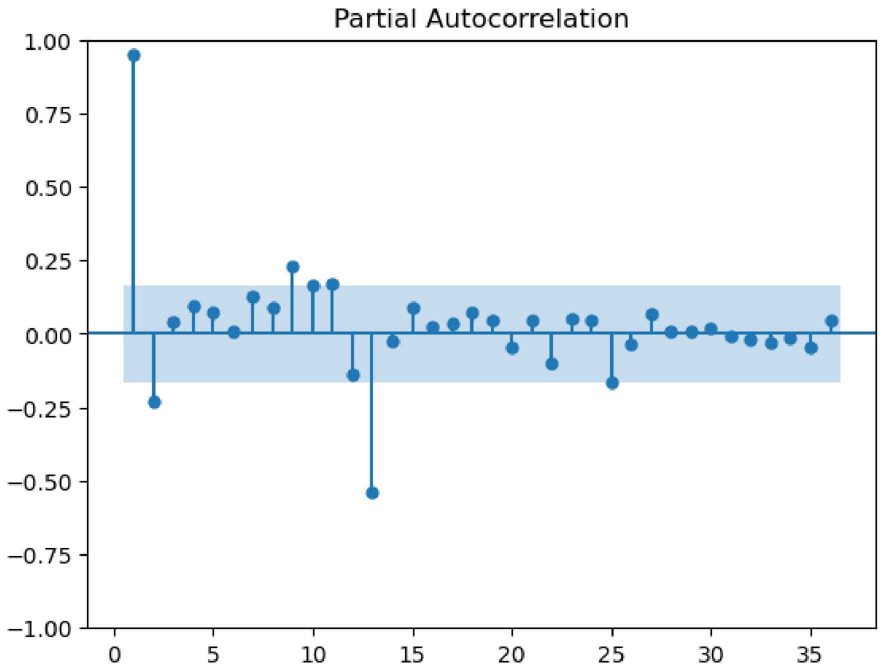




PACF: Partial Autocorrelation Function

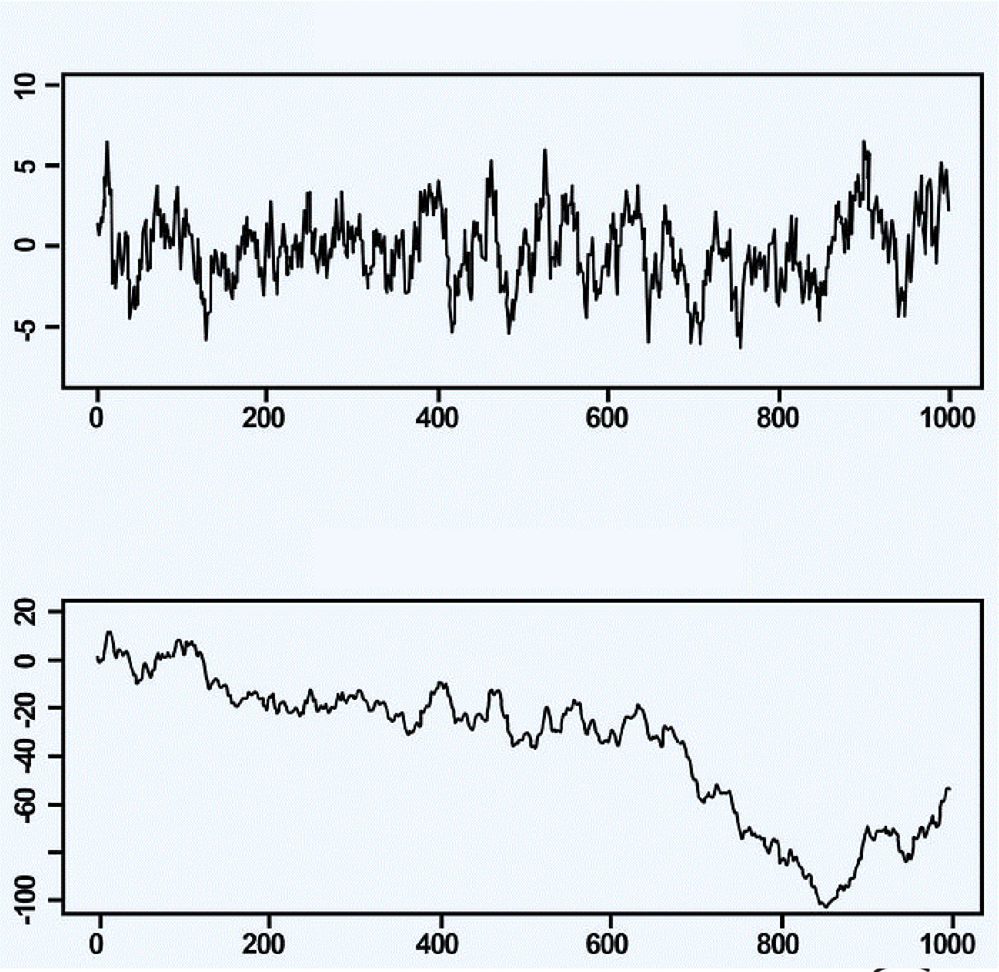
- PACF measures the direct correlation between time series and its lag, controlling for intermediate lags.
- y_t and y_{t-2} might be correlated, simply because they are both connected to y_{t-1}





Stationarity

- Stationary vs Non-Stationary Data. What makes a data set Stationary?
- In a stationary timeseries, the statistical properties do not depend on the time
- Predictability (long-horizon): Stationary time series are easier to forecast because you can assume that future statistical properties will not change.
- This doesn’t mean we cannot forecast non-stationary data!



Pedram Jahangiry

Differencing

- Differencing: Computing the difference between consecutive observations.
- Differencing helps to stabilize the mean of a time series by removing changes in the level

Original data

Time1	10
Time2	12
Time3	8
Time4	14
Time5	7

1st differenced

Time1	NA
Time2	2
Time3	-4
Time4	6
Time5	-7

2nd differenced

Time1	NA
Time2	NA
Time3	-6
Time4	10
Time5	-13

Module 1- Part 1

1. Timeseries Basics

   - What is sequence data?
   - Time series tasks
   - Trend, Seasonality, Residual
   - ACF and PACF
   - Stationarity

2. Forecasting strategies

   - One-step, Multi-step, Multi-output
   - Univariate vs Multivariate
   - Benchmarks

3. Modeling

   - White Noise
   - Econometrics
   - Machine Learning
   - Deep Learning

4. Can we beat wall street?

   - Is stock price Random Walk? (EMH)
   - Does Uni-variate Model work?
   - Can Multi-variate Models help?
   - So What?

Pedram Jahangiry

Module 1- Part 2

1. Timeseries Basics

- What is sequence data?
- Time series tasks
- Trend, Seasonality, Residual
- ACF and PACF
- Stationarity

2. Forecasting strategies

- One-step, Multi-step, Multi-output
- Univariate vs Multivariate
- Benchmarks

3. Modeling

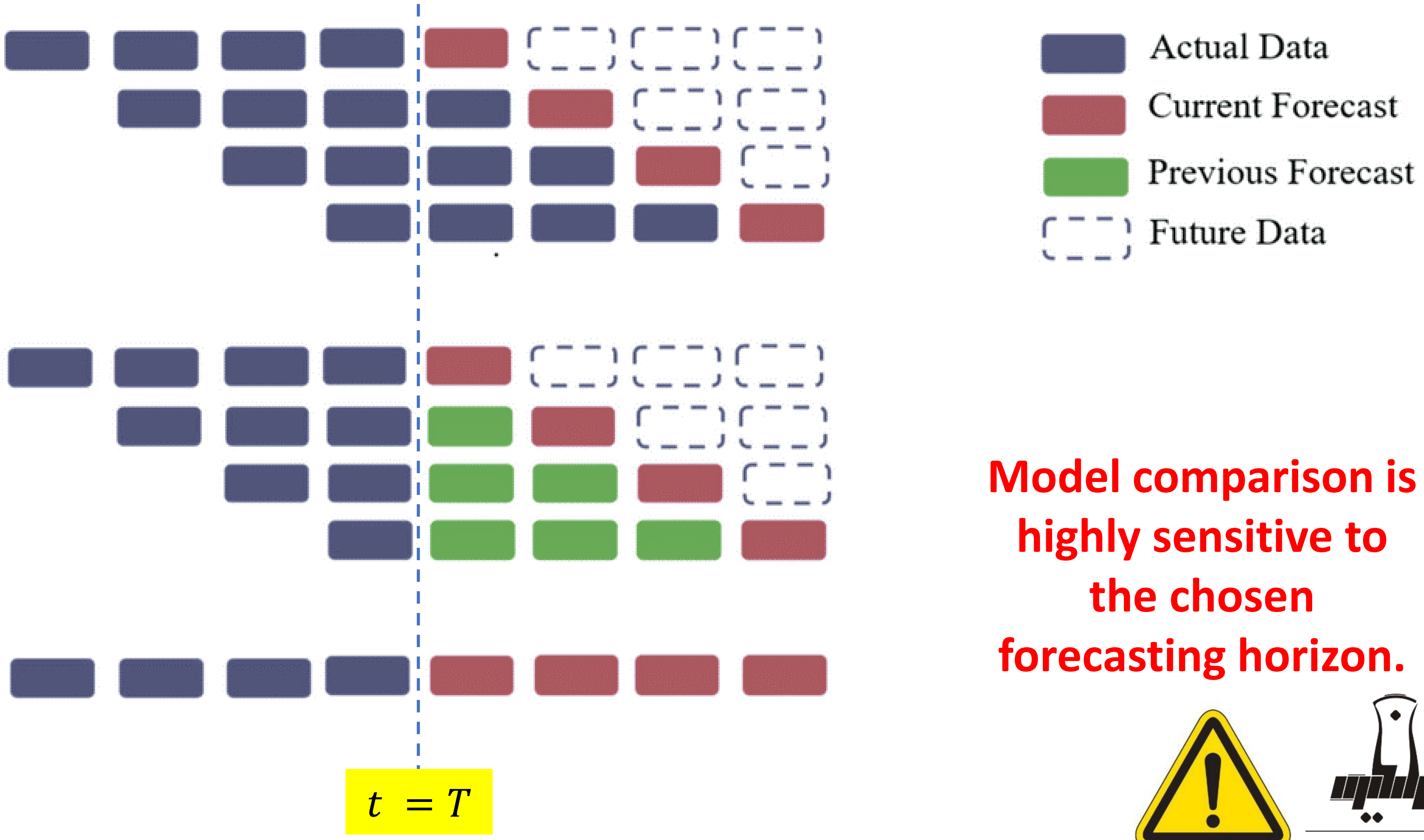
- White Noise
- Econometrics
- Machine Learning
- Deep Learning

4. Can we beat wall street?

- Is stock price Random Walk? (EMH)
- Does Uni-variate Model work?
- Can Multi-variate Models help?
- So What?

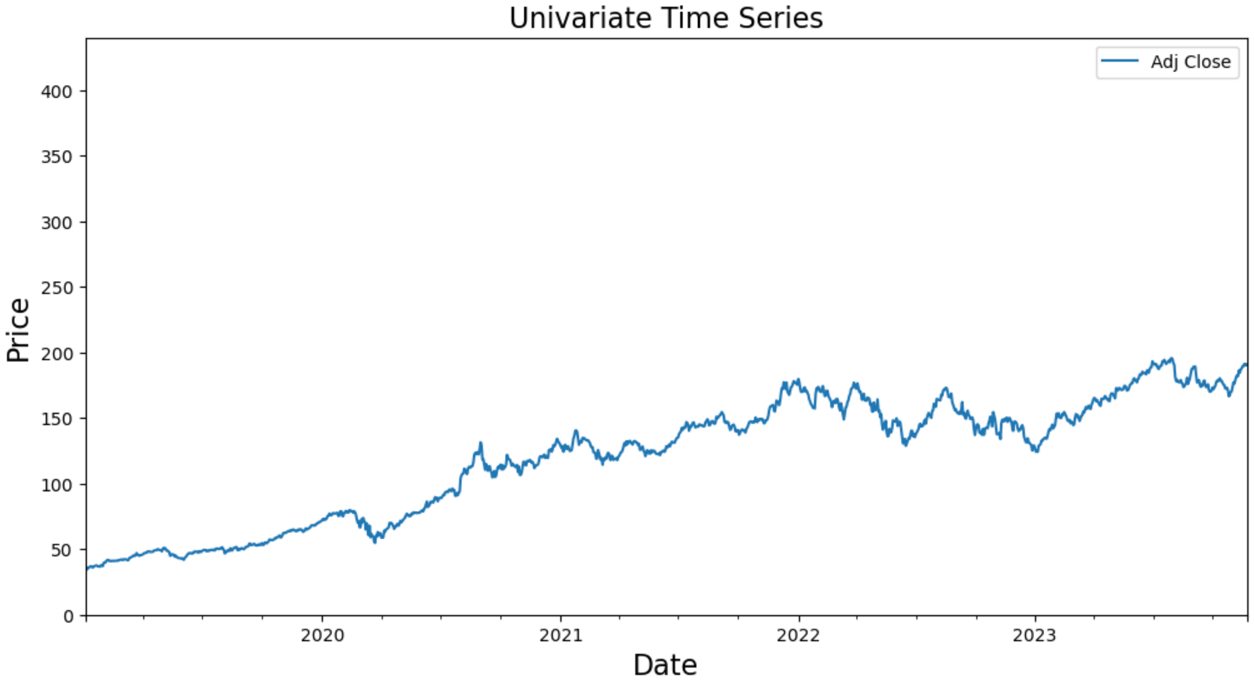
Pedram Jahangiry

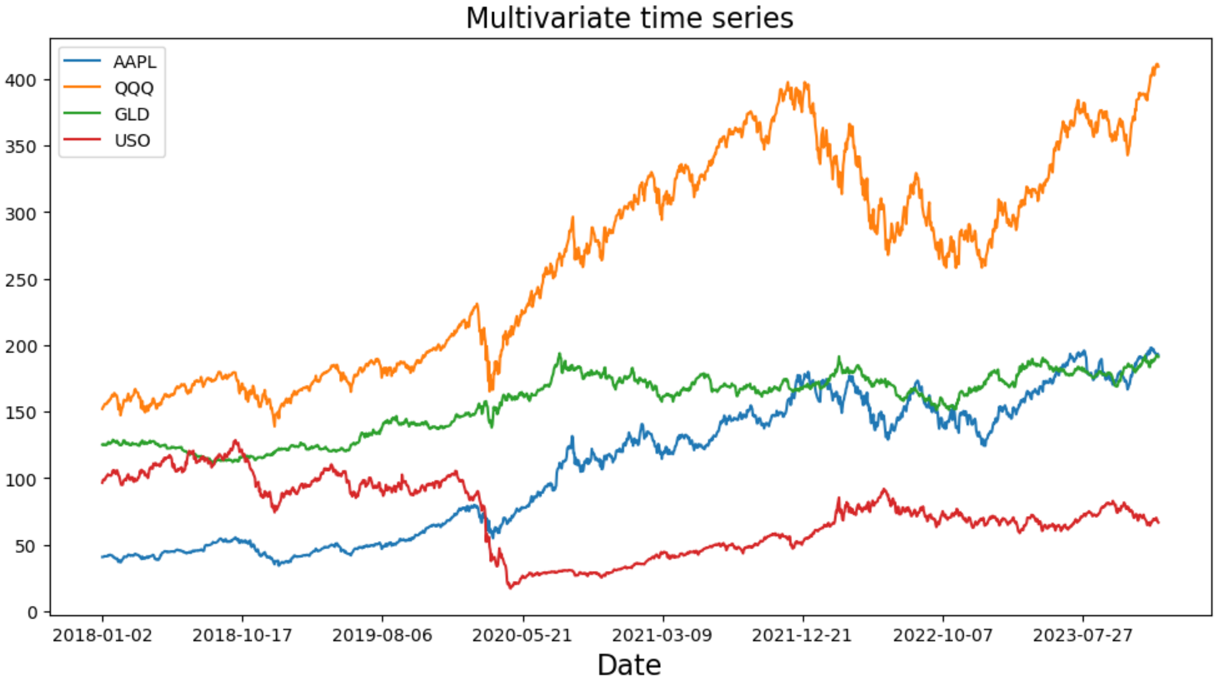
One-step, Multi-step, Multi-output Forecasts



Model comparison is highly sensitive to the chosen forecasting horizon.

