Time Series Analysis and Forecasting

Sourav Karmakar

souravkarmakar29@gmail.com

Outline

Section 1

Introduction to Time Series

Definition

A **time series** is a sequence of data points collected at successive, equally spaced points in time.

- It is a set of observations $\{y_t\}$, where t is the time index ($t = 1, 2, ..., T$).
- The primary goal of time series analysis is to understand the underlying structure and patterns in the data to make predictions about the future.

Examples of Time Series Data

Here are some common examples of time series data:

- **Economics:** Monthly unemployment rates, quarterly Gross Domestic Product (GDP), daily stock prices.
- **Business:** Monthly sales revenue, weekly website traffic, daily number of units sold.
- **Meteorology:** Daily temperature, annual rainfall.

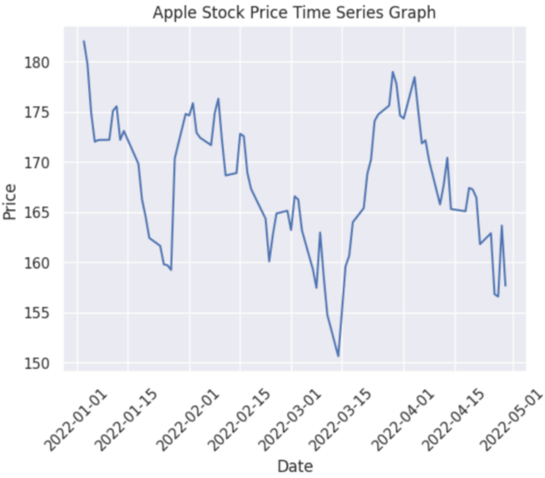


Figure: A typical time series plot.

Examples of Time Series Data

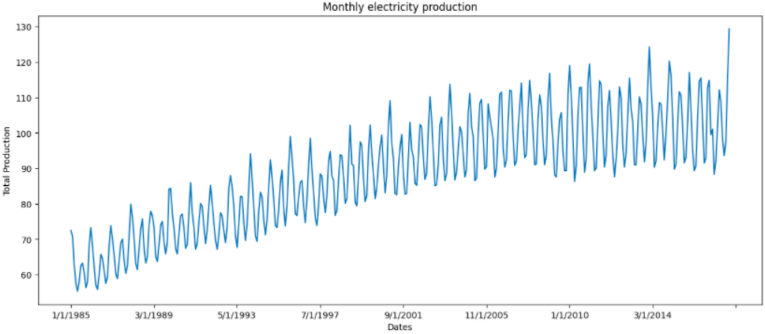


Figure: Another time series plot.

Section 2

Components of a Time Series

Classical Decomposition Model

A time series can be thought of as a combination of different components. We will focus on the **additive model**:

Additive Model

$$Y_t = T_t + S_t + C_t + I_t$$

- Y_t: The observed value at time t.
- T_t: The **Trend** component.
- S_t: The **Seasonal** component.
- C_t: The **Cyclical** component.
- I_t: The **Irregular** (or residual) component.

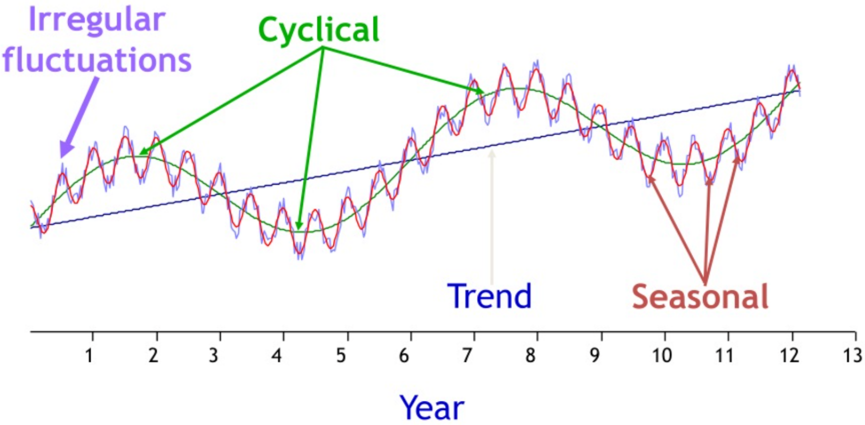
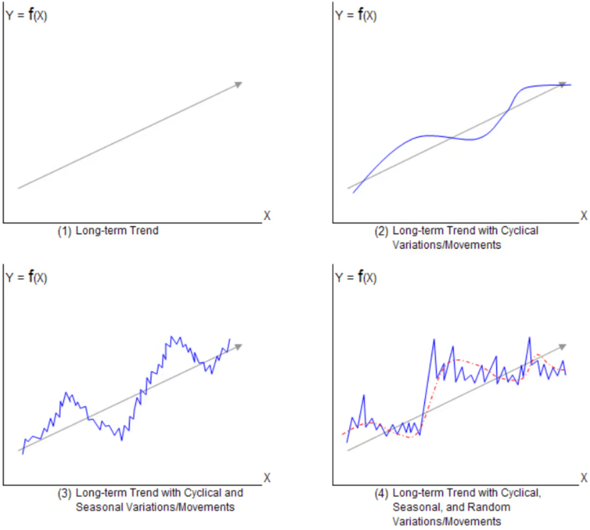


