Business Forecasting

ADIA Course

Day 1 – Session 2

Madhurima Panja

Sorbonne University, Abu Dhabi, UAE

Research Areas: Time Series Forecasting and Machine Learning

June 30, 2025

Time Series Analysis vs. Forecasting

Time Series Analysis

Purpose: Understand the data

- Reveals the underlying structure (functional form)
- Helps interpret trends, seasonality, and influencers
- Explains why things happen
    1. Is the series mean-reverting?
    2. Why does it grow explosively?
- Evaluates predictability

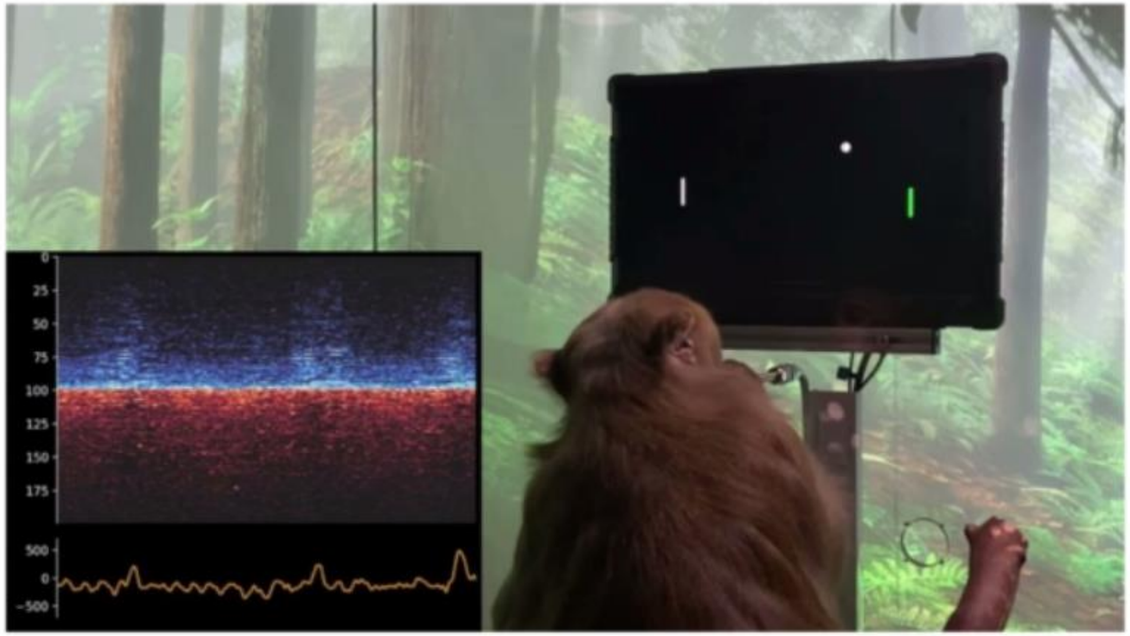
Forecasting

Purpose: Predict the future

- Answers questions like:
    1. "What will happen next?"
    2. "What's the value tomorrow?"

Time Series Classification

- Predicting a number: regression
- Predicting a category: classification
- E.g. Neuralink is reading your brain signals - hungry or tired
- E.g. Smartphone accelerometer - walking, sleeping, sitting



