Time Series Analysis

Concepts, Methods, and Applications

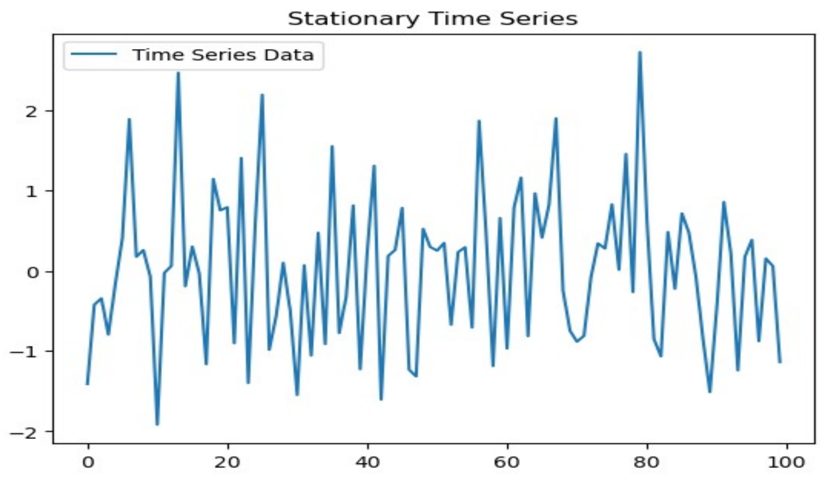
Abdellah El Fallahi

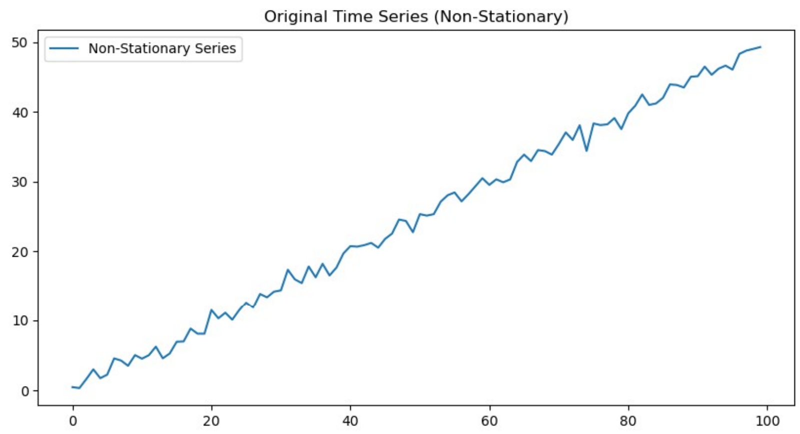
2020-2026

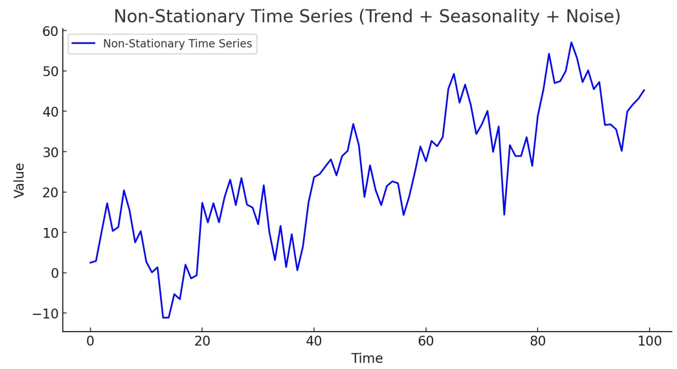
What is Time Series Analysis?

- A **time series** is a sequence of data points collected or recorded at successive time intervals. It represents how a variable changes over time.

Type of Time series







Vérifier la stationnarité (avant tout)

Avant de choisir un modèle, on confirme que la série est stationnaire.

- Tests statistiques :
  - Test ADF (Augmented Dickey-Fuller)
  - Test KPSS
  - Test PP (Phillips–Perron)
- Analyse graphique :
  - Série sans tendance
  - ACF décroît rapidement
  - Variance constante

Méthodes classiques pour séries stationnaires

- **Modèles linéaires**
- 👉 **Les plus adaptés et les plus utilisés**
- **AR(p)** – AutoRegressive
- **MA(q)** – Moving Average
- **ARMA(p, q)** – combinaison AR et MA
- ✔ Conditions :
- Série stationnaire
- Pas de saisonnalité
- Choix de p et q :
- ACF → MA(q)
- PACF → AR(p)
- AIC / BIC pour validation

Lissage et filtrage

Utiles pour débruiter une série stationnaire.