Figure: Different components of time series

Understanding the Components

- **Trend (T_t):** The long-term direction of the series. Is it generally increasing, decreasing, or stable over time?
- **Seasonality (S_t):** A fixed and known periodic pattern. For example, higher sales every December.
- **Cycle (C_t):** Patterns that are not of a fixed period, usually tied to broader economic conditions. Cycles are longer and more variable than seasonal patterns.
- **Irregular (I_t):** Random, unpredictable fluctuations. This is the noise left over after the other components are removed.

Time Series Components: Example



Smoothing: Simple Moving Average (SMA)

Intuition: The idea is to smooth out short-term fluctuations and highlight longer-term trends. We calculate the average of a fixed number of recent observations.

Formula

The k-period moving average at time t is:

$$\text{SMA}_t = \frac{1}{k}\sum_{i=0}^{k-1} y_{t-i} = \frac{y_t + y_{t-1} + \ldots + y_{t-k+1}}{k}$$

Example (3-Period SMA)

Consider the data: 10, 12, 11, 15, 14.

- SMA at t=3: $(10 + 12 + 11)/3 = 11$
- SMA at t=4: $(12 + 11 + 15)/3 = 12.67$
- SMA at t=5: $(11 + 15 + 14)/3 = 13.33$

Simple Moving Average: Example

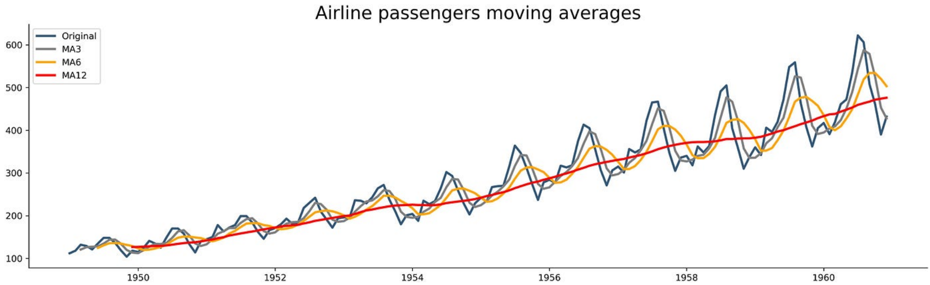


Figure: Simple Moving Average

Problems with Simple Moving Average

Despite its simplicity, SMA has two main drawbacks:

1. **Lagging Effect:** The average always "lags" behind the trend. If the data is increasing, the SMA will always be below the actual values.

2. **Equal Weighting:** It gives equal importance to all data points within the window (k), regardless of how old they are. Intuitively, the most recent data should be more relevant for forecasting.

Exponentially Weighted Moving Average (EWMA)

Intuition: EWMA addresses the shortcomings of SMA by assigning more weight to recent observations and exponentially decreasing weights to older ones.

Formula

The EWMA at time t is calculated recursively:

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

where α is the smoothing parameter ($0 < \alpha \leq 1$).

Example (EWMA with $\alpha = 0.5$)

Data: 10, 12, 11. Let's assume initial forecast $\hat{y}_0 = 10$.