Univariate vs Multivariate





- Using a single variable's historical data to predict future values for that same variable

- Incorporates multiple variables to forecast future values, considering the relationship between them.
- **Cannot** be used for multi-step forecasting!
- Less stable in LR due to increased forecast error.

Forecasting Benchmarks

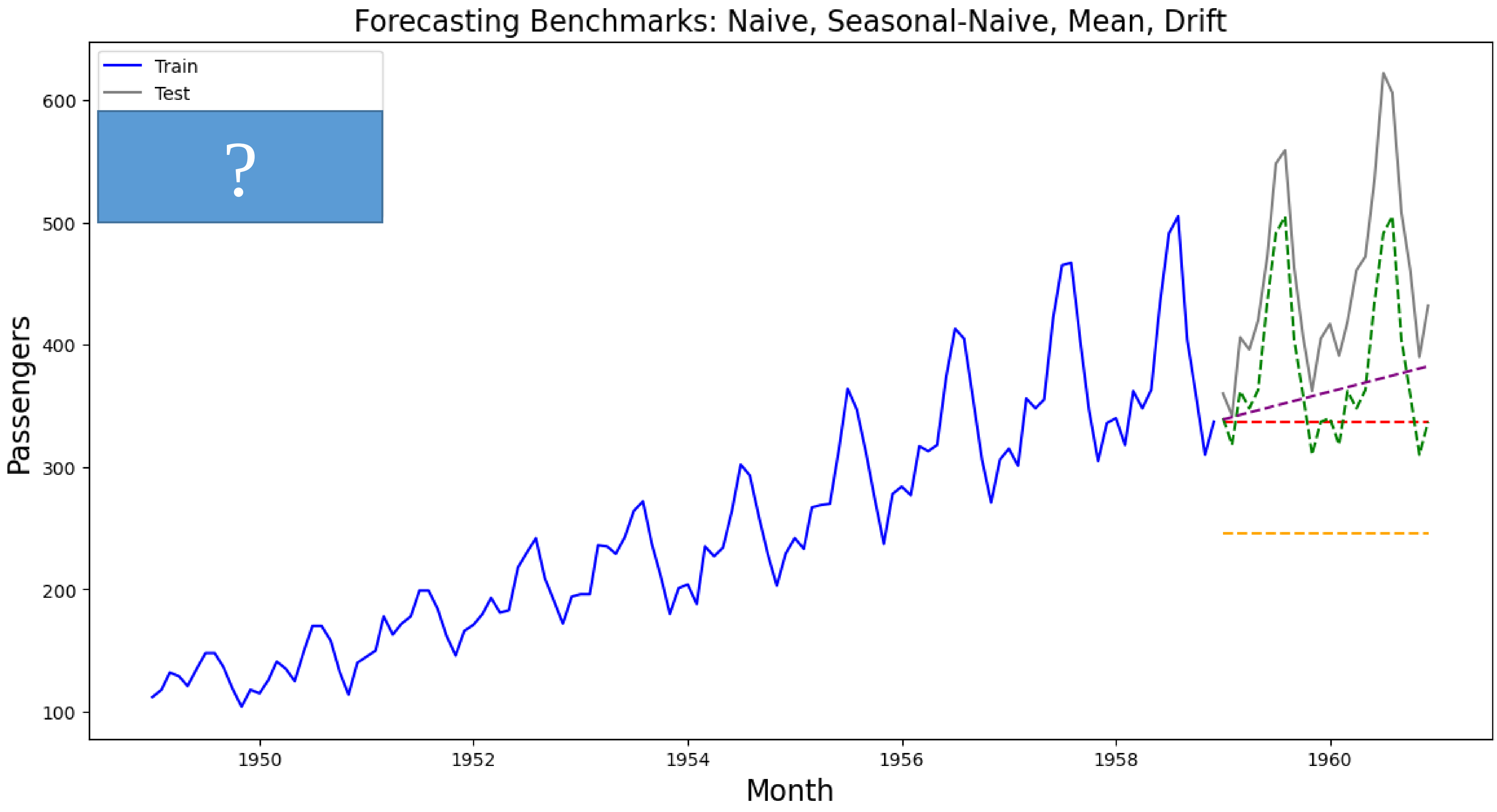
- Some forecasting methods are extremely simple and surprisingly effective.

1. Mean method: Predicts all future values to be the average of the historical data.
2. Naïve method: Predicts the next value to be the current value
3. Seasonal Naïve: Predicts the next value by using the same value from the previous season
4. Drift method: Predicts future values by extending a linear trend fitted to the historical data.

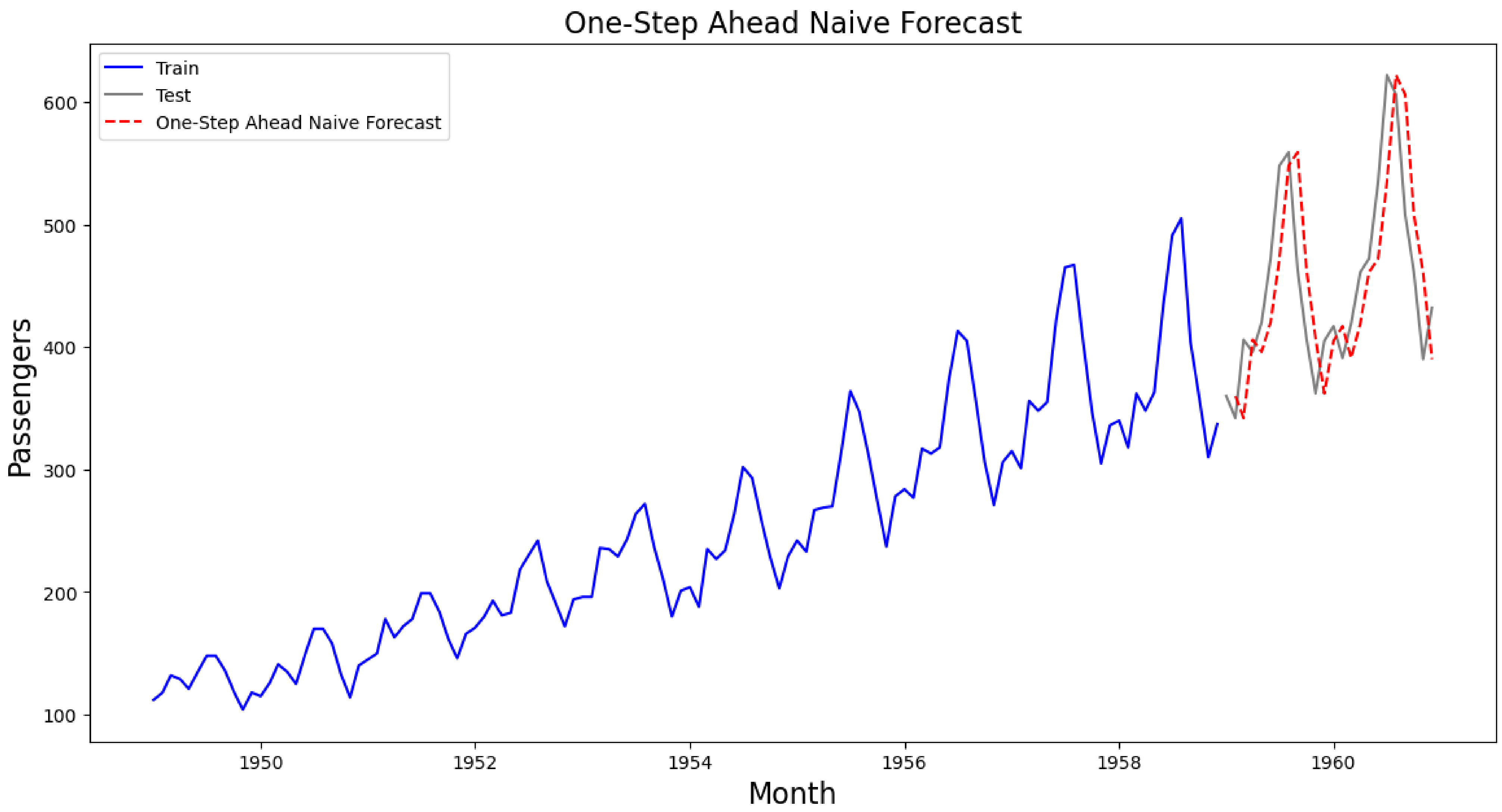
- These methods will serve as benchmarks rather than the method of choice.
- If our model cannot beat the benchmark, it is not worth considering!!!

Pedram Jahangiry

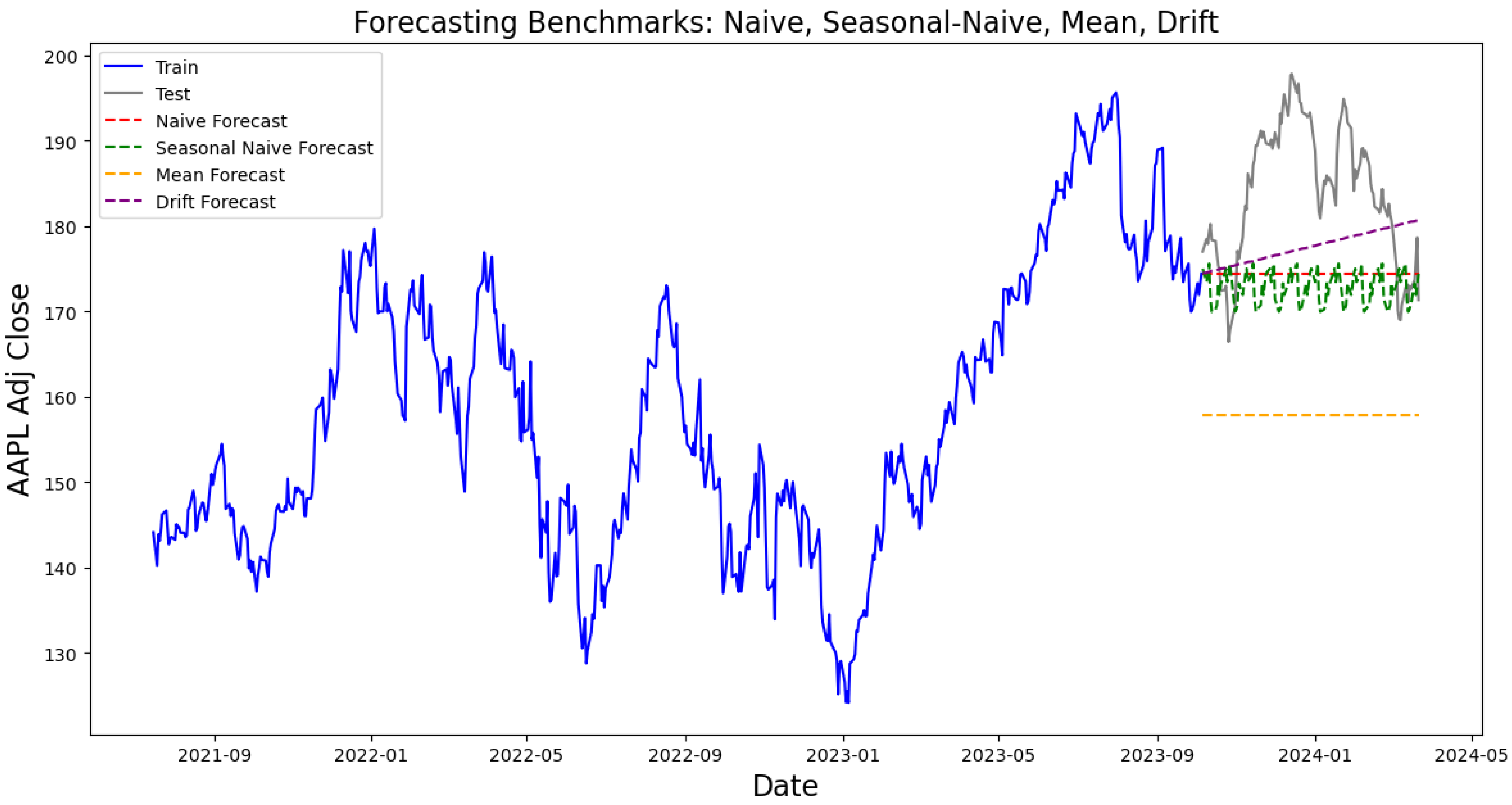
Forecasting Benchmarks (Multi-Horizon)