- Moyenne mobile centrée
- Filtre de Wiener
- Filtre passe-bas / passe-haut
- Filtre de Kalman (cas simple)

Méthodes fréquentielles

- Quand la série stationnaire contient des cycles.
- Analyse spectrale
- Périodogramme
- Transformée de Fourier (FFT)

Méthodes Machine Learning (si stationnaire)

- La stationnarité facilite l’apprentissage.
- Régression linéaire avec lags
- Support Vector Regression (SVR)
- Random Forest
- kNN
- Variables explicatives typiques :
  - $y_{t-1}, y_{t-2}, \ldots$
- rolling mean / rolling std

Séries stationnaires mais bruitées

Pour modéliser le bruit :

- White Noise
- ARMA de faible ordre
- Modèles GARCH *(si variance conditionnelle variable)*

Cas particulier : variance non constante

La moyenne est stationnaire mais la variance change.

- **ARCH / GARCH**
- Très utilisé en finance (volatilité)

Tableau récapitulatif

Caractéristique	Méthode recommandée
Série stationnaire simple	AR / MA / ARMA
Bruit blanc	White Noise
Cycles	Analyse spectrale
Variance variable	GARCH
Relations non linéaires	ML
Prévision court terme	ARMA

Comment rendre une séries stationnaire

Identifier pourquoi la série n'est pas stationnaire

Avant toute transformation, on **diagnostique le problème** :

- Tendance (croissante / décroissante)
- Saisonnière
- Variance non constante
- Tests statistiques (ADF, KPSS)

Éliminer la tendance

a) Différenciation (méthode la plus utilisée)

- On calcule :
- $Z_t = y_t - y_{t-1}$ Différence d'ordre 1 → supprime tendance linéaire
- Différence d'ordre 2 → tendance plus complexe
  - $Z_t = y_t - 2y_{t-1} + y_{t-2}$
- Très utilisée dans **ARIMA**

b) Suppression explicite de la tendance

- Régression linéaire sur le temps
- On garde les résidus

Éliminer la saisonnalité

a) Différenciation saisonnière

- $Z_t = y_t - y_{t-s}$ où s est la période (12, 4, 7, …)

b) Décomposition

  - Additive ou multiplicative
  - On retire la composante saisonnière

Stabiliser la variance

- Si la variance augmente avec le temps 👇
- **Transformations courantes**
  - Logarithme : $\log(y_t)$
  - Racine carrée : sqrt{y_t}
  - Box-Cox
- 📌 Très fréquent en économie et finance

Vérifier la stationnarité (obligatoire)

Après transformation :

- **ADF** : doit rejeter l'hypothèse de racine unitaire
- **KPSS** : doit accepter l'hypothèse de stationnarité
- ACF décroît rapidement
- Série oscille autour d'une moyenne constante

Cas pratiques (quoi faire ?)

Problème détecté	Solution
Tendance	Différenciation
Saisonnière	Différenciation saisonnière
Variance non constante	Log / Box-Cox
Tendance + saison	Diff + diff saisonnière
Bruit fort	Lissage

Rproject:

```
y_diff <- diff(y)
y_season <- diff(y, lag=12)
adf.test(y_diff)
```

Python

```
y_diff = y.diff().dropna()adfuller(y_diff)
```

Importance of Time Series Analysis

Studying time series is important for several reasons, especially in fields like finance, economics, engineering, and science.

1. Understanding Trends and Patterns

- Time series analysis helps identify long-term trends, seasonal patterns, and cyclical behaviors in data. This is useful for making informed decisions based on historical data.

2. Forecasting and Prediction

- Time series models are used to predict future values based on past data. This is essential in stock market forecasting, weather prediction, demand forecasting, and many other applications.

3. Anomaly Detection

- Time series analysis helps detect unusual patterns or anomalies, which is important in fraud detection, network security, and quality control.

4. Causal Relationships and Dependencies

- It helps in understanding how different factors influence each other over time. For example, in economics, time series analysis can determine how interest rates affect inflation.