Types of Forecasting Tasks

1. **One-Step Forecasting**

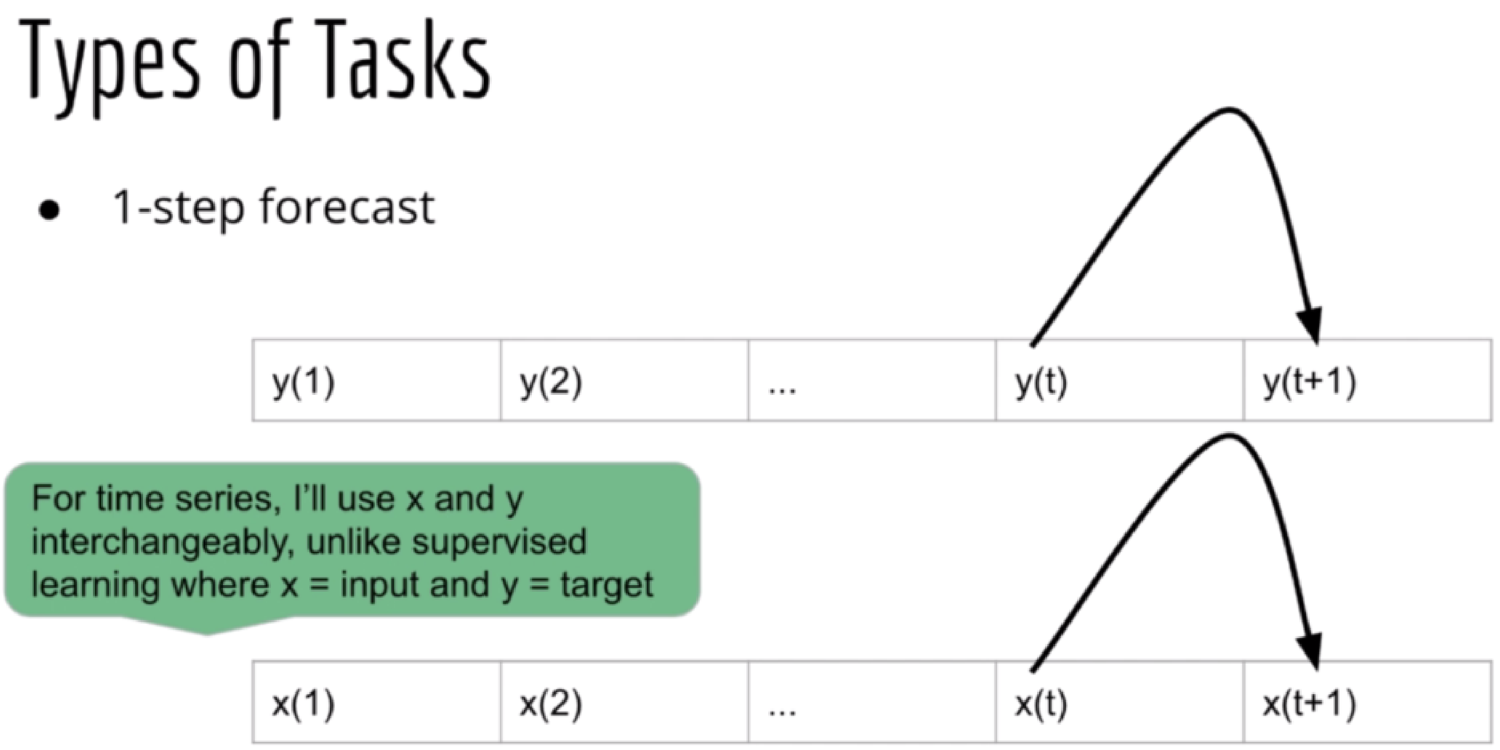
2. **Multi-Step Forecasting**

   - Incremental Multi-step

   - Multi-output Multi step

One-Step Forecast

Here we will forecast for only one time - Very next frame.



Multi-Step Forecast

The **most of the time** we would be needing such forecast.

— The term used is: "**Forecast horizon**" = Multiple steps in the future.

Now, there are **2** ways to produce such "**Multi step**" forecasts.

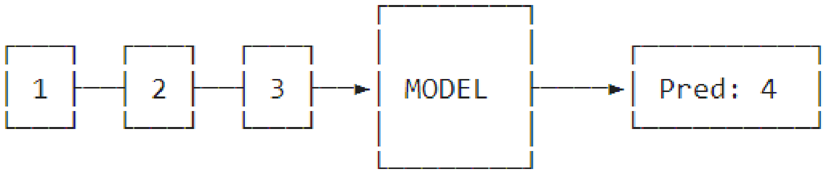
1. **Incremental method** (which can be done with any 1 step predictor)

2. **Multi-output forecast** (limited to certain models)

Remember: These are just the methods not the types of models.

Incremental Multi-Step Forecast

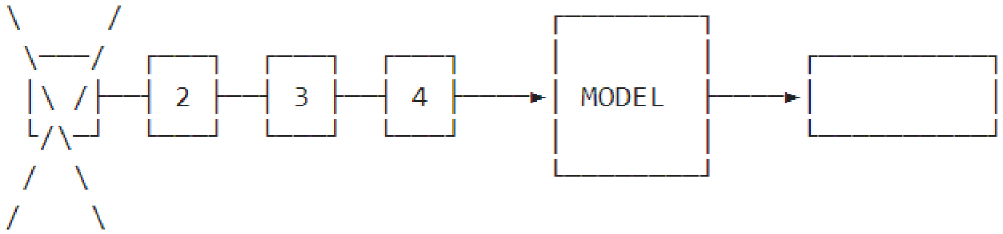
The model can be trained on the fixed **n** days. So, suppose



Here, our **p = 3** and **h = 2**.