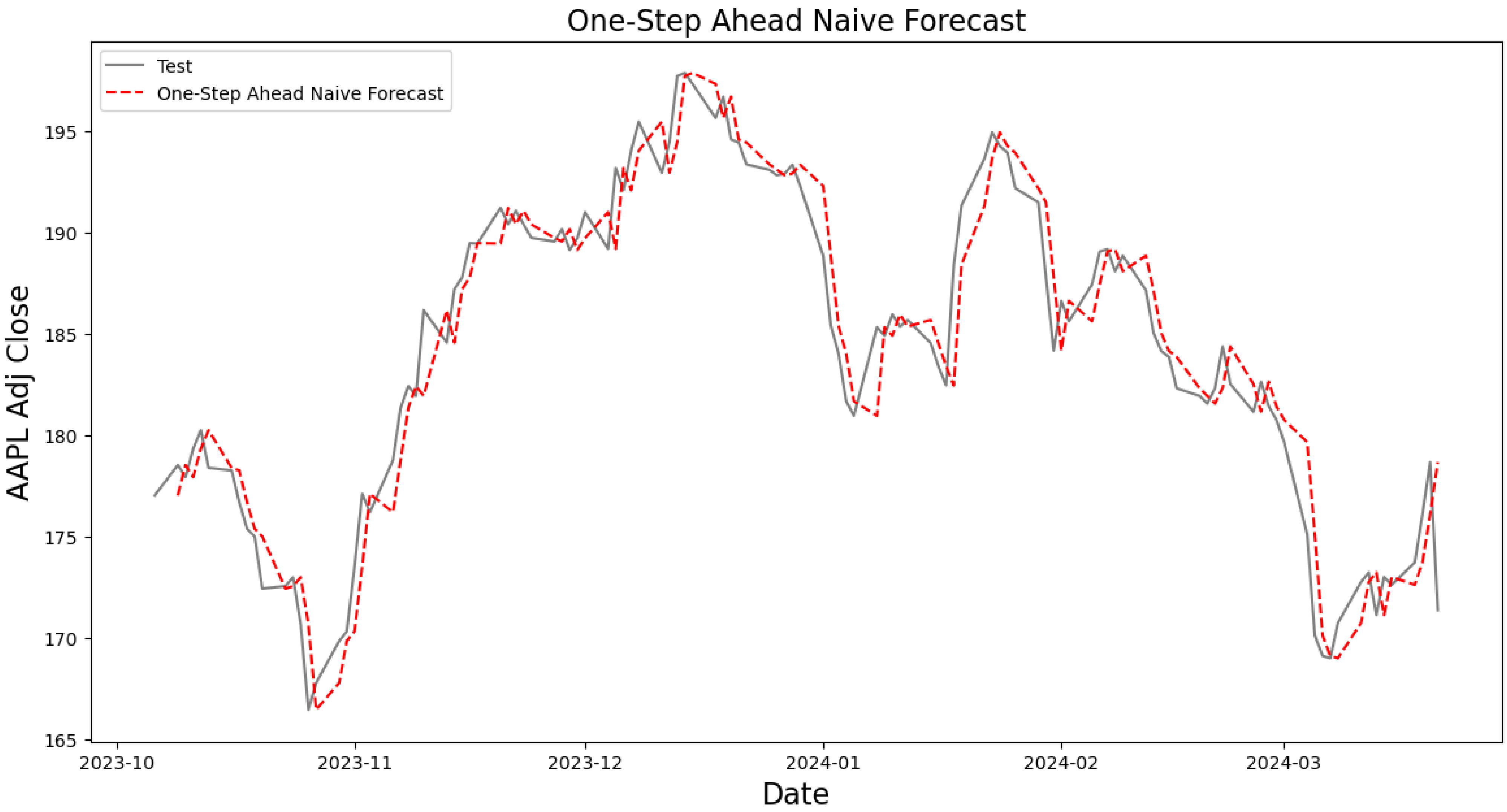
Forecasting Benchmarks (One-step)



Forecasting Benchmarks



Forecasting Benchmarks



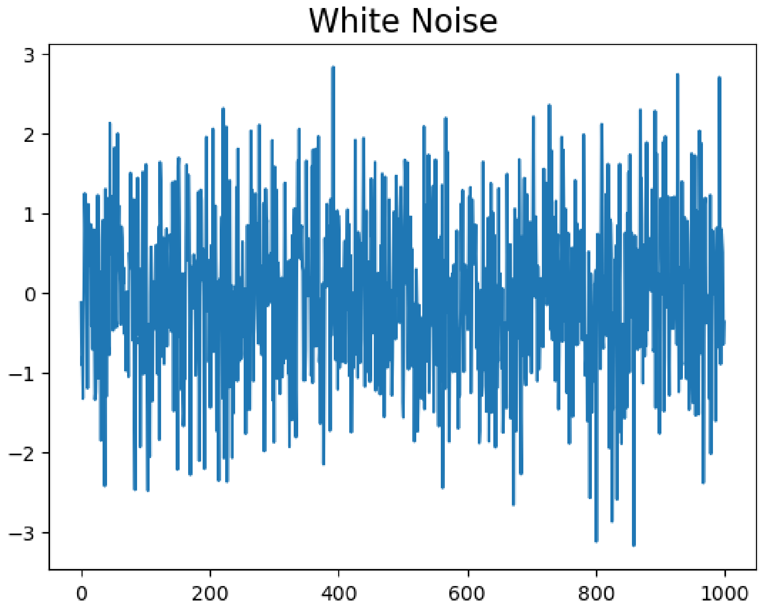
Module 1- Part 3

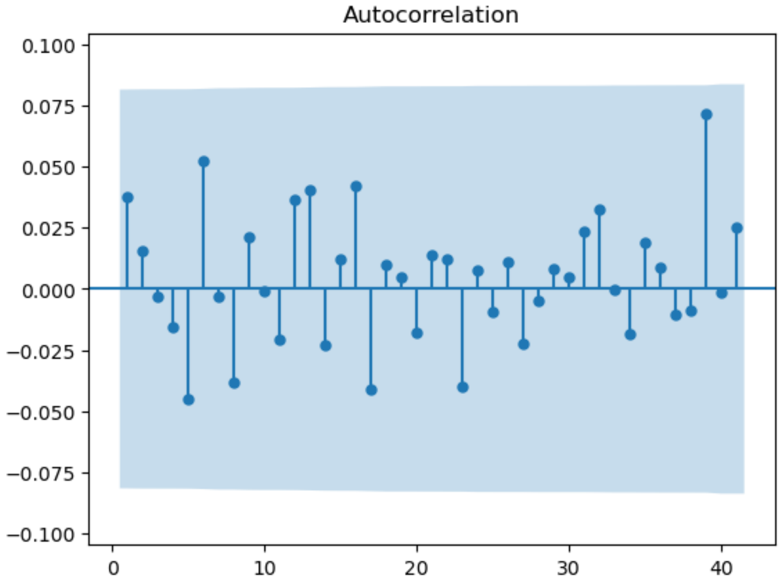
1. Timeseries Basics - What is sequence data? - Time series tasks - Trend, Seasonality, Residual - ACF and PACF - Stationarity	3. Modeling - White Noise - Econometrics - Machine Learning - Deep Learning
2. Forecasting strategies - One-step, Multi-step, Multi-output - Univariate vs Multivariate - Benchmarks	4. Can we beat wall street? - Is stock price Random Walk? (EMH) - Does Uni-variate Model work? - Can Multi-variate Models help? - So What?

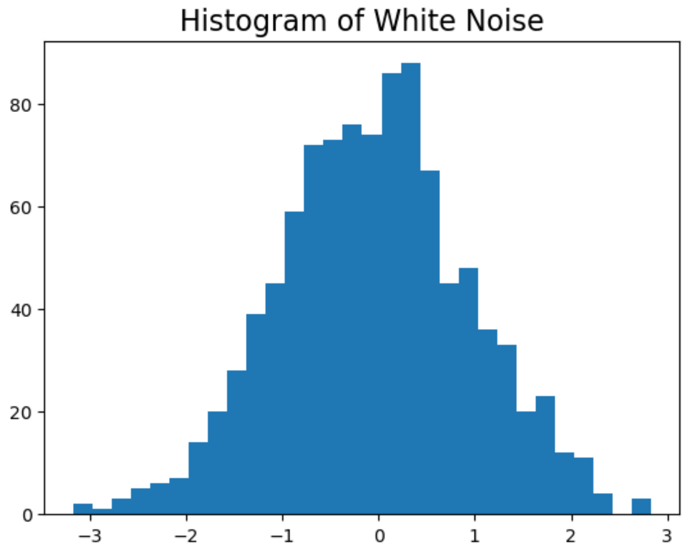
Pedram Jahangiry

The Importance of White Noise in Modeling

- White noise refers to a random process with:
- Zero mean, Constant variance and No autocorrelation.
- Role in timeseries modeling: Diagnostic tool! Indicator of a good fit!

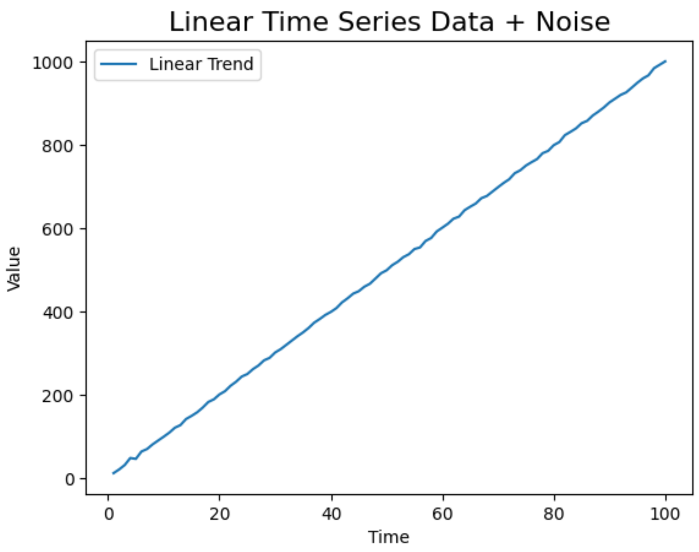


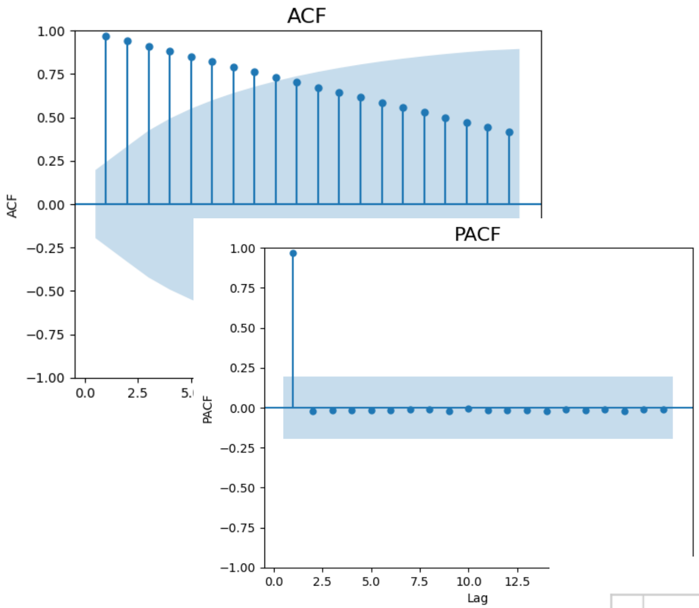


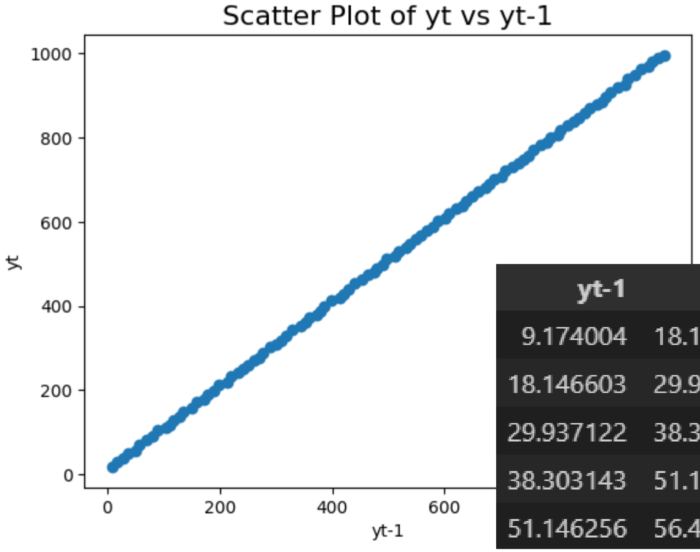


Pedram Jahangiry

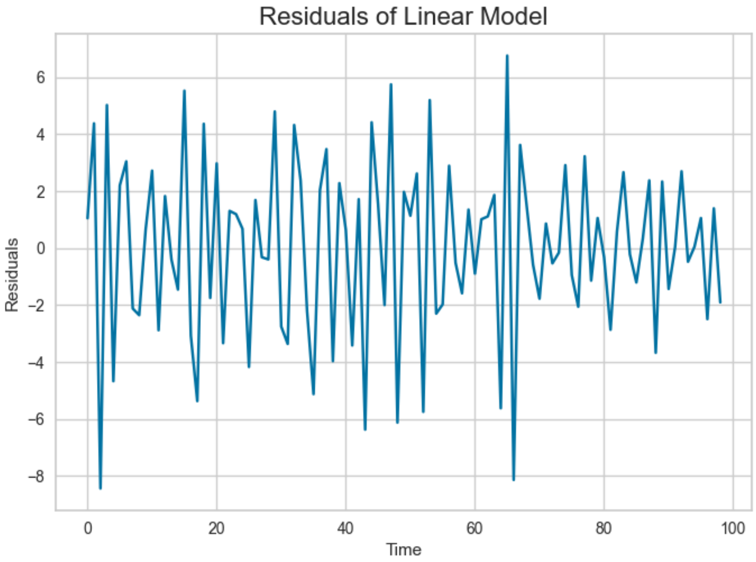
Linear vs Non-Linear (data or model?)







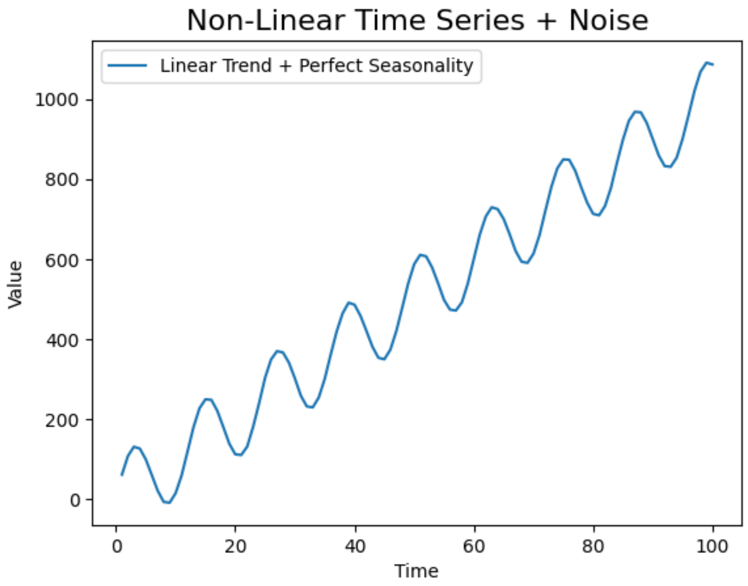
yt-1	yt
9.174004	18.146603
18.146603	29.937122
29.937122	38.303143
38.303143	51.146256
51.146256	56.428267