Importance of Time Series Analysis

5. Control and Optimization

- Time series models are used in control systems and optimization problems, such as managing inventory in supply chain management or adjusting parameters in industrial processes.

6. Risk Management

- In finance and insurance, time series models help assess risks, volatility, and uncertainties in markets.

7. Policy Making and Strategic Planning

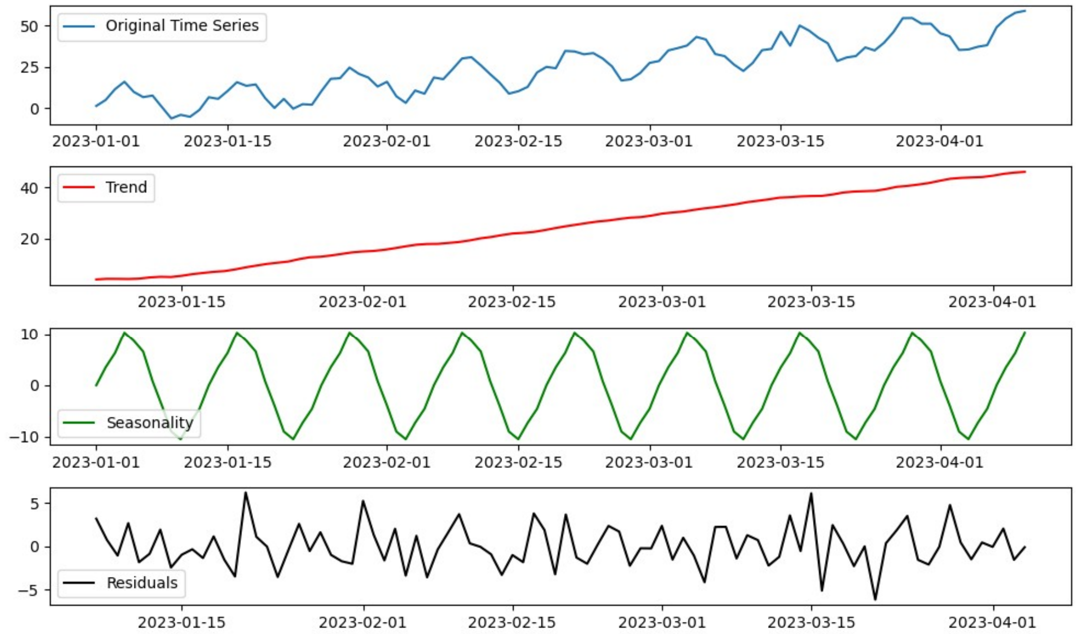
- Governments and businesses use time series analysis to make data-driven policy and strategic decisions.

8. Improving Machine Learning Models

- Time series data is widely used in machine learning for applications like speech recognition, energy consumption forecasting, and healthcare analytics.

Components of Time Series

- 1. Trend
- 2. Seasonality
- 3. Cyclic Patterns
- 4. Random Noise



Examples of Time Series Data

- Stock Prices,
- Weather Data,
- Sales Figures,
- Economic Indicators

Data Collection & Sources

- Financial,
- Weather,
- Economic,
- and Business Data

Data Preprocessing in time series

- Handling missing values,
- smoothing techniques,
- transformations.

Méthodes descriptives & exploratoires

- Décomposition :
- Tendance (Trend)
- Saisonnière (Seasonality)
- Résiduelle (Noise)
- Moyennes mobiles
- Lissage exponentiel simple
- Analyse graphique (plots, seasonal plots)
- Autocorrélation (ACF) et autocorrélation partielle (PACF)

Méthodes statistiques classiques

- **Modèles linéaires**
  - AR (AutoRegressive)
  - MA (Moving Average)
  - ARMA
  - ARIMA
  - SARIMA (avec saisonnalité)
  - SARIMAX (avec variables exogènes)
- **Lissage exponentiel**
  - Simple Exponential Smoothing
  - Holt (trend)
  - Holt-Winters (trend + saisonnalité)

Méthodes probabilistes et état-espace

Adaptées aux systèmes dynamiques.

- Modèles état-espace
- Filtre de Kalman
- Modèles bayésiens dynamiques
- Hidden Markov Models (HMM)

Méthodes de Machine Learning

⚠ Utilisées quand les relations sont **non linéaires**.