- $\hat{y}_1 = 0.5 \times 10 + (1 - 0.5) \times 10 = 10$
- $\hat{y}_2 = 0.5 \times 12 + (1 - 0.5) \times 10 = 11$
- $\hat{y}_3 = 0.5 \times 11 + (1 - 0.5) \times 11 = 11$

Exponentially Weighted Moving Average (EWMA)

Why is it called **exponential**?

$$\hat{y}_t = \alpha y_t + (1-\alpha)\hat{y}_{t-1} \;;\; \hat{y}_0 = y_0$$

$$\Rightarrow \hat{y}_1 = \alpha y_1 + (1-\alpha)\hat{y}_0 = \alpha y_1 + (1-\alpha) y_0$$

$$\Rightarrow \hat{y}_2 = \alpha y_2 + (1-\alpha)\hat{y}_1 = \alpha y_2 + (1-\alpha)(\alpha y_1 + (1-\alpha) y_0)$$
$$\Rightarrow \hat{y}_2 = \alpha y_2 + \alpha(1-\alpha) y_1 + (1-\alpha)^2 y_0$$

$$\Rightarrow \hat{y}_3 = \alpha y_3 + (1-\alpha)\hat{y}_2 = \alpha y_3 + (1-\alpha)(\alpha y_2 + \alpha(1-\alpha) y_1 + (1-\alpha)^2 y_0)$$
$$\Rightarrow \hat{y}_3 = \alpha y_3 + (1-\alpha)\hat{y}_2 = \alpha y_3 + \alpha(1-\alpha) y_2 + \alpha(1-\alpha)^2 y_1 + (1-\alpha)^3 y_0$$

$$\hat{y}_t = \alpha \;\left[\sum_{i=0}^{t-1} (1-\alpha)^i y_{t-i}\right] + (1-\alpha)^t y_0$$

Significance of α

A higher α gives more weight to recent data, making the model more responsive to changes.

Adjusted EWMA

Intuition: The standard recursive EWMA has an initialization bias. To correct this, we can represent the EWMA as a weighted average where the weights for all available past points are explicitly calculated and normalized.

Adjusted EWMA Formula

The value at time t is calculated as the weighted average of all observations up to that point:

$$\hat{y}_t = \frac{\sum_{i=0}^{t} w_i y_{t-i}}{\sum_{i=0}^{t} w_i}$$

where the weight for the i^{th} observation in the past is $w_i = (1-\alpha)^i$.

Example (Adjusted EWMA with $\alpha = 0.8$)

Data: 10, 12.

- $\hat{y}_0 = \frac{(1-\alpha)^0 y_0}{(1-\alpha)^0} = \frac{1\times 10}{1} = 10.$
- $\hat{y}_1 = \frac{(1-\alpha)^1 y_0 + (1-\alpha)^0 y_1}{(1-\alpha)^1 + (1-\alpha)^0} = \frac{0.2\times 10 + 1\times 12}{0.2+1} = \frac{14}{1.2} \approx 11.67.$

Choosing α: Span

Intuition: Directly choosing α can be abstract. Instead, we can specify the smoothing factor using more intuitive concepts related to the "memory" of the moving average.

Span (S): Describes the "lookback period," similar to a simple moving average. A larger span means more smoothing (smaller α).

Formula

$$\alpha = \frac{2}{S+1}$$

Example

If Span $S = 19$, then $\alpha = 2/(19+1) = 0.1$.

Half-life (H): The time it takes for an observation's weight to decay to half its original value.

Formula

$$\alpha = 1 - \exp\left(\frac{-\ln(2)}{H}\right)$$

Example

If Half-life $H = 5$, then $\alpha = 1 - \exp(-\ln(2)/5) \approx 0.1295$.

Choosing α: Centre of Mass

Center of Mass (C): An alternative way to specify the average age of the data.

Formula

$$\alpha = \frac{1}{C + 1}$$

Example

If Center of Mass $C = 9$, then $\alpha = 1/(9 + 1) = 0.1$.

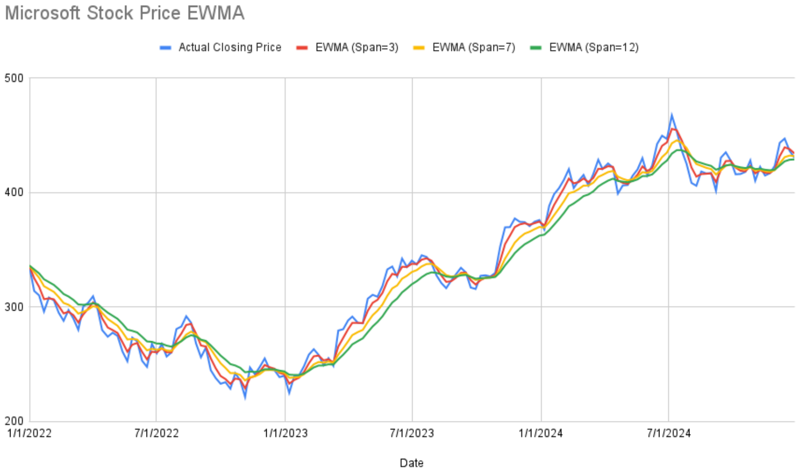


Figure: Exponentially Weighted Moving Average

Section 3

Time Series Forecasting

What is Forecasting and Why is it Hard?