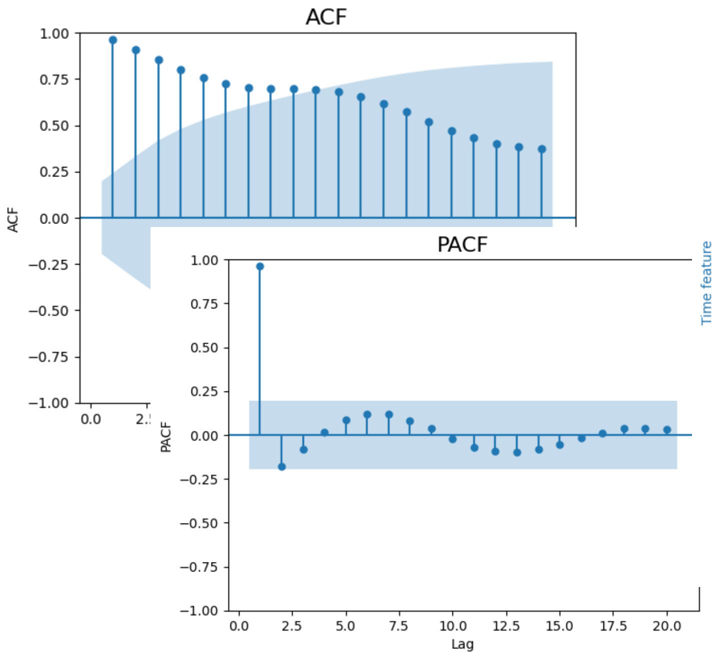


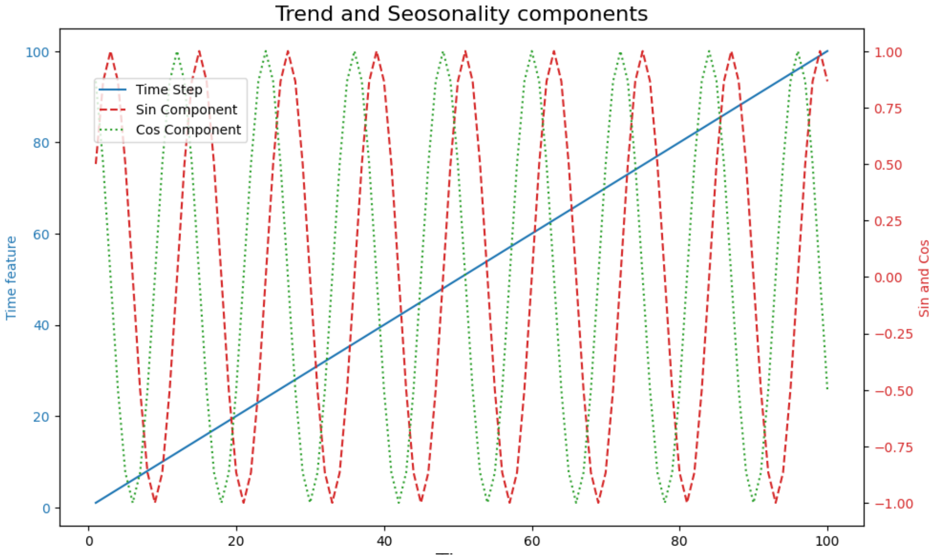
- Can you think of a simple linear model?
- True model: $y_t = 10 + y_{t-1} + \epsilon_t$

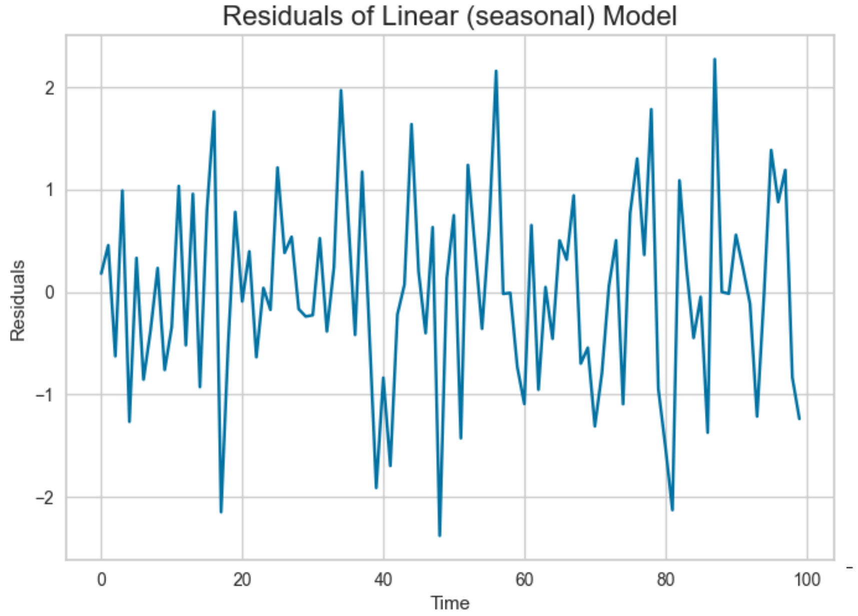
Pedram Jahangiry

Linear vs Non-Linear (data or model?)



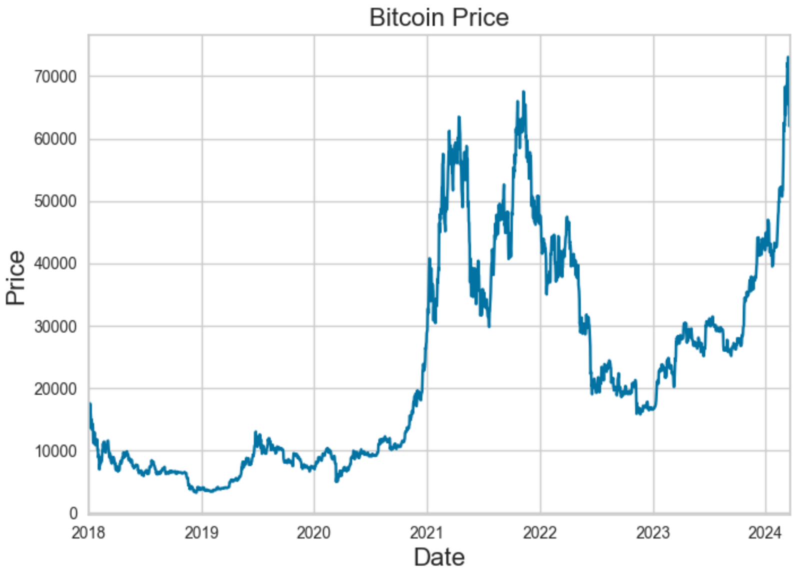


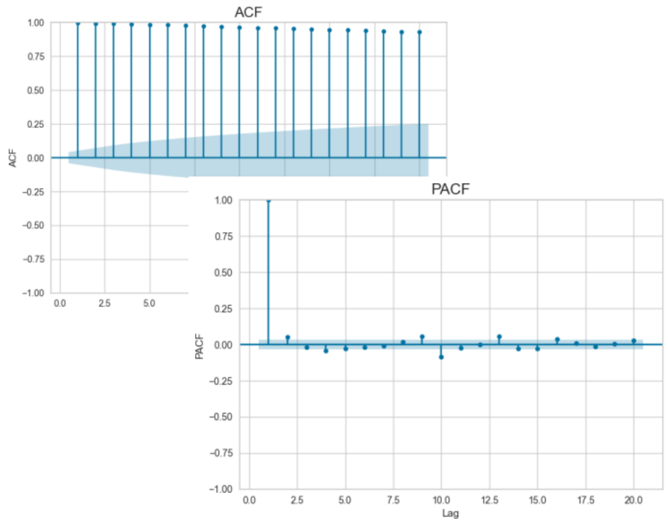


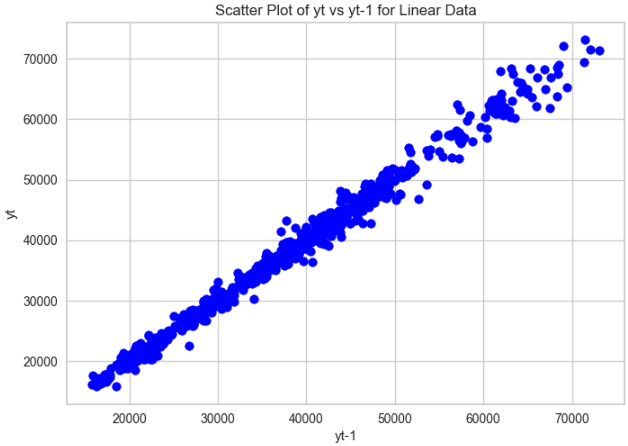


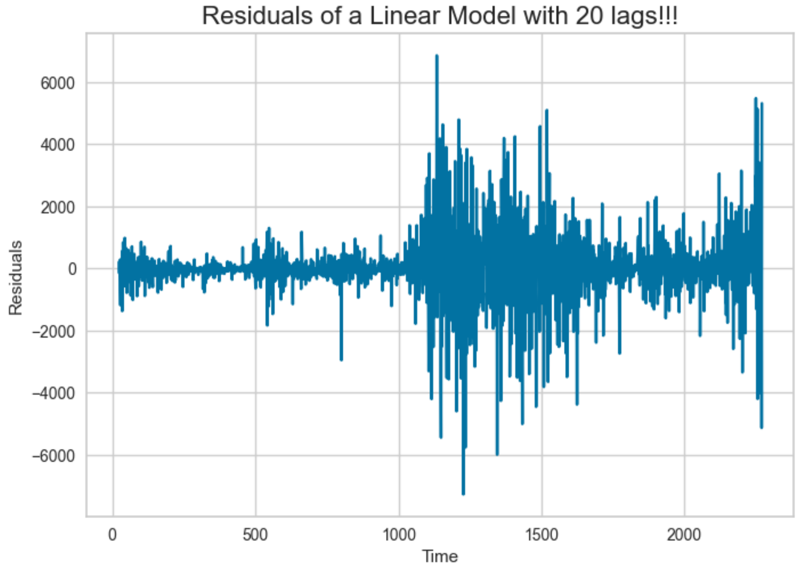
- Can you think of a simple linear model?
- True model: $y_t = 10t + 100\sin(..) + \epsilon_t$

Linear vs Non-Linear (data or model?)









- Can you think of a simple linear model?
- True model: $y_t = I\ wish\ :)$

Basic steps in a forecasting task

Step1: Problem definition

- Forecasting type and horizon (one-step, multi-step, multi-output forecasts), …

Step 2: Data Collection

- Time horizon, structural changes, data type, …

Step 3: Exploratory Analysis

- Trend, seasonality, outliers, …

Step 4: Model Selection and Training

- Traditional **vs** machine learning **vs** deep learning

Step 5: Model Evaluation and Comparison

- MSE, RMSE, MAE, MAPE, …

Modeling, a simple example

- Quantifying airline passengers!

- Let's build a model:
    - Cross sectional vs timeseries?
    - Univariate vs Multivariate?

$$air\ passengers = \beta_0 + \beta_1 gdp + \beta_2 unemp + \beta_3 saving + \cdots + u$$

✓ Can you interpret this model?

✓ Can you make predictions using your model?

✓ Can you make forecasts into the future? What are the challenges?

Pedram Jahangiry

Modeling, a simple example (cont'd)

- Focusing on Forecasting for one country:

$$air\ passengers_t = \beta_0 + \beta_1 gdp_t + \beta_2 unemp_t + \beta_3 saving_t + \cdots + u_t$$

- This explanatory model is contemporaneous (static). How can we make it a dynamic forecasting model?

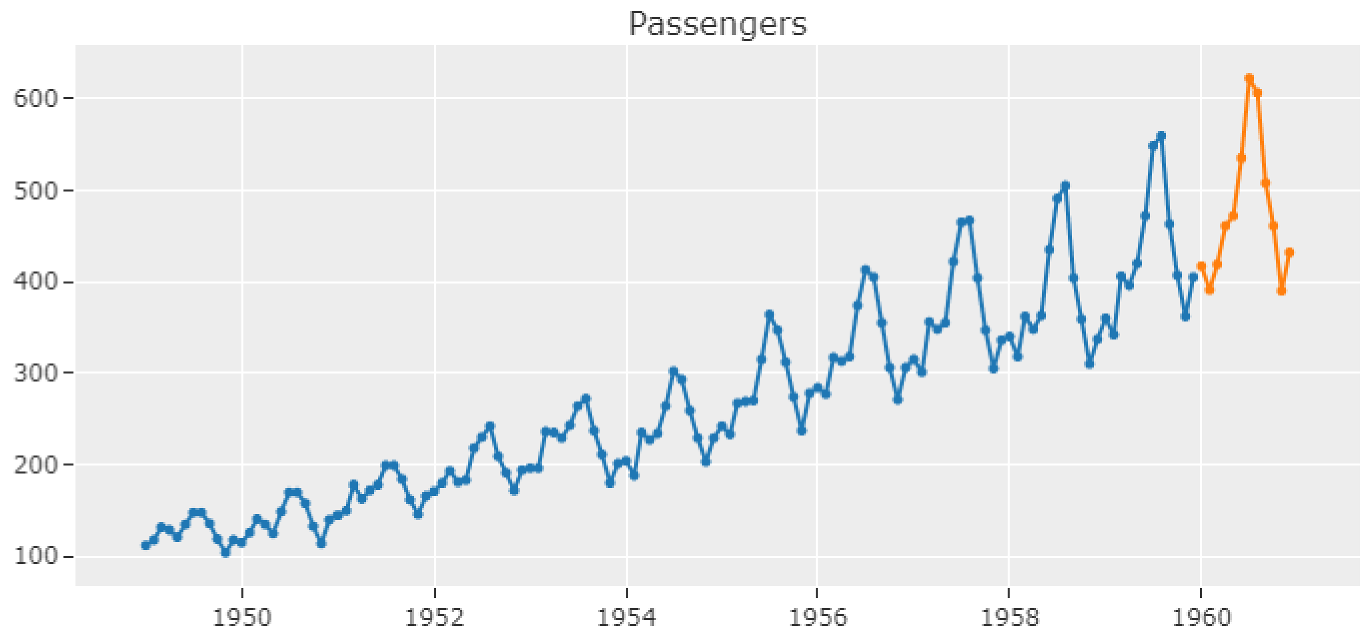
$$air\ passengers_{t+1} = \beta_0 + \beta_1 gdp_t + \beta_2 unemp_t + \beta_3 saving_t + \cdots + u_t$$

- Assumption: everything is reflected in passengers' number already! So, why not →

$$passengers_{t+1} = \mathrm{f}(passengers_t, passengers_{t-1}, \ldots)$$

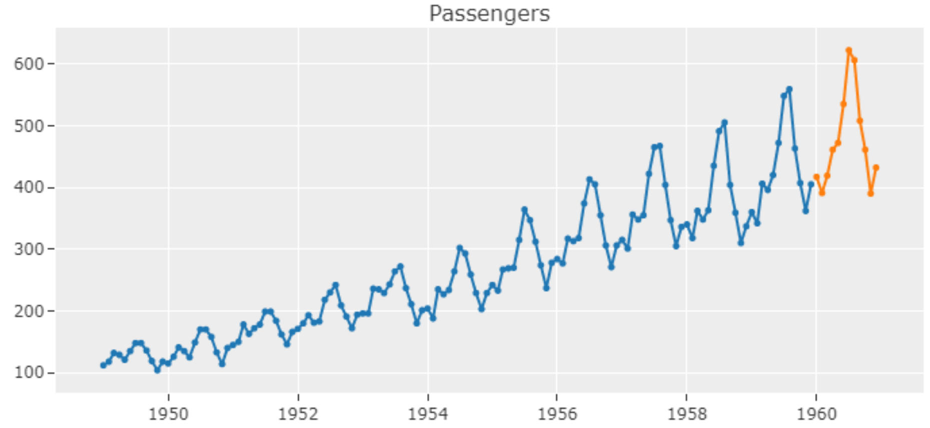
Modeling, a simple example (cont'd)

$$passengers_{t+1} = \mathrm{f}(passengers_t, passengers_{t-1}, \dots)$$



- Goal: Identify existing patterns in data to project future patterns!

Why do we need ML/DL?



$$passengers_{t+1} = \mathrm{f}(passengers_{t}, passengers_{t-1}, \ldots)$$

- How do we identify the functional form of f( )?
- How about this:

$$passengers_{t+1} = c + \phi_1\, passengers_t + \phi_2\, passengers_{t-1} + \ \ldots + \epsilon_t$$

- What if we are looking for more complex relationships?
- Or a non-parametric approach?