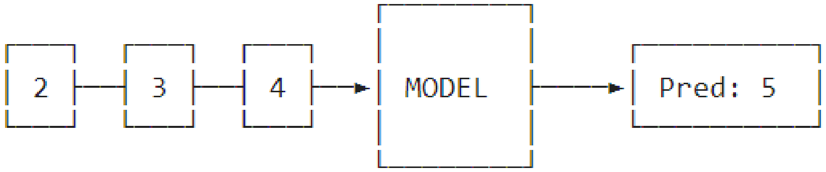
- **p** = How many days do you want to base your forecast on?
- **h** = How many days do you want to forecast?

Now, in the first iteration we have forecast 4th day. But still one day is yet to be forecast so we will take day 4 as the input.



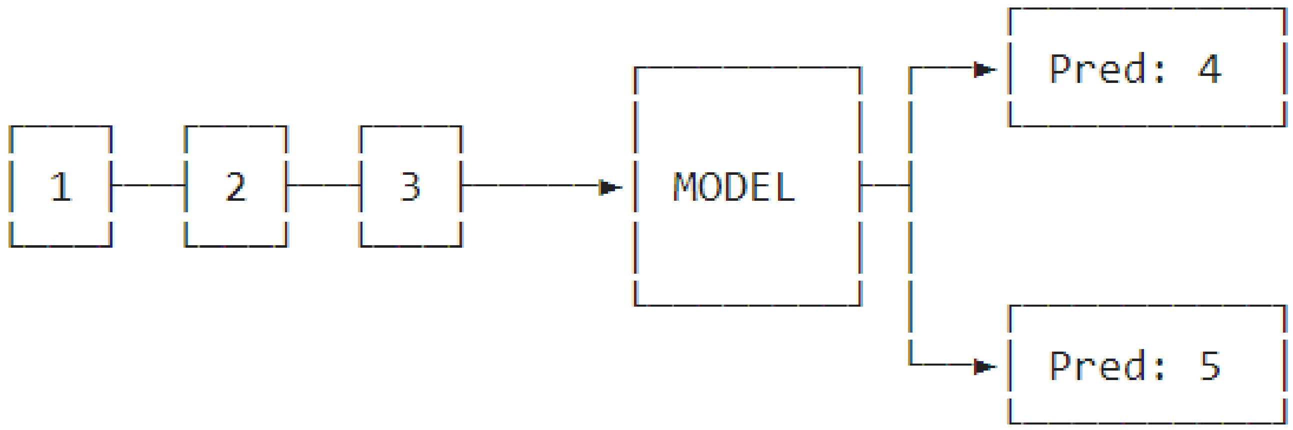
So, instead of increasing the number of days, we will **eleminate the first ones** to keep the p consistant.

This is how **incremental multi-step** works in a nutshell.



Multi Output Forecast

Here, we feed the time-series in and get the **h days out and at once**.



Multi-step Forecasting

Short -term	Involves predicting events only a few time periods (days, weeks, months) into future
Medium - term	Involves predicting events to one or two years into future
Long - term	Involves predicting events to many years into future

Type	Uses
Short –term & Medium-term	Budgeting and selecting new research for development projects
Long - term	Strategic planning

Multi-step Forecasting

Short -term	Involves predicting events only a few time periods (days, weeks, months) into future
Medium - term	Involves predicting events to one or two years into future
Long - term	Involves predicting events to many years into future

Type	Uses
Short –term & Medium-term	Budgeting and selecting new research for development projects
Long - term	Strategic planning

Financial Terms in Time Series

Stock Return

It is the same what we call "percentage-change".