Definition

Forecasting is the process of making predictions about the future based on past and present data.

Challenges and Pitfalls

Forecasting is inherently difficult because the future is uncertain. Key challenges include:

- **Randomness:** Every time series contains some level of unpredictable, random noise.
- **Changing Patterns:** The underlying trend or seasonality can change over time (e.g., due to a new product launch or a market shock).
- **Overfitting:** Creating a model that is too complex and captures the noise in the data, leading to poor performance on new data.
- **Assuming the Future is like the Past:** Models rely on historical patterns, which may not hold true in the future.

Evaluating Forecast Quality: Key Metrics

We compare the forecasted values ($\hat{y}_t$) with the actual values (y_t).

- **Mean Absolute Error (MAE):** The average of the absolute errors.

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$$

- **Root Mean Squared Error (RMSE):** Expresses the error in the original units. Penalizes larger errors more heavily.

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

- **Mean Absolute Percentage Error (MAPE):** The average percentage error. It is scale-independent.

$$\text{MAPE} = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|\right) \times 100\%$$

Metrics: A Numerical Example

Example

Suppose we have the following actual and forecasted values:

- Actuals (y_t): 10, 12, 15
- Forecasts ($\hat{y}_t$): 11, 13, 13
- Errors (e_t): -1, -1, 2

Calculate Metrics:

- **MAE:** $\frac{|-1|+|-1|+|2|}{3} = \frac{4}{3} \approx 1.33$
- **RMSE:** $\sqrt{\frac{(-1)^2+(-1)^2+(2)^2}{3}} = \sqrt{\frac{6}{3}} \approx 1.414$
- **MAPE:** $\left(\frac{1}{3}\left[\left|\frac{-1}{10}\right| + \left|\frac{-1}{12}\right| + \left|\frac{2}{15}\right|\right]\right) \times 100\%$

  $= \left(\frac{0.1+0.0833+0.1333}{3}\right) \times 100\% \approx 10.55\%$

Section 4

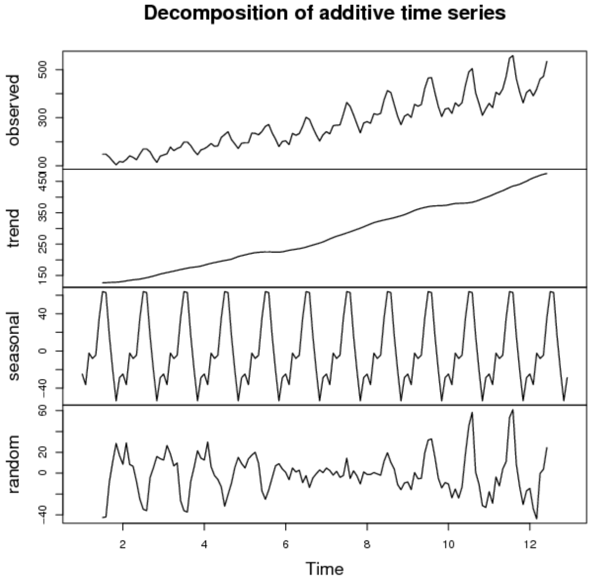
ETS Decomposition and Forecasting

ETS Models: Error, Trend, Seasonality

ETS is a family of models that explicitly model the components we've discussed.

- **E**rror: How the random fluctuations enter the model (e.g., additively).
- **T**rend: The nature of the trend (e.g., None, Additive).
- **S**easonality: The nature of the seasonal pattern (e.g., None, Additive).

This framework leads to a variety of models, such as ETS(A,N,N), which means Additive Error, No Trend, and No Seasonality. This is also known as Simple Exponential Smoothing.


Decomposition of additive time series
observed
trend
seasonal
random
100
300
500
150
250
350
450
−40
0
40
−40
0
20
60
2
4
6
8
10
12
Time

Simple Exponential Smoothing (SES)

Intuition: A simple but powerful forecasting method for data with **no clear trend or seasonality**. It's essentially the EWMA we just saw.