Pedram Jahangiry

Predictive modeling vs forecasting?

$$passengers = \beta_0 + \beta_1 gdp + \beta_2 unemp + \beta_3 saving + \cdots + u$$

	Predictive Modeling	Forecasting
Definition	Creating models to analyze data and predict an outcome.	Making informed predictions about future events based on past data.
Focus	Any point in time (past, present, future)	Primarily focused on future outcomes.
Variables	Multivariate	Typically, univariate

- Forecasting is a subset of predictive modeling.
- We can use econometrics, ML or DL for predictive modeling.

Pedram Jahangiry

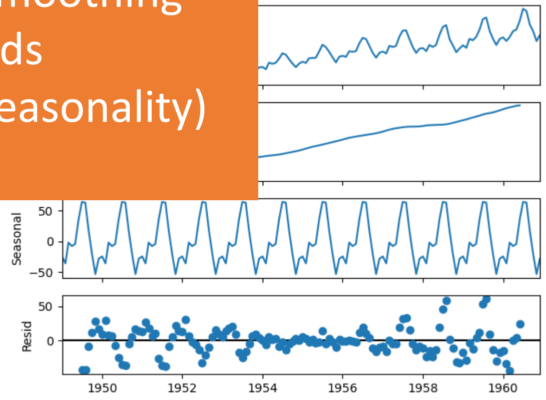
Econometrics

- Decomposition-Based: Focuses on modeling through **decomposition** and **autocorrelations**.
- Parametric Structure: Defined mathematical form; interpretable components.
- Statistical Tools: Confidence intervals, hypothesis testing for insights.

ETS

Exponential Smoothing methods

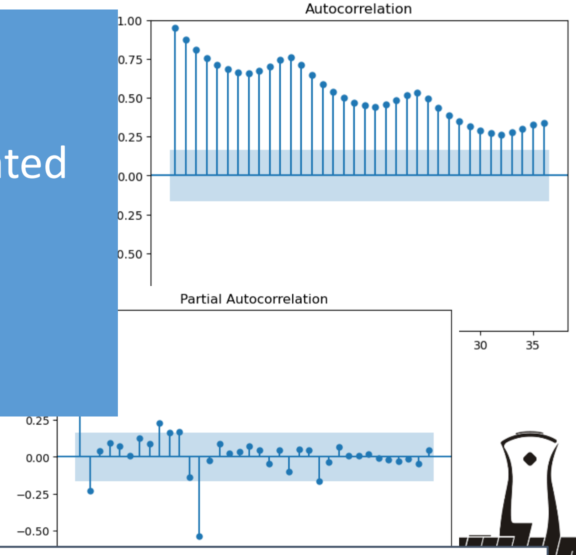
(Error, Trend, Seasonality)



ARIMA

Autoregressive Integrated Moving Averages

SARIMAX

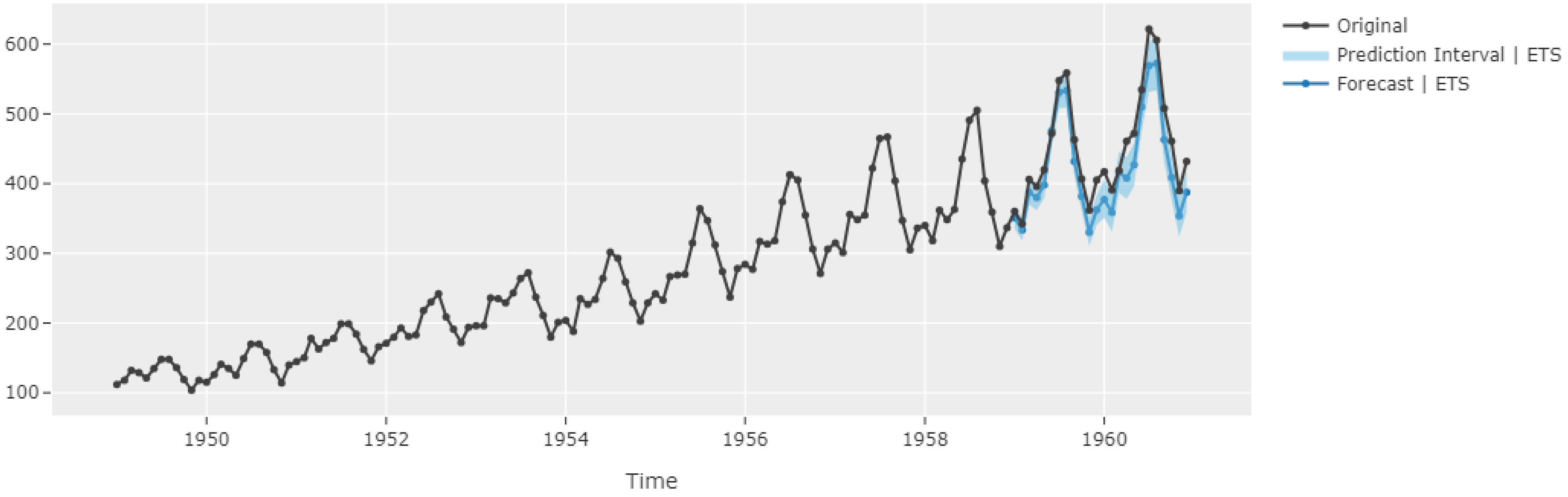


$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Pedram Jahangiry

Econometrics Example: ETS

- ETS decomposes time series into Error, Trend, and Seasonal components.



Econometrics Limitations

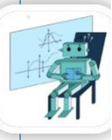
- Cannot handle strong, complex non-linear relationships
- **Transition to Machine Learning**: ML/DL models are inherently designed to capture complex non-linear patterns in time series

Pedram Jahangiry

Machine Learning

- **Data Transformation**: preprocessing of time series data into a supervised learning format, creating features and a target variable.
- **Flexibility**: Can capture complex nonlinear relationships without explicit mathematical models for the underlying processes.
- **Feature Importance**: Offers insights into which features (lags, external variables) are most predictive.
- **Challenges**: Overfitting, plus Requires special cross validation and bootstrapping

Supervised
- Regression
- Classification

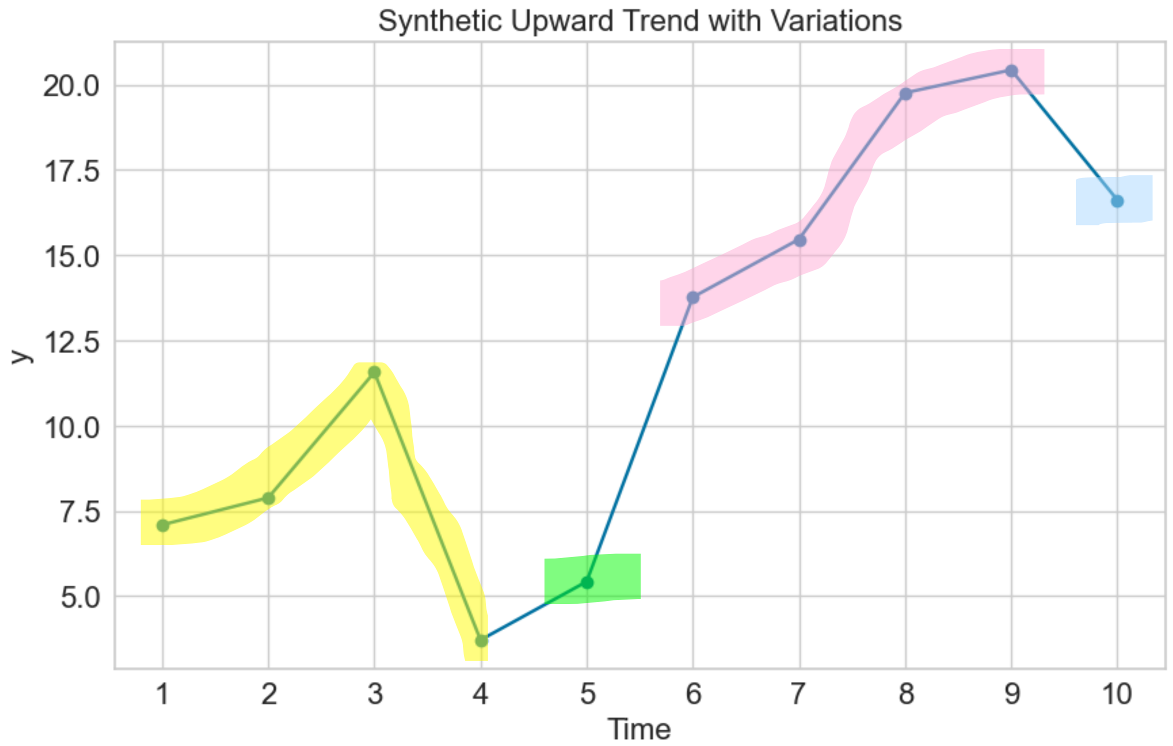


Unsupervised
- Clustering
- Anomaly detection

Semi-supervised

Self-supervised

ML Data Transformation (Single-output)



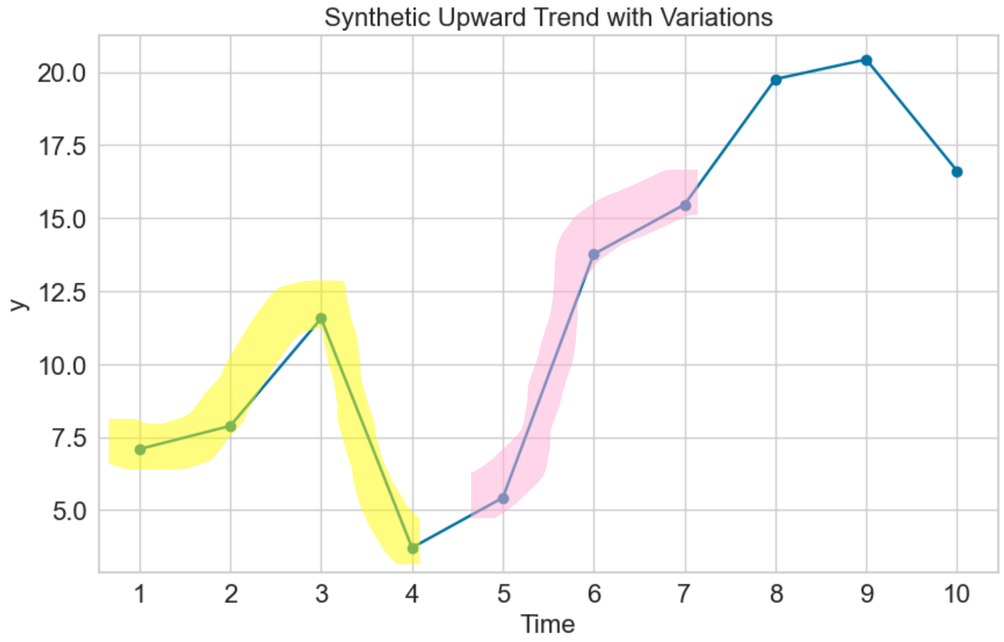
Features (X)				Target (y)
y_{t-4}	y_{t-3}	y_{t-2}	y_{t-1}	y_t
y_1	y_2	y_3	y_4	y_5
y_2	y_3	y_4	y_5	y_6
y_3	y_4	y_5	y_6	y_7
y_4	y_5	y_6	y_7	y_8
y_5	y_6	y_7	y_8	y_9
y_6	y_7	y_8	y_9	y_{10}

TS raw data → TS Supervised data

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}

Pedram Jahangiry

ML Data Transformation (Multi-output)



Features (X)				Target (y)		
y_{t-4}	y_{t-3}	y_{t-2}	y_{t-1}	y_t	y_{t+1}	y_{t+2}
y_1	y_2	y_3	y_4	y_5	y_6	y_7
y_2	y_3	y_4	y_5	y_6	y_7	y_8
y_3	y_4	y_5	y_6	y_7	y_8	y_9
y_4	y_5	y_6	y_7	y_8	y_9	y_{10}

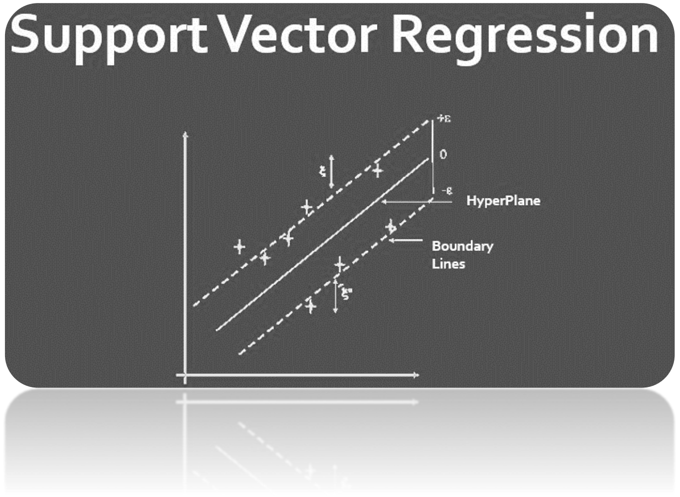
TS raw data

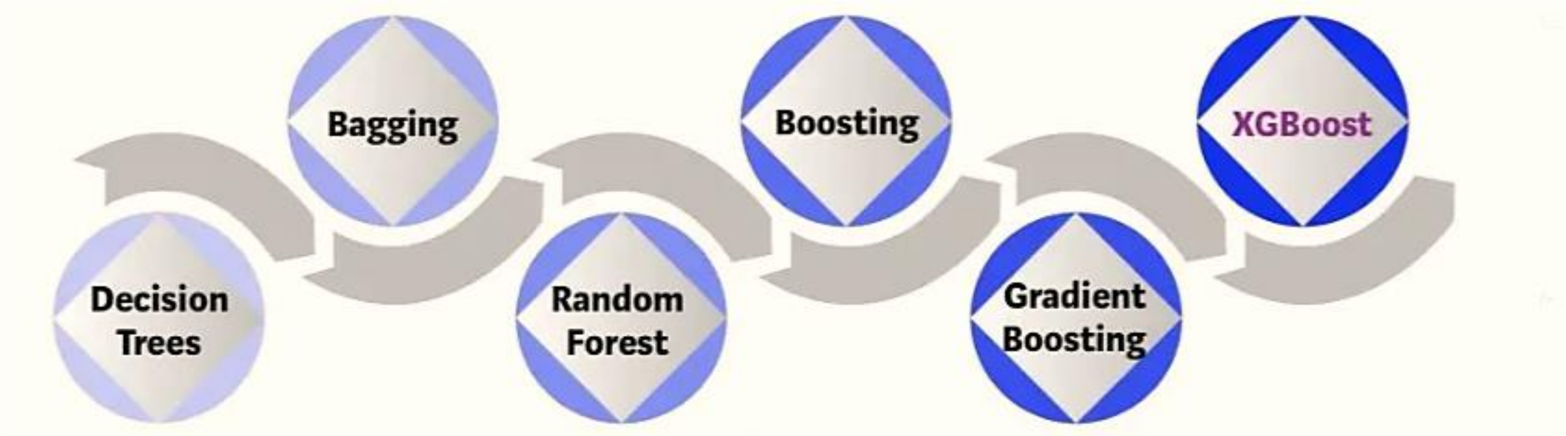
y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}