- Régression linéaire avec features temporelles
- Random Forest
- Gradient Boosting (XGBoost, LightGBM)
- Support Vector Regression (SVR)
- k-Nearest Neighbors (kNN)

⚠ Nécessitent souvent du **feature engineering** :

- Lags
- Rolling mean
- Rolling std
- Variables calendaires (jour, mois, fêtes…)

Deep Learning pour séries temporelles

Très efficaces pour grandes bases de données.

- RNN (Recurrent Neural Networks)
- LSTM
- GRU
- CNN 1D pour séries temporelles
- Transformers pour time series
- Temporal Fusion Transformer (TFT)

Méthodes de détection d'anomalies

Pour surveillance et maintenance prédictive.

- Z-score
- IQR
- Isolation Forest
- Autoencoders
- STL + détection sur résidus

Méthodes hybrides

- Combinaison de plusieurs approches.
- ARIMA + Neural Networks
- Décomposition + ML
- Statistical model + Deep Learning

Choisir la bonne méthode

Situation	Méthode recommandée
Série courte, régulière	ARIMA / Holt-Winters
Forte saisonnalité	SARIMA / STL
Non-linéarité	ML / LSTM
Données massives	Deep Learning
Explicabilité requise	Modèles statistiques
Temps réel	Kalman / online learning

Outils & bibliothèques courantes

- **R** : forecast, tsibble, fable, TTR
- **Python** : statsmodels, prophet, sktime, tsfresh, tensorflow, pytorch

Visualization of Time Series Data

- Line plots,
- Seasonal decomposition,
- Trends.

Smoothing models

- **Traditional Statistical Models**
  - **ARIMA (AutoRegressive Integrated Moving Average)**: A widely used model for forecasting stationary time series.
  - **SARIMA (Seasonal ARIMA)**: An extension of ARIMA that captures seasonality.
  - **VAR (Vector Autoregression)**: Used for multivariate time series forecasting.
  - **Holt-Winters (Exponential Smoothing)**: Effective for time series with trend and seasonality.

Smoothing models: Machine Learning Models

- **Random Forest / XGBoost / LightGBM**: These tree-based models can handle time series forecasting by treating the problem as a supervised learning task with lag features.
- **Support Vector Regression (SVR)**: Can be applied to time series forecasting but requires feature engineering.

Smoothing Techniques: Deep Learning Models

- **Recurrent Neural Networks (RNNs)**: Designed to handle sequential data.
- **Long Short-Term Memory (LSTM)**: A type of RNN that can capture long-term dependencies in time series.
- **Gated Recurrent Units (GRU)**: Similar to LSTMs but computationally more efficient.
- **Temporal Convolutional Networks (TCN)**: Uses convolutional layers instead of recurrent layers to capture time dependencies.
- **Transformers (e.g., Time Series Transformer, Informer)**: Adapted from NLP to handle long-range dependencies in time series data.

Hybrid and Specialized Models

- **Facebook Prophet**: Designed for business forecasting with automatic trend and seasonality detection.
- **DeepAR (Amazon)**: A probabilistic forecasting model using deep learning.
- **N-BEATS (Neural Basis Expansion Analysis for Time Series)**: A deep learning model specifically designed for time series forecasting.
- **Neural ODEs (Ordinary Differential Equations)**: A newer approach to modeling time series using continuous-time representations.

Autoregressive (AR) Model

- AR models the relationship between a variable and its past values.

AR(p) Model Equation:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t$$

where:

- X_t = The observed time series at time t
- c = Constant term (optional)
- ϕ_i = Autoregressive coefficients (parameters)
- p = Order of the AR process (number of lagged terms)
- $\epsilon_t \sim N(0, \sigma^2)$ = White noise (random error)

Moving Average (MA) Model

- MA captures the relationship between a variable and past errors.

MA(q) Model Equation:

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

where:

- X_t = The observed time series at time t
- μ = Mean of the series (optional, often assumed to be 0 for simplicity)
- $\epsilon_t \sim N(0, \sigma^2)$ = White noise error terms
- θ_i = MA coefficients (parameters)
- q = Order of the MA process (number of lagged error terms included)

ARMA (Autoregressive Moving Average)

- Combination of AR(p) and MA(q) models.