Model Equations

The model has two equations:

$$\text{Smoothing equation: } l_t = \alpha y_t + (1-\alpha) l_{t-1}$$
$$\text{Forecast equation: } \hat{y}_{t+h|t} = l_t$$

Where:

- l_t is the estimated **level** (or smoothed value) of the series at time t.
- α is the smoothing parameter for the level.
- The forecast for any future period ($h \geq 1$) is simply the last estimated level, l_t.

Double Exponential Smoothing (Holt's Method)

Intuition: An extension of SES for data that exhibits a **trend**. It adds a second smoothing equation to model the trend.

Model Equations

$$
\begin{aligned}
\text{Level: } & l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \\
\text{Trend: } & b_t = \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1} \\
\text{Forecast: } & \hat{y}_{t+h|t} = l_t + h \cdot b_t
\end{aligned}
$$

Where:

- l_t is the level and b_t is the **trend** (slope) at time t.
- α is the smoothing parameter for the level.
- β is the smoothing parameter for the trend.

The forecast is now a line starting at the current level and rising or falling according to the trend.

Triple Exponential Smoothing (Holt-Winters)

Intuition: Extends Holt's method to capture **seasonality**. It adds a third smoothing equation to model the seasonal component.

Model Equations (Additive Seasonality)

Let m be the seasonal period (e.g., 12 for monthly data).

$$\begin{aligned}
\text{Level: } & l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \\
\text{Trend: } & b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\
\text{Seasonal: } & s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \\
\text{Forecast: } & \hat{y}_{t+h|t} = l_t + h \cdot b_t + s_{t+h-m(k+1)}
\end{aligned}$$

Where:

- s_t is the seasonal component at time t.
- γ is the smoothing parameter for seasonality.

Section 5

Understanding Autocorrelation

Intuition: How is a time series related to its own past?

- **Autocovariance:** Measures the linear relationship between a series and a lagged version of itself. A positive value means higher values are followed by higher values.

Formula

$$\gamma_k = \mathrm{Cov}(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)]$$

Its value depends on the scale of the data, making it hard to interpret.

- **Autocorrelation (ACF):** The normalized version of autocovariance. It measures the *strength* of the relationship on a scale from -1 to 1.

Formula

$$ACF(k) = \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\mathrm{Cov}(y_t, y_{t-k})}{\mathrm{Var}(y_t)}$$

Lag-1 autocorrelation 0.8 means strong positive correlation with the previous observation.

Example

Consider the time series: $y = [10, 12, 11, 13]$. The mean $\mu = 11.5$.

1. Calculate Autocovariance at lag 0 (the Variance):

- $\gamma_0 = \frac{1}{4}\sum(y_t - \mu)^2$
- $\gamma_0 = \frac{(10-11.5)^2+(12-11.5)^2+(11-11.5)^2+(13-11.5)^2}{4}$
- $\gamma_0 = \frac{2.25+0.25+0.25+2.25}{4} = \mathbf{1.25}$

2. Calculate Autocovariance at lag 1:

- $\gamma_1 = \frac{1}{4}\sum_{t=2}^{4}(y_t - \mu)(y_{t-1} - \mu)$
- $\gamma_1 = \frac{(12-11.5)(10-11.5)+(11-11.5)(12-11.5)+(13-11.5)(11-11.5)}{4}$
- $\gamma_1 = \frac{(-0.75)+(-0.25)+(-0.75)}{4} = \mathbf{-0.4375}$

3. Calculate Autocorrelation at lag 1:

- $ACF(1) = \rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-0.4375}{1.25} = \mathbf{-0.35}$

Definition

Partial Autocorrelation Function (PACF) at lag$=k$ measures the **direct correlation** between an observation y_t and its lagged value y_{t-k} after removing the linear influence of the intermediate lags $(y_{t-1}, y_{t-2}, \ldots, y_{t-k+1})$.