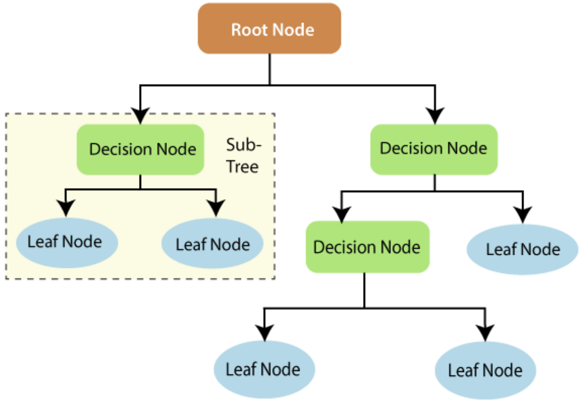
→ TS Supervised data (h=3)

Pedram Jahangiry

Machine Learning Models







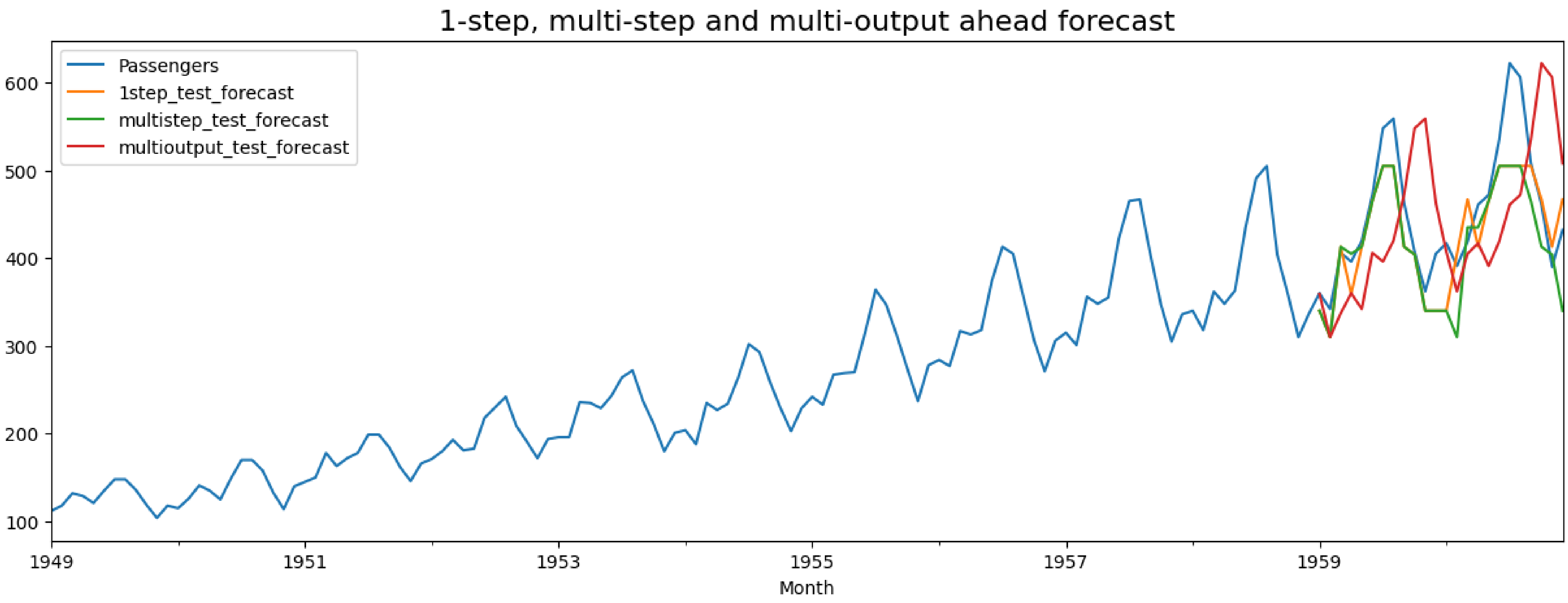
dmlc XGBoost

LightGBM

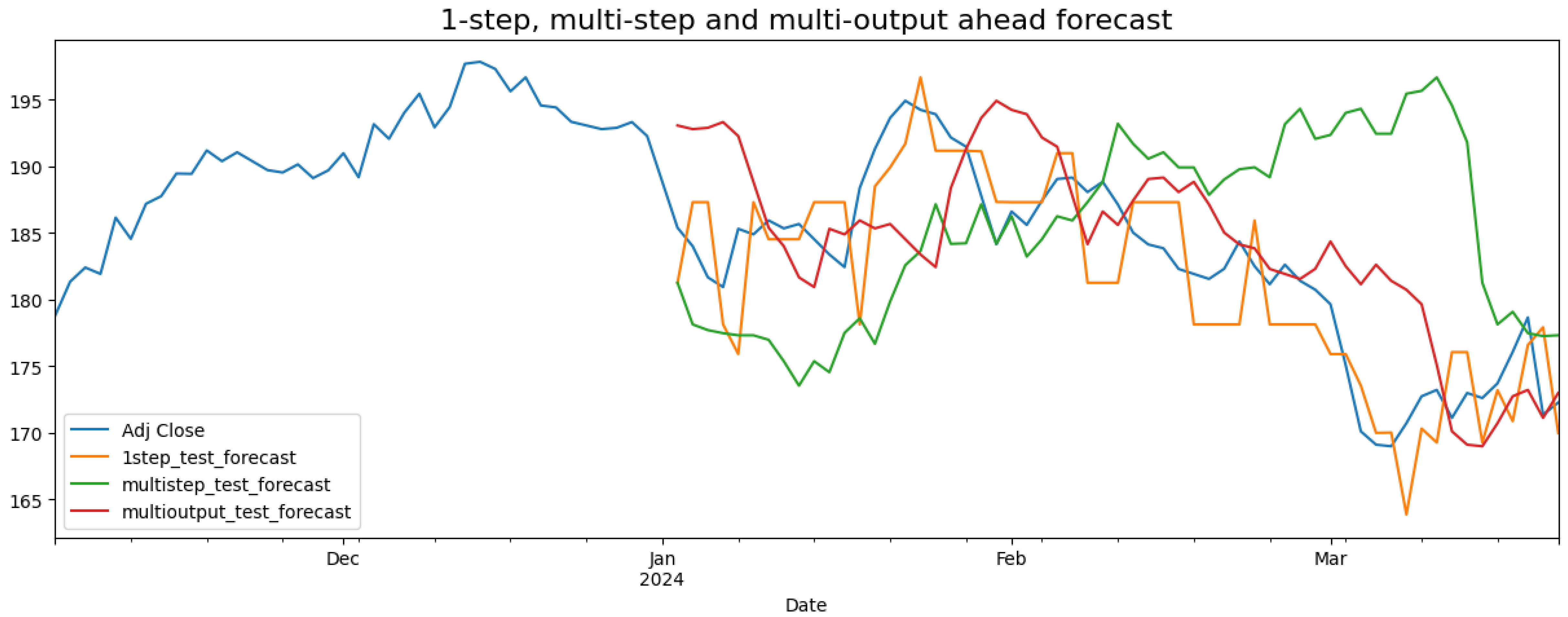
CatBoost

Pedram Jahangiry

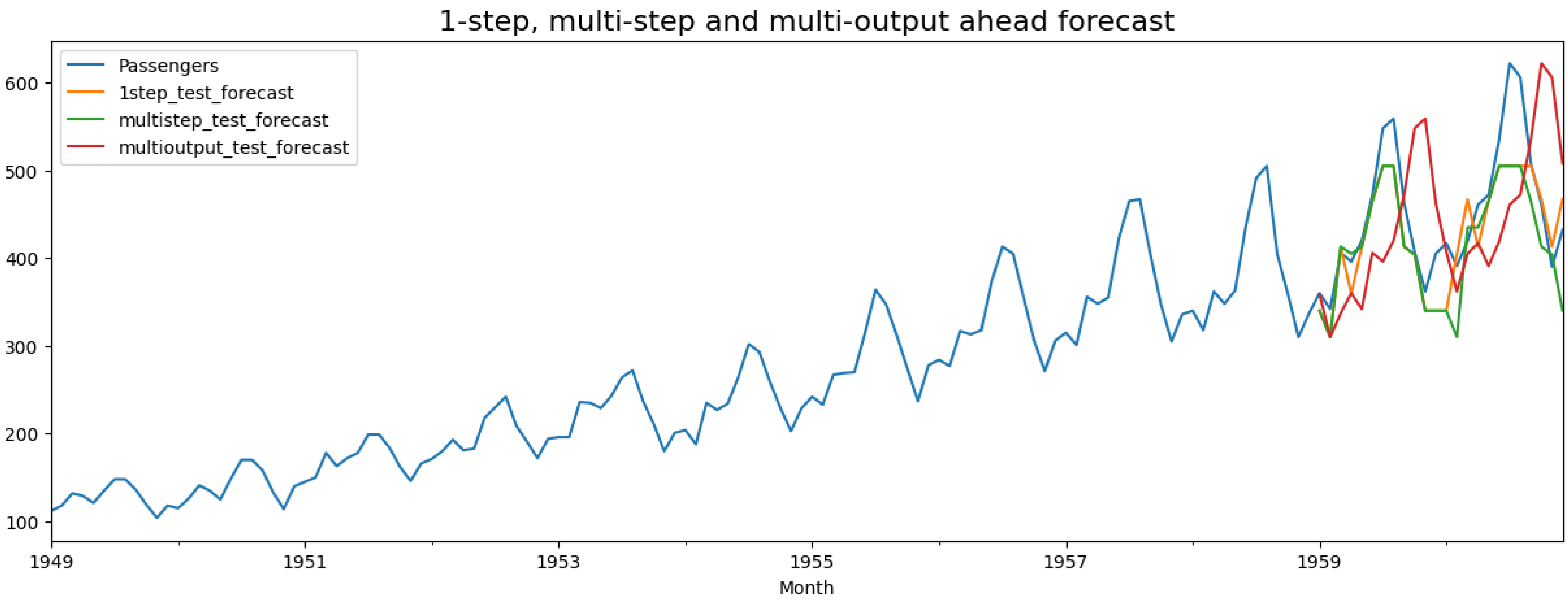
ML Example: Random Forest with 12 lags



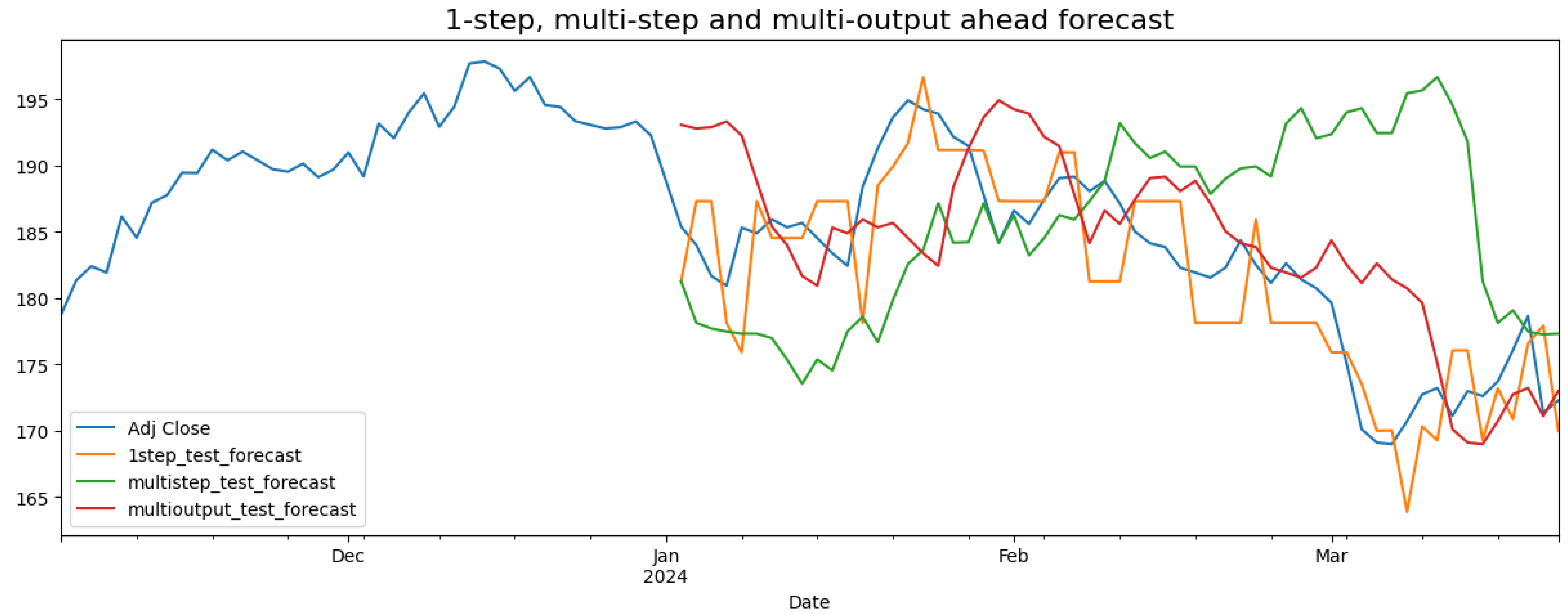
ML Example: Random Forest with 21 lags



ML Example: Random Forest with 12 lags

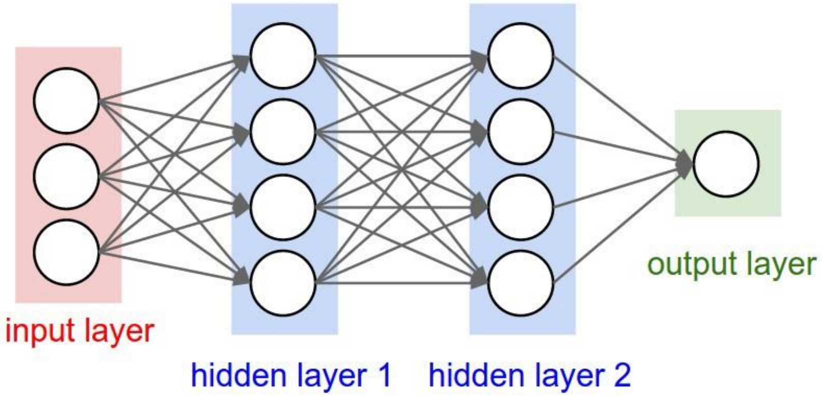


ML Example: Random Forest with 21 lags

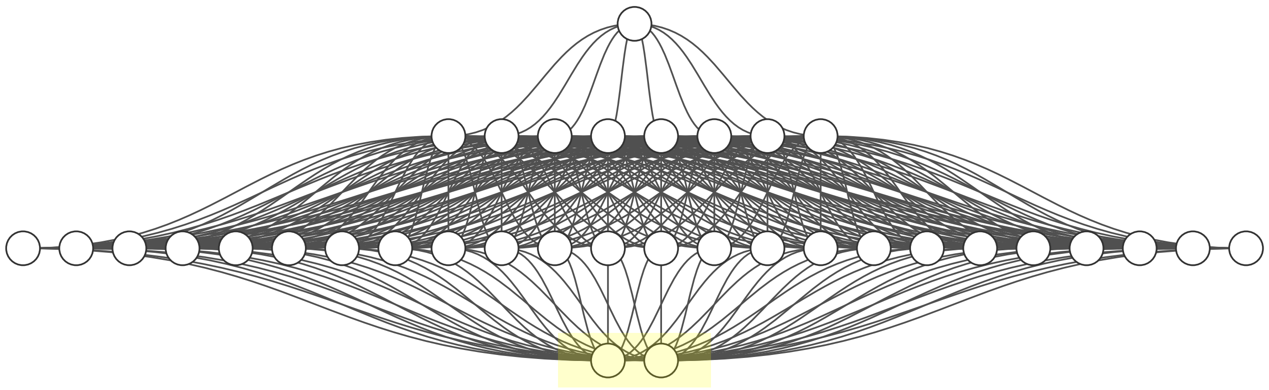


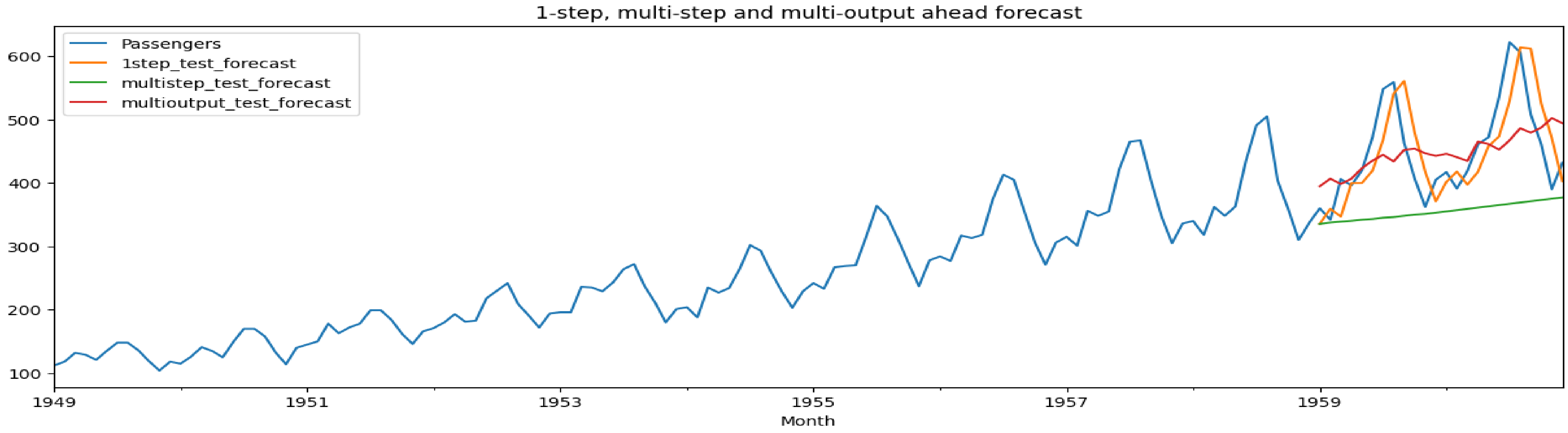
Deep Learning

- Just like ML, DL models require Data Transformation.
- Non-Parametric Nature: DL models do not assume a predefined form for the time series model, allowing them to learn complex and highly nonlinear patterns directly from data.
- High Capacity and Flexibility: Can model very complex relationships and interactions, including multivariate and high-dimensional time series. Well-suited for large datasets.

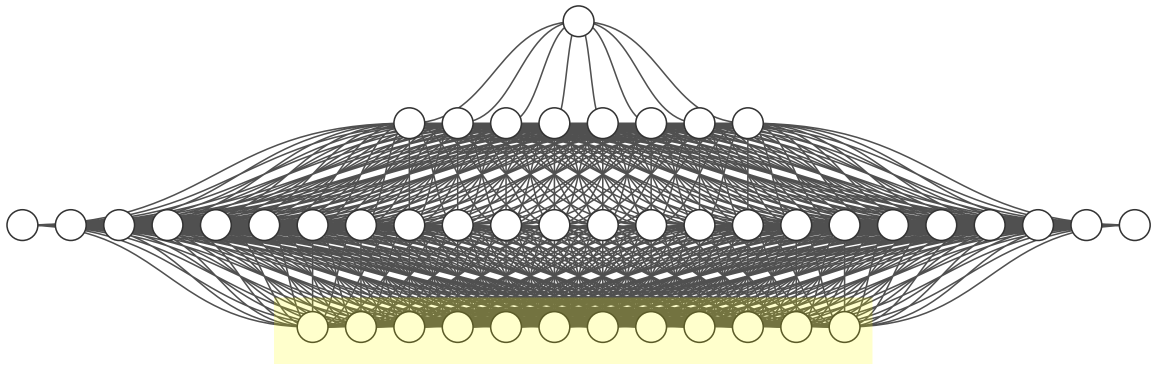


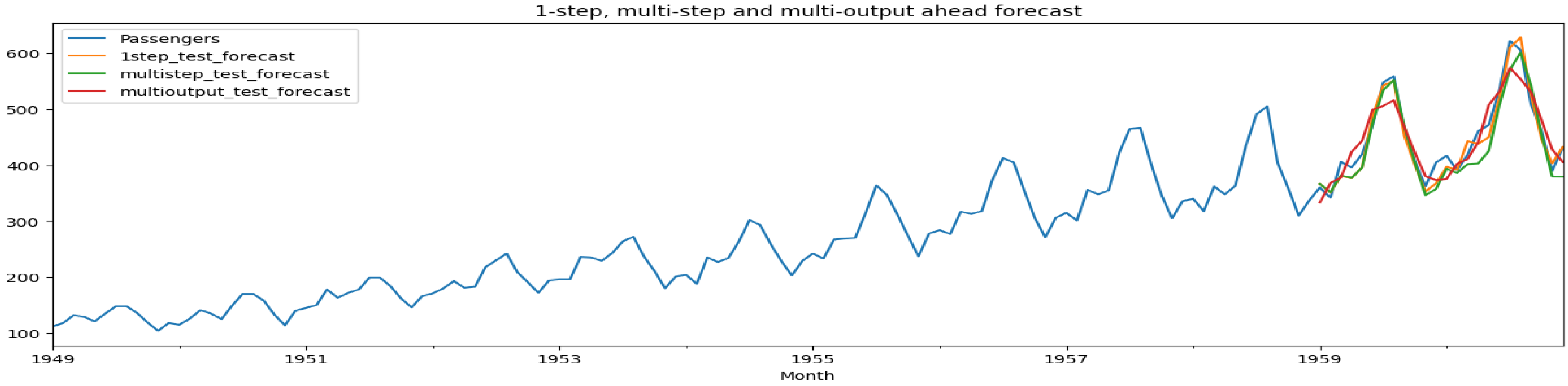
DL Example: DNN with 2 lags





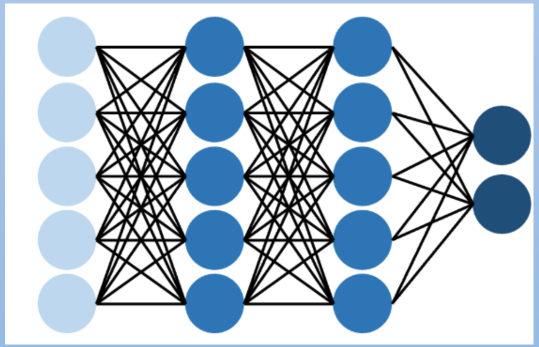
DL Example: DNN with 12 lags



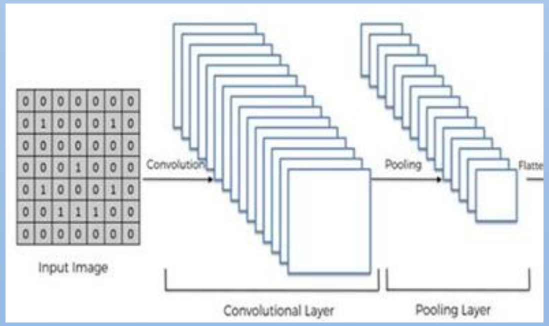


Deep Learning Architectures