"How much change is there in the stock price since we have bought it?"

$$return = \frac{\text{Price}_{\text{current}} - \text{Price}_{\text{buy}}}{\text{Price}_{\text{buy}}}$$

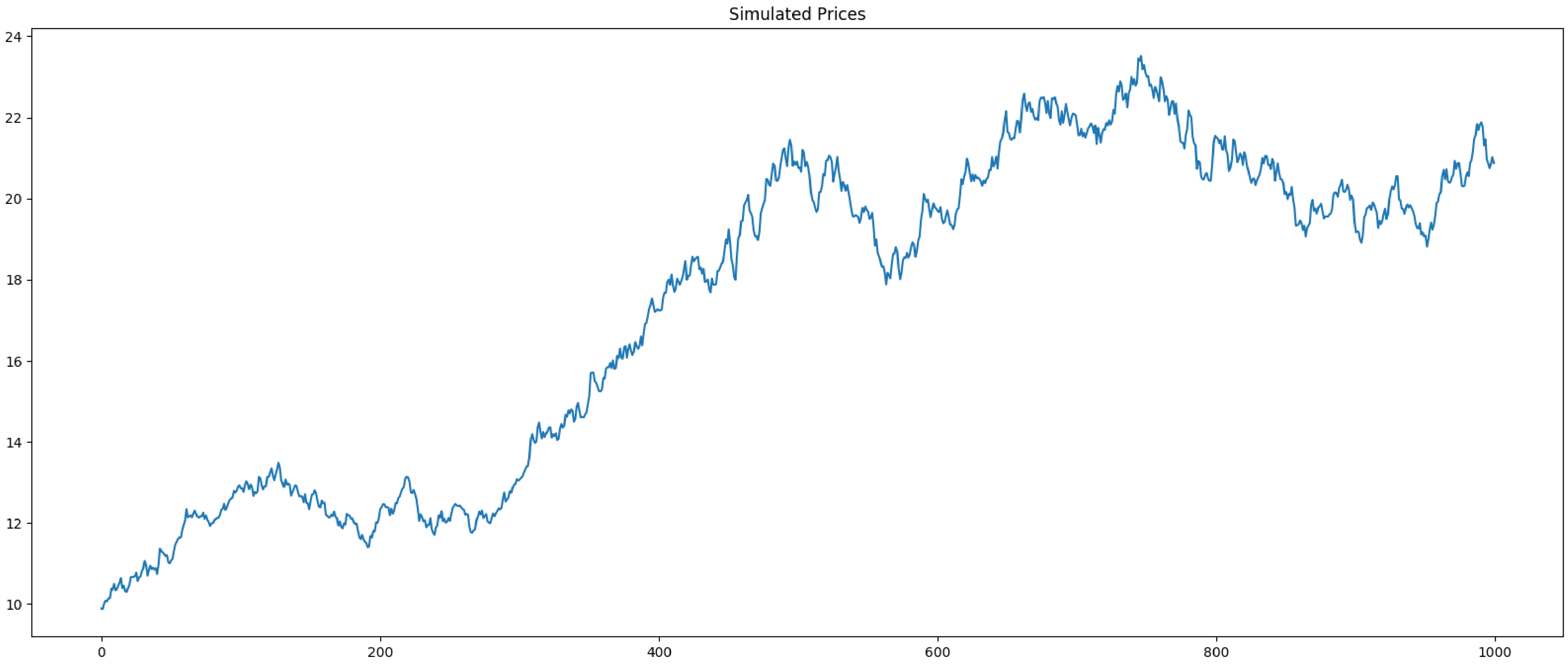
Gross Return

It gives the "single number" that we can use **to directly** get the final price.

$$GrossReturn = \frac{\text{Price}_{\text{current}}}{\text{Price}_{\text{buy}}}$$

Meaning, it will return the *"multiplier"*.

Financial Simulation



Random Walk

Introduction

- In time series analysis, a random walk is a stochastic process where future values are determined by previous values plus a random shock.
- Understanding the random walk is crucial for modeling and forecasting in various fields.

What is it?

- It is about *those* stocks which follow a **random walk** ie. its price is not predictable.
- "*Sometimes the best-fitting model is in fact a random walk*"
- It is either to go up or down randomly having 50-50 chance on each direction
- It is ***impossible*** predict the next as there is 50-50 chances

Characteristics of a Random Walk

- **Constant Drift:**
  - A random walk typically exhibits a constant drift or trend over time.
  - The drift reflects the average rate of change in the series.
- **Unpredictable Movement:**
  - Future values of a random walk are unpredictable.
  - Each step depends solely on the current position and a random shock, making forecasting challenging.
- **No Auto-correlation:**
  - A random walk typically has no autocorrelation.
  - The correlation between observations at different lags is close to zero.