Example

Consider the series: $y = [10, 12, 11, 13]$. From our previous calculation, we know the autocorrelations:

- $\rho_1 = -0.35$
- $\rho_2 = 0.3$

1. PACF at lag 1: The direct correlation with lag 1 has no intermediate lags to remove, so it's the same as the ACF.

- $\text{PACF}(1) = \rho_1 =$ **-0.35**

Example

2. PACF at lag 2: We measure the correlation between y_t and y_{t-2} after accounting for the effect of y_{t-1}.

Formula

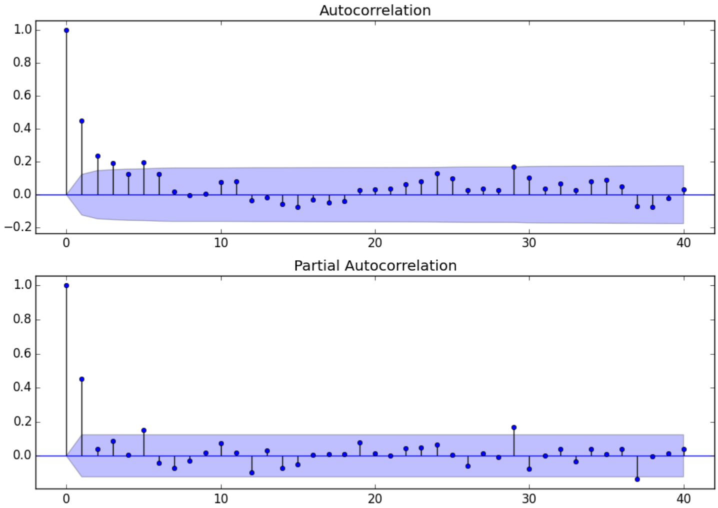
$$\text{PACF}(2) = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

- $\text{PACF}(2) = \frac{0.3-(-0.35)^2}{1-(-0.35)^2} = \frac{0.3-0.1225}{1-0.1225} = \frac{0.1775}{0.8775} \approx \mathbf{0.202}$

These are two critical diagnostic plots used in ARIMA modeling.

- **Autocorrelation Function (ACF):** A plot of the autocorrelations (ρ_k) against the lags ($k = 1, 2, ...$).
  - It shows the **total correlation** (direct and indirect) between an observation and its past values. An effect at lag 2 might be present only because lag 2 is correlated with lag 1.
- **Partial Autocorrelation Function (PACF):**
  - It measures the **direct correlation** between an observation and its value at lag k, after removing the influence of the intermediate lags ($1, 2, ..., k - 1$).

ACF and PACF Plots



Section 6

AR, MA, ARMA, and ARIMA Models

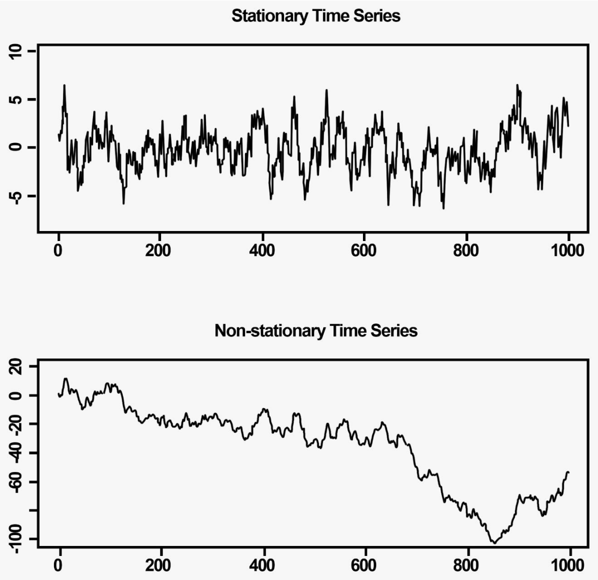
Definition

A time series is **stationary** if its statistical properties (such as mean, variance, and autocorrelation) are constant over time.

- A stationary series has no predictable long-term patterns (no trend or seasonality).
- It "hovers" around a constant mean.
- Most real-world time series are **non-stationary**.
- ARIMA models require the time series to be stationary.

Why is stationarity important? A stationary series is easier to model and forecast because its fundamental properties don't change over time.

Stationary and Non-Stationary Time Series: Example



Achieving Stationarity: Differencing

Intuition: If a series has a trend, it is non-stationary because its mean changes over time. We can make it stationary by taking the difference between consecutive observations. This is called **differencing**.