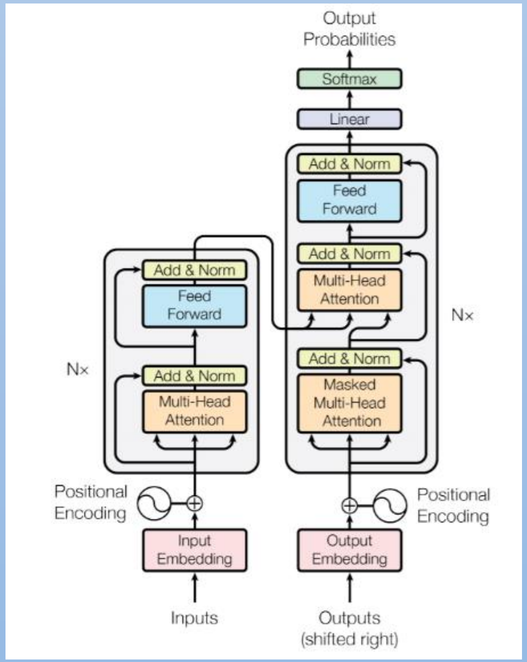
Standard NN



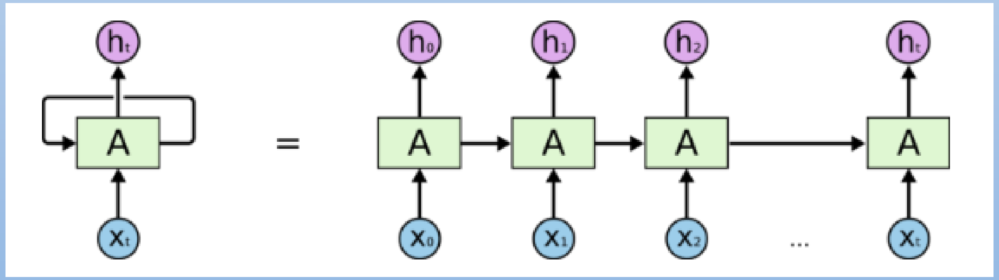
Convolutional NN



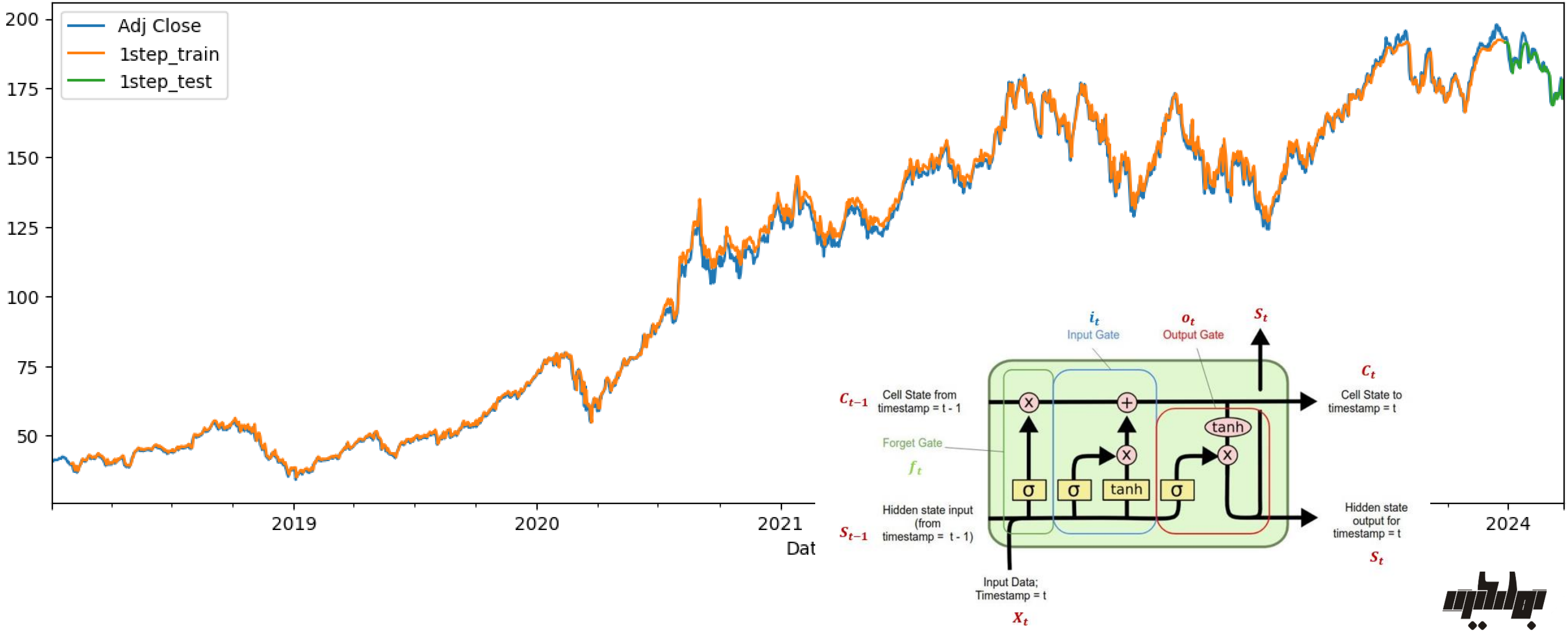
Transformers



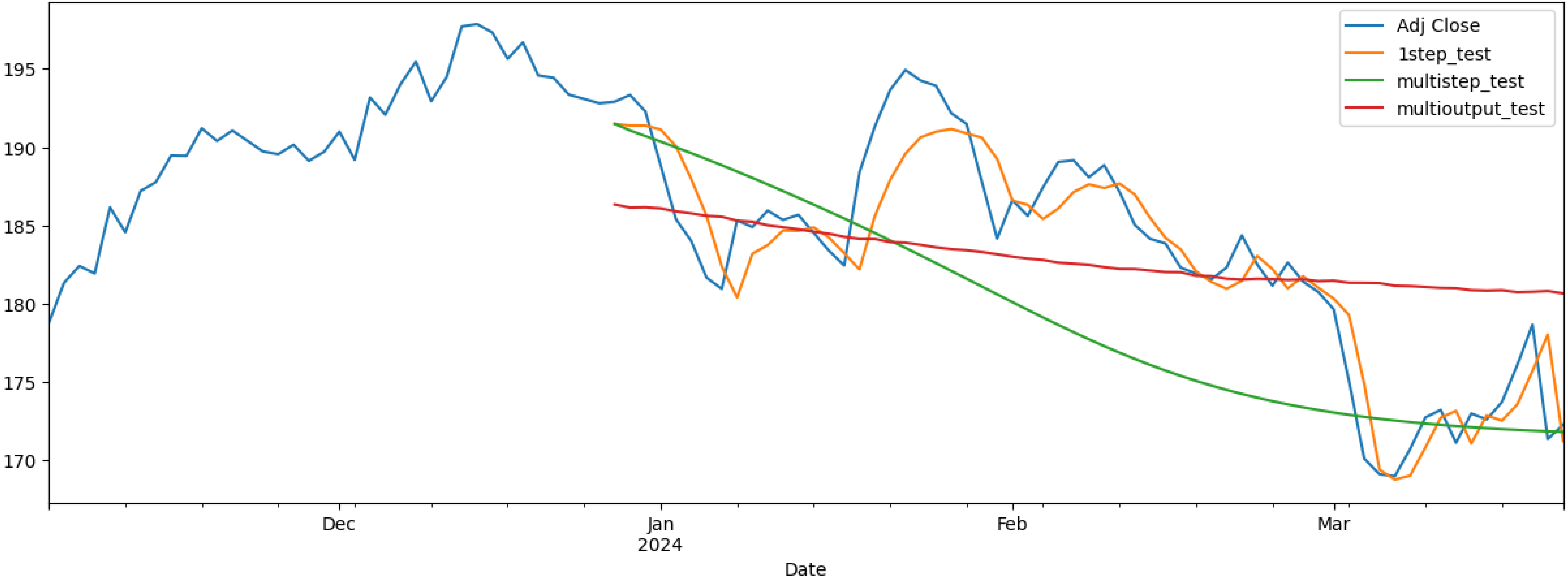
Recurrent NN



DL Example: LSTM with 21 lags



DL Example: LSTM with 21 lags



Modeling Comparison

Aspect	Econometric Models	Machine Learning (ML)	Deep Learning (DL)
Feature Engineering	Requires explicit modeling of seasonality and trend	Captures complex patterns with less need for manual engineering	Often automates feature engineering
Model Complexity	**Lower**; focuses on data generation process	**Moderate**; can handle non-linear interactions	**High**; suited for complex and high-dimensional data
Interpretability	**High**; interpretable parameters, statistical tests	Moderate; provides feature importance	Low; considered a 'black box' approach
Computational Intensity	Generally lower	Varies; dependent on model complexity and data size	High; requires significant computational resources
Data Suitability	Works well when the underlying process is **well understood**	Effective for structured datasets with **complex relationships**	Ideal for large datasets, including **unstructured** data
Core Models	ARIMA, Exponential Smoothing (ETS)	Random Forest, Gradient Boosting	RNNs, LSTMs, CNNs