Gaussian Random Walk

- It is where the up and down values come from the Gaussian distribution
- Meaning, Adding the np.randn0 as the noise
- So, here the prices don't go 1 unit up or down but go any value from the Gaussian distribution, thus called: Gaussian Random Walk

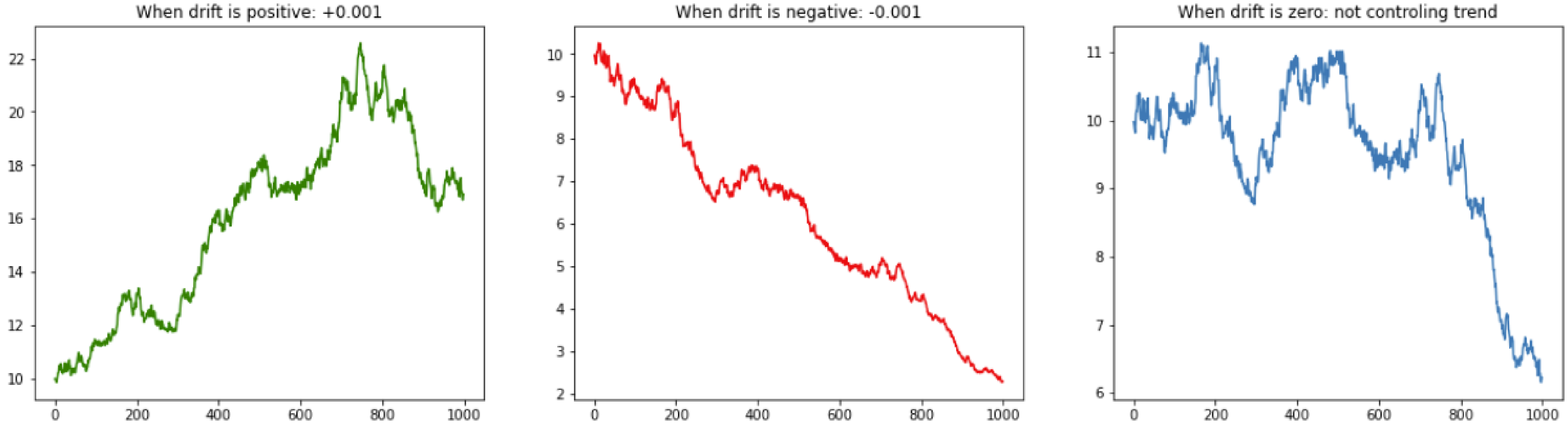
$$new = old + e$$

- Since there is unpredictability, we can only know that the error $e \in N(0, \sigma^2)$
- **Something interesting that we can do with log**
  The general formulae for new price (in random walk):

$$new = old + \mu + e$$

- μ = Drift - This would control the trend of the time series.
- e = The noise $\in N(0, \sigma^2)$

Financial Simulation with changing μ



So, from the charts above - we can see that changing the drift term changes the overall trend of the time series.

Naïve Forecasting Methods

Average Method

- Here, the forecasts of all future values are equal to the average (or "mean") of the historical data. If we let the historical data be denoted by $y_1, \ldots, y_T$, then we can write the forecasts as

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T.$$

- The notation $\hat{y}_{T+h|T}$ is a short-hand for the estimate of y_{T+h} based on the data $y_1, \ldots, y_T$.

Naive Method

- For naïve forecasts, we simply set all forecasts to be the value of the last observation. That is,

$$\hat{y}_{T+h|T} = y_T.$$

- This method works remarkably well for many economic and financial time series.

- Random walk forecasts.

Seasonal Naive Method

- A similar method is useful for highly seasonal data. In this case, we set each forecast to be equal to the last observed value from the same season (e.g., the same month of the previous year). Formally, the forecast for time $T + h$ is written as

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)},$$

where $m =$ the seasonal period, and k is the integer part of $(h-1)/m$ (i.e., the number of complete years in the forecast period prior to time $T + h$). This looks more complicated than it really is.

- For example, with monthly data, the forecast for all future February values is equal to the last observed February value. With quarterly data, the forecast of all future Q2 values is equal to the last observed Q2 value (where Q2 means the second quarter). Similar rules apply for other months and quarters, and for other seasonal periods.

Forecast Evaluation

Forecast Evaluation

Performance metrics such as mean absolute error (MAE), root mean square error (RMSE), and mean absolute percent error (MAPE) are used to evaluate the performances of different forecasting models for the unemployment rate data sets:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \,;}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \,;$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|,$$

Where y_i is the actual output, $\hat{y}_i$ is the predicted output, and n denotes the number of data points.

By definition, the lower the value of these performance metrics, the better is the performance of the concerned forecasting model.