First-Order Differencing

$$y'_t = y_t - y_{t-1}$$

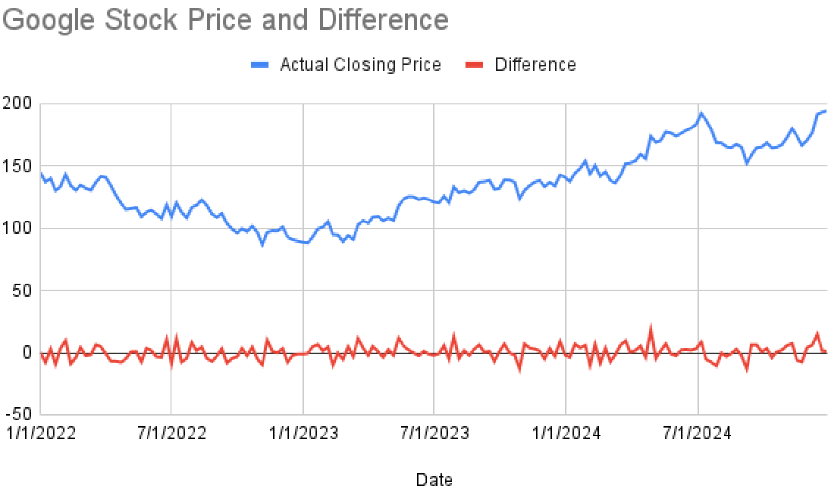
This process removes the trend, as we are now modeling the *change* from one period to the next, which is often stable.

Example

- Non-stationary series (with trend): 2, 4, 6, 8, 10, 12
- Differenced series (stationary): 2, 2, 2, 2, 2

Sometimes, we may need to difference the data more than once to achieve stationarity (second-order differencing: $y''_t = y'_t - y'_{t-1}$).

Differencing: Example



Detecting the stationarity in Time Series

- **Statistical Tests**
  - Most of the times identifying the stationarity of a time series can't be done just by visualizing it.
  - Hence, we need to perform Statistical Tests in order to check if the time series is stationary.
- **The Dickey Fuller Tests**
  - The Dickey-Fuller test was the first statistical test developed to test the null hypothesis that a unit root is present in an autoregressive model of a given time series, and that the process is thus not stationary. The original test treats the case of a simple lag-1 AR model.
- **The KPSS Tests**
  - Another prominent test for the presence of a unit root is the KPSS test. [Kwiatkowski et al, 1992] Conversely to the Dickey-Fuller family of tests, the null hypothesis assumes stationarity around a mean or a linear trend, while the alternative is the presence of a unit root.
- There are other tests to determine the stationarity of time series models.

Intuition: The idea is that the current value of the series, y_t, can be explained as a function of its past values. It's like a regression of the series on itself.

AR(p) Model

An autoregressive model of order p, denoted AR(p), is written as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t$$

Where:

- p is the order (number of past values or "lags" used).
- ϕ_i are the model parameters.
- ϵ_t is white noise (random error).
- c is the intercept (or drift) of the model.

AR(0) and AR(1) process

- The simplest of AR process is **AR(0)**, which has no dependence between the terms. In fact, **AR(0)** is just random residual error (also called **white noise**)

- **AR(1)** equation: $y_t = c + \phi_1 y_{t-1} + \epsilon_t$
  - Only the previous term in the process and the noise term contribute to the output.
  - If $|\phi_1|$ is close to 0, then the process still looks like a white noise.
  - If $\phi_1 < 0$ and $c = 0$ the y_t tends to oscillate between positive and negative values.
  - If $\phi_1 = 1$ then the process is equivalent to a random walk.
  - For a stationary autoregressive time series, the absolute values of the model parameters should always be less than 1. i.e. $|\phi_i| < 1$.

Moving Average (MA) Models

Intuition: The current value y_t is a function of past *forecast errors* (ϵ_t). It assumes that the random shocks or errors from the past influence the present observation.

MA(q) Model

A moving average model of order q, denoted MA(q), is written as:

$$y_t = c + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + ... + \theta_q\epsilon_{t-q}$$

Where:

- q is the order (number of past error terms used).
- θ_i are the model parameters.
- ϵ_t is white noise.
- c is the intercept (or drift). In case of MA process, c is exactly the mean of the process.

Note: This is different from the Simple Moving Average for smoothing!

Moving Average (MA) Models

- Although the equation of moving average (MA) model looks like a regression model, the difference is that the ϵ_t is not observable.
- Contrary to AR model, finite MA model is always stationary. (Yes, AR process can be non-stationary, for example: random walk)
- Finite MA models are stationary because the observation is just a weighted moving average over past forecast errors.
- The simplest MA process is **MA(0)**: $y_t = c + \epsilon_t$, which is just random residual error (**white noise**), same as AR(0) process.
- **MA(1)** equation: $y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1}$

ARMA Model

An ARMA(p, q) model simply combines the AR(p) and MA(q) components. It models the series using both past values and past errors.

$$y_t = c + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} + \epsilon_t$$

ARIMA Model

An ARIMA(p, d, q) model is the complete package. It combines ARMA with differencing.