ARMA(p, q) Model Equation:

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

where:

- X_t = Observed time series at time t
- c = Constant term (optional)
- ϕ_i = AR coefficients (how past values influence the present)
- θ_j = MA coefficients (how past error terms influence the present)
- p = Order of the **AR** part (number of lagged observations)
- q = Order of the **MA** part (number of lagged error terms)
- $\epsilon_t \sim N(0, \sigma^2)$ = White noise error term

ARIMA (Autoregressive Integrated Moving Average)

- ARIMA = AR + Differencing + MA.

ARIMA(p, d, q) Model Equation

$$\Phi_p(B)(1-B)^d X_t = c + \Theta_q(B)\epsilon_t$$

where:

- X_t = Observed time series at time t
- B = **Backshift operator** ($BX_t = X_{t-1}$)
- d = **Degree of differencing** (number of times differencing is applied to make the series stationary)
- p = **AR order** (number of lagged observations)
- q = **MA order** (number of lagged error terms)
- $\Phi_p(B)$ = Autoregressive polynomial $(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)$
- $\Theta_q(B)$ = Moving Average polynomial $(1 + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q)$
- $\epsilon_t \sim N(0, \sigma^2)$ = White noise

ACF and PACF Interpretation

Identification through ACF and PACF

Definition

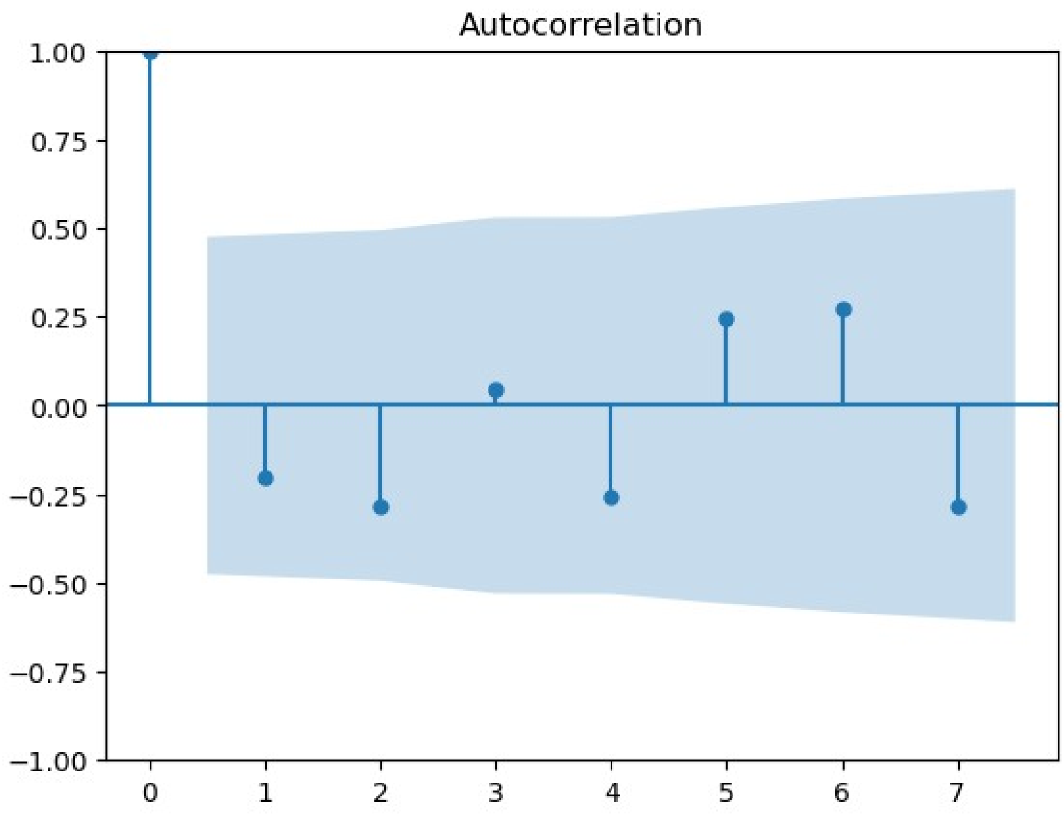
Model	*ACF*	*PACF*
$AR(1)$	Geometric Decay	Cutoff after Lag 1
$AR(p)$	Geometric Decay	Cutoff after Lag p
$MA(1)$	Cutoff after Lag 1	Geometric Decay
$MA(q)$	Cutoff after Lag q	Geometric Decay
$ARMA(p, q)$	Geometric Decay	Geometric Decay

ARMA model's parameters

- The **Autocorrelation Function (ACF)** and **Partial Autocorrelation Function (PACF)** help in selecting the appropriate values for **p** (AR order) and **q** (MA order) in an **ARMA(p, q) model**.