Python packages

Package	Metrics	ML	DL	Focus
statsmodels	✓✓✓✓			Statistics, Econometrics
scikit learn	✓	✓✓✓✓		General Machine Learning
Keras			✓✓✓✓	General Deep Learning
PYCARET	✓✓✓	✓✓✓		Auto ML, Rapid prototyping, Comparison
Darts	✓✓✓	✓	✓✓	Advanced timeseries and forecasting
GLUONTS	✓	✓	✓✓	Large Scale probabilistic Models
NIXTLA	✓✓✓	✓✓✓	✓✓✓	Workflows, SOTA methods

Pedram Jahangiry

Module 1- Part 4

1. Timeseries Basics
    - What is sequence data?
    - Time series tasks
    - Trend, Seasonality, Residual
    - ACF and PACF
    - Stationarity

2. Forecasting strategies
    - One-step, Multi-step, Multi-output
    - Univariate vs Multivariate
    - Benchmarks

3. Modeling
    - White Noise
    - Econometrics
    - Machine Learning
    - Deep Learning

4. Can we beat wall street?
    - Is stock price Random Walk? (EMH)
    - Does Uni-variate Model work?
    - Can Multi-variate Models help?
    - So What?

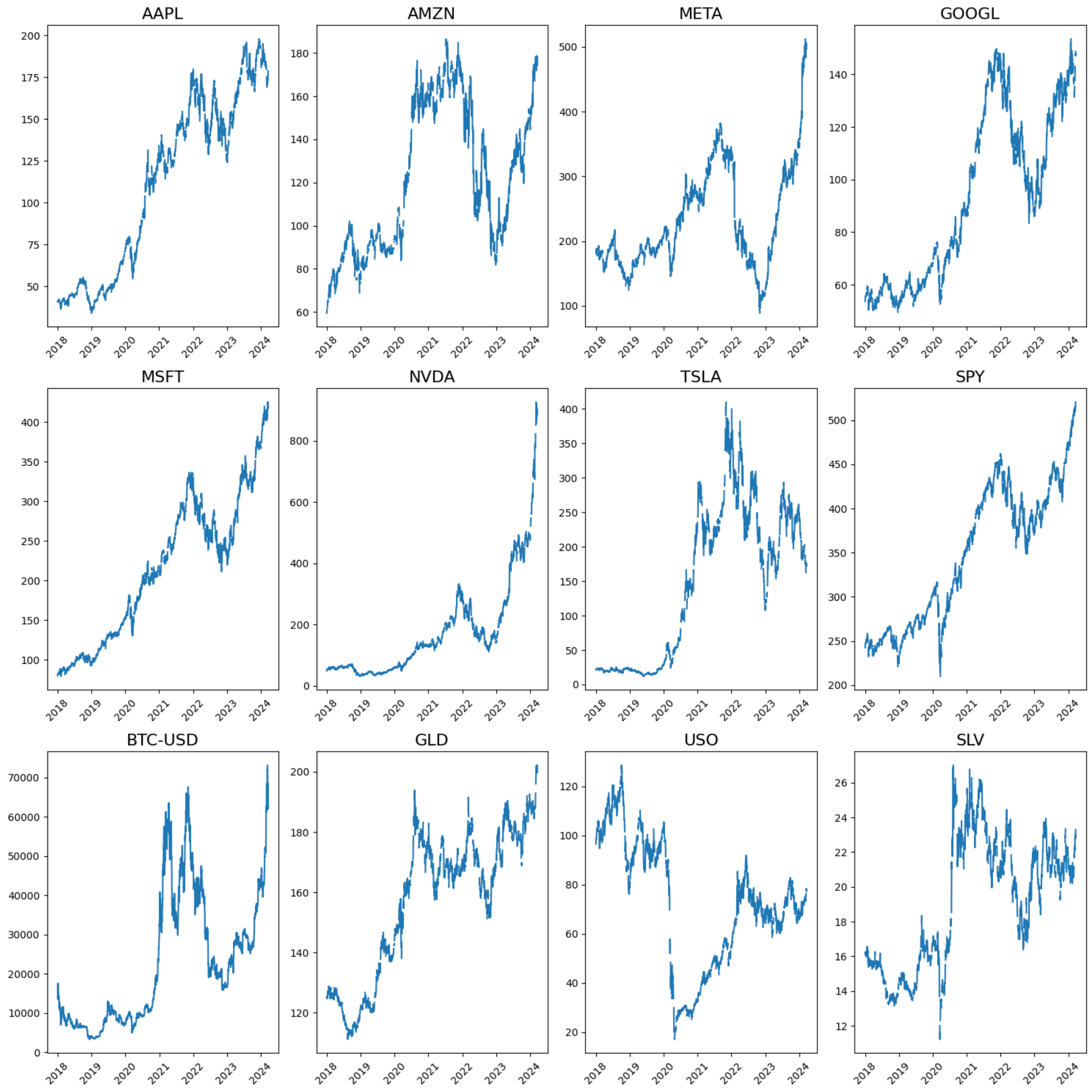
Pedram Jahangiry

Is stock price Random Walk?

- Efficient Market Hypothesis (EMH): Stock prices reflect all available information.
- Empirical evidence: Statistical tests often fail to reject the random walk hypothesis. Many forecasting models struggle to **consistently** outperform naive benchmarks.
- Unpredictable Events: Markets are driven by news, economic shifts, and geopolitical events, many of which are inherently unpredictable.

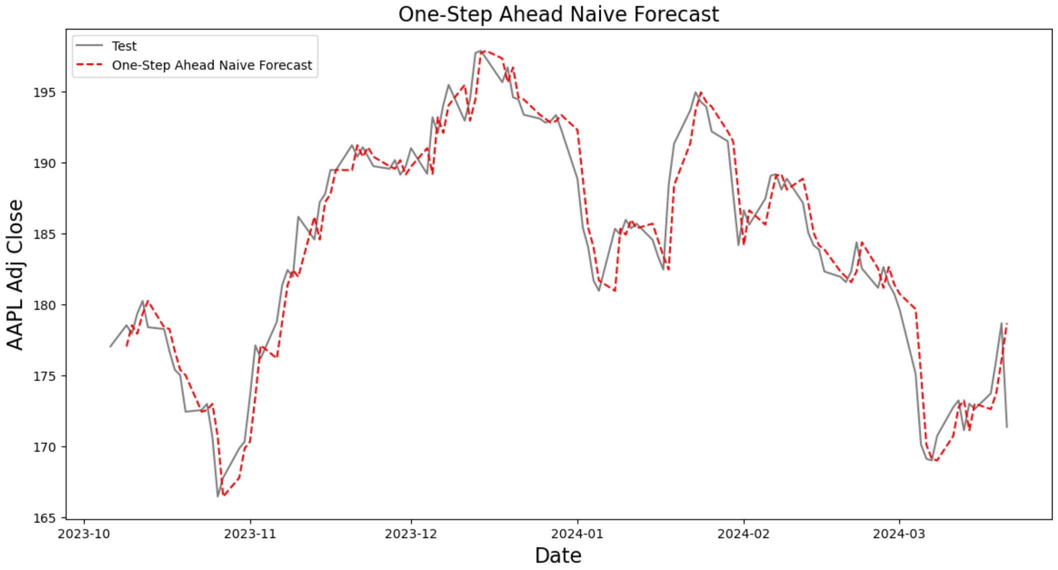
Pedram Jahangiry

4. Beating Wall Street



Does Uni-variate Model Work?

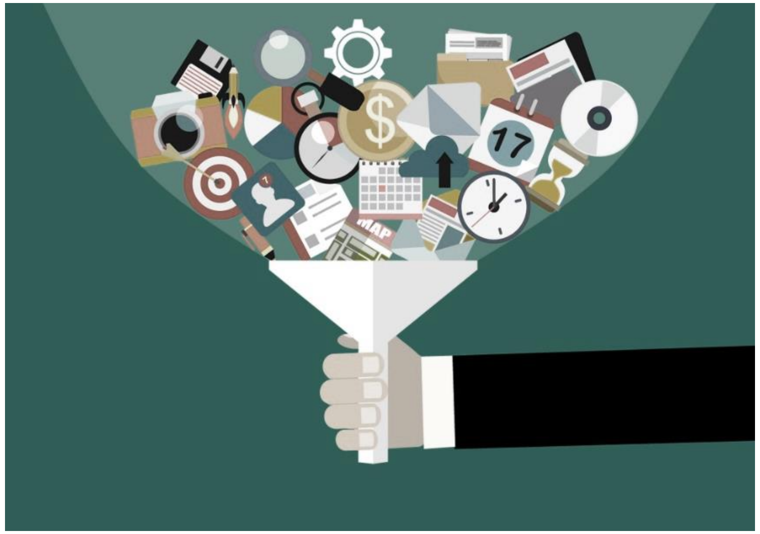
- Hard to beat naïve forecaster!
- Complex models might offer marginal gains, but the effort/cost of constant updating is substantial.
- Trading strategy is a whole other challenge: bet sizing, transaction costs, and turning forecasts into profitable decisions.



Pedram Jahangiry

Can Multi-variate Models Help?

- Potential for Improvement by using Alternative data
- Feature Engineering is Key: The success of multi-variate models often hinges on identifying and extracting strong, predictive features from the diverse data.
- Increased Complexity and accessibility challenge: Building and maintaining multi-variate models is often computationally more demanding
- Trading strategies remains complex due to market efficiency and transaction costs.



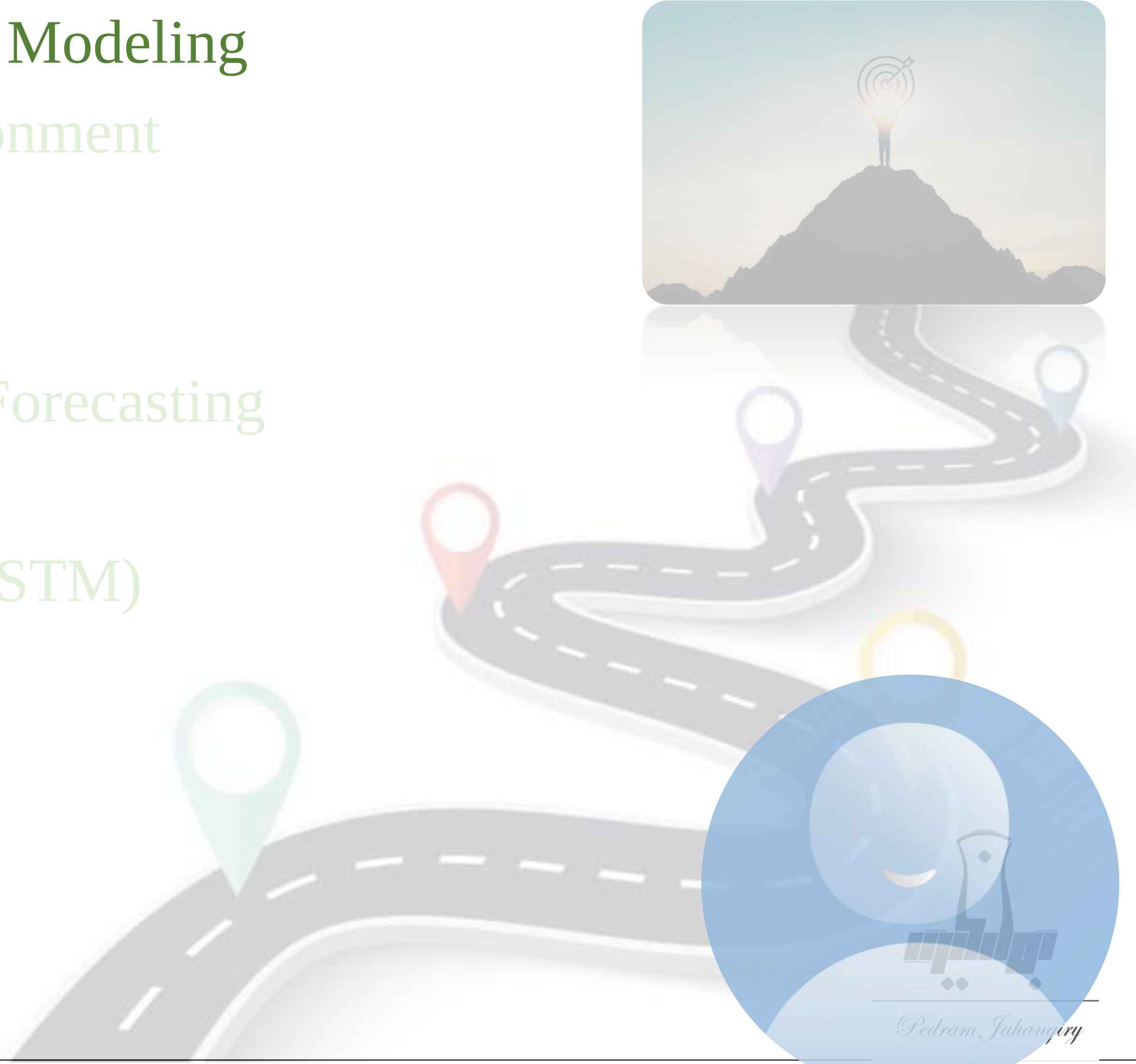
Pedram Jahangiry

DALL-E

DALL-E

Road map!

- ✓ Module 1- Demystifying Timeseries Data and Modeling
- Module 2- Setting up Deep Forecasting Environment
- Module 3- Exponential Smoothing
- Module 4- ARIMA models
- Module 5- Machine Learning for Time series Forecasting
- Module 6- Deep Neural Networks
- Module 7- Deep Sequence Modeling (RNN, LSTM)
- Module 8- Prophet (Forecasting at Scale)



Where to find the materials?