- **AR(p):** Autoregressive part.
- **I(d):** Integrated part. 'd' is the degree of differencing required to make the series stationary.
- **MA(q):** Moving Average part.

It is one of the most widely used models for forecasting non-stationary time series.

Identifying Model Orders with ACF/PACF

For a **stationary** time series, the patterns in the ACF and PACF plots help us choose the order for AR and MA models.

AR(p) Model

- **ACF:** Tails off gradually.
- **PACF:** Cuts off sharply after lag **p**.

MA(q) Model

- **ACF:** Cuts off sharply after lag **q**.
- **PACF:** Tails off gradually.

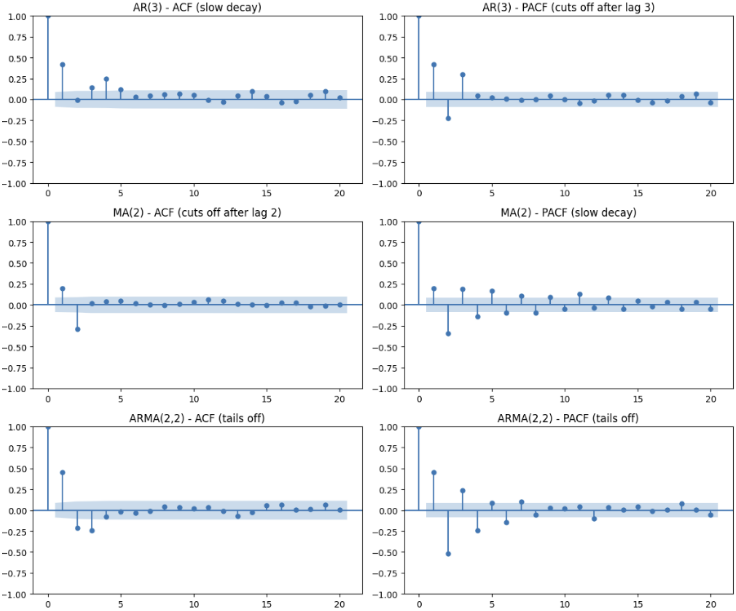
Conclusion

Use the plot that "cuts off" to identify the model order. If the PACF cuts off at lag p, you likely have an AR(p) model. If the ACF cuts off at lag q, you likely have an MA(q) model.

ARMA(p,q) Model

- **ACF:** Tails off gradually.
- **PACF:** Tails off gradually.

Identifying Model Orders with ACF/PACF: Example



ETS Models

- Based on decomposing the series into **Trend** and **Seasonality**.
- Intuitive and easy to interpret the components.
- Can handle different types of trends and seasonality.
- Often performs very well in practice.

ARIMA Models

- Based on the idea of **autocorrelation**.
- Requires the series to be **stationary**, achieved via differencing.
- More flexible in modeling complex correlation structures.
- Can be more difficult to select the right model (finding p, d, q).

Key Takeaway: They are two different but powerful philosophies for modeling time series. Neither is universally better than the other.

Section 7

Seasonal ARIMA (SARIMA) Models

Seasonal ARIMA (SARIMA)

Intuition: ARIMA models have trouble with seasonal data because the correlation at a seasonal lag (e.g., lag 12 for monthly data) is not captured by the standard AR and MA terms.

The SARIMA model extends ARIMA by adding seasonal components.

SARIMA(p, d, q)(P, D, Q)m Model

- **(p, d, q):** The non-seasonal part of the model (just like ARIMA).
- **(P, D, Q)m:** The *seasonal* part of the model.
  - **P:** Seasonal AR order.
  - **D:** Seasonal differencing.
  - **Q:** Seasonal MA order.
  - **m:** The seasonal period (e.g., m=12 for monthly, m=4 for quarterly).

This model can handle both trend and seasonality by applying differencing at both the non-seasonal and seasonal levels.

Questions?

References

Hyndman, R. J., & Athanasopoulos, G. (2021).
Forecasting: principles and practice, 3rd edition.
OTexts: Melbourne, Australia.
https://otexts.com/fpp3