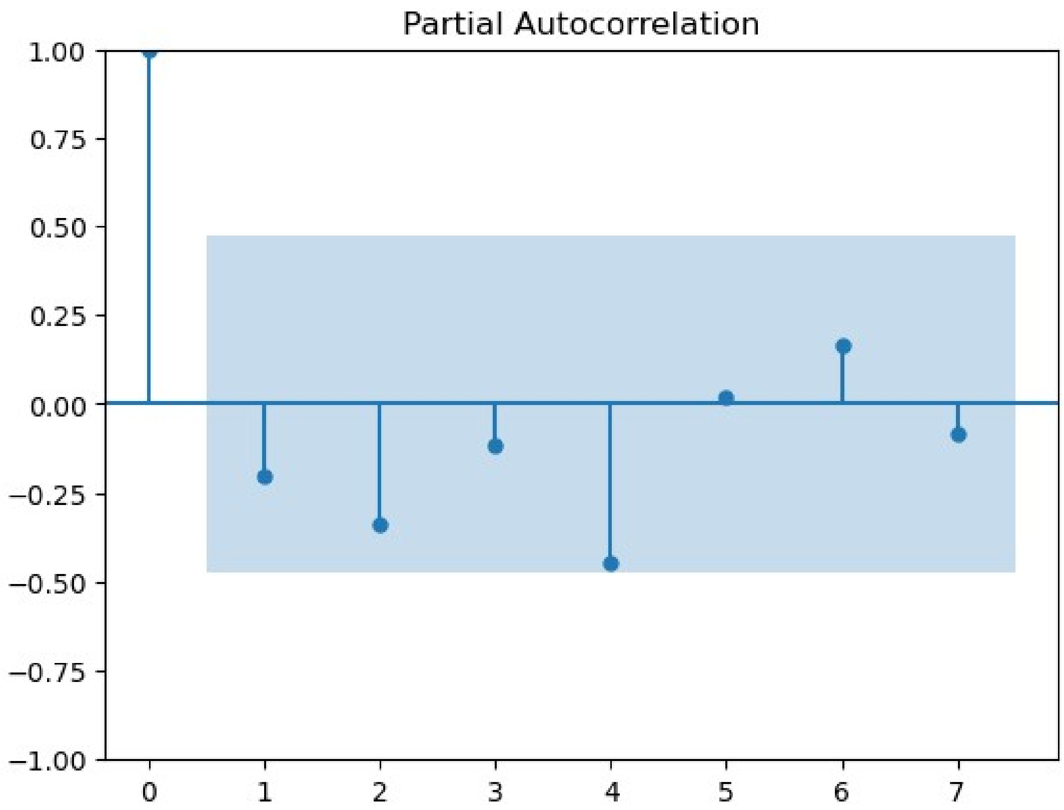
ACF

Autocorrelation Function (ACF): Measures the correlation between a time series and its past values at different lags. It helps to determine the order of a MA model



PACF

Partial Autocorrelation Function (PACF): Measures the direct correlation between a time series and its past values, removing the effects of intermediate lags. It hepls to dertmine the ordre of a AR model



Model Selection Based on ACF & PACF

- Examine ACF & PACF plots to determine if the time series follows **AR(p)**, **MA(q)**, or **ARMA(p, q)**.
- Fit different ARMA models and compare **AIC/BIC** for the best model.

AIC and BIC

- **AIC (Akaike Information Criterion):** Measures the relative quality of a statistical model, penalizing overfitting.
- **BIC (Bayesian Information Criterion):** Similar to AIC but penalizes complexity more strictly.
- The AIC and BIXC are used to compare two models, the model with small AIC and BIC is better

AIC and BIC

- The **Akaike Information Criterion (AIC)** and **Bayesian Information Criterion (BIC)** are used to assess the goodness of fit of statistical models while penalizing model complexity.

AIC

1. **Akaike Information Criterion (AIC)**

$$AIC = -2\ln(L) + 2k$$

- L = Maximum likelihood of the model
- k = Number of estimated parameters
- A lower **AIC** value indicates a better-fitting model.

$$\mathrm{AIC} = n \cdot \ln\left(\frac{\mathrm{SSE}}{n}\right) + 2k$$

Where:

- n = Number of observations (data points).
- SSE = Sum of Squared Errors or Residual Sum of Squares (RSS).
- k = Number of parameters (model parameters including the intercept).

BIC

2. **Bayesian Information Criterion (BIC)**

$$BIC = -2\ln(L) + k\ln(n)$$

- n = Number of observations
- k = Number of estimated parameters
- BIC penalizes complexity more than AIC, favoring simpler models when sample size n is large.

BIC (Bayesian Information Criterion):

$$\mathrm{BIC} = n \cdot \ln\left(\frac{\mathrm{SSE}}{n}\right) + k \cdot \ln(n)$$

Where:

- n = Number of observations.
- SSE = Sum of Squared Errors (RSS).
- k = Number of model parameters.
- $\ln$ = Natural logarithm.

Mathematical formulation of SSE and RSS

Mathematical Formulation of SSE and RSS

Given a set of **observed values** $y_1, y_2, \ldots, y_n$ and the **predicted values** from a model $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$, the **SSE** (or **RSS**) is calculated as the sum of the squared differences between each observed value and its corresponding predicted value:

$$\text{SSE} = \text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

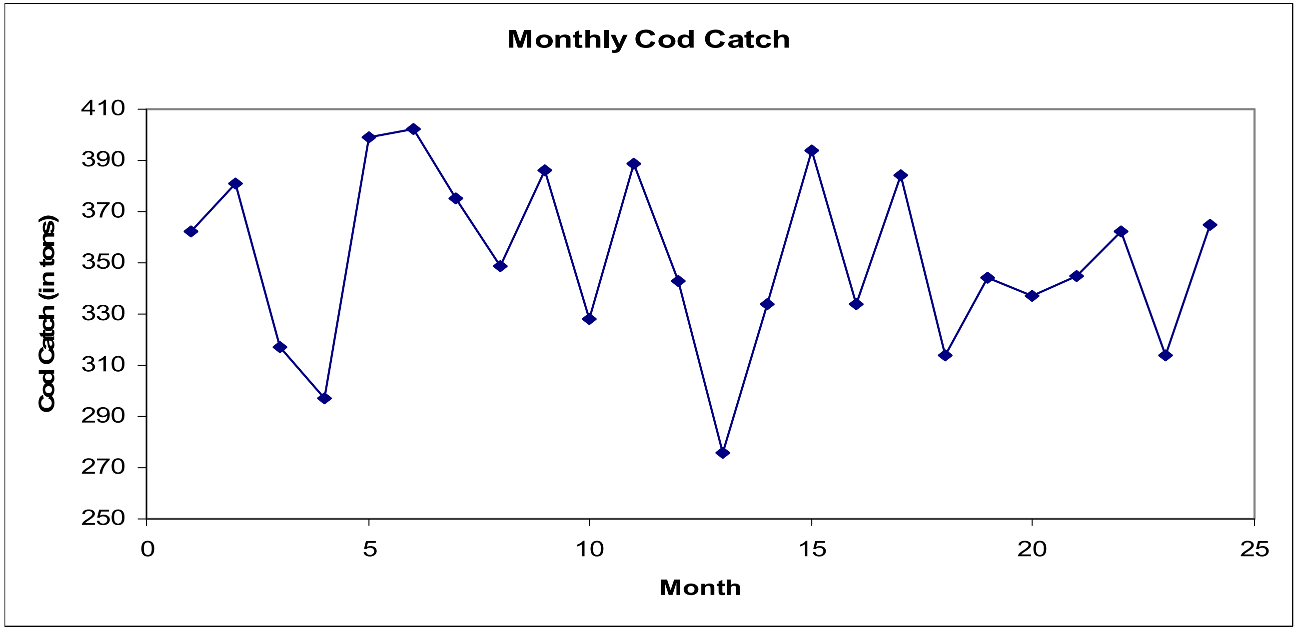
Where:

- y_i = The observed value (actual data point) at time i.
- $\hat{y}_i$ = The predicted value from the model at time i.
- n = The number of observations in the dataset.

Selection Models

- Model seclection based on:
  - p-values
  - Residual sum of squares and R^2=(TSS-RSS)/TSS
  - AIC
  - BIC
  - AICc

Simple Exponential Smoothing



Simple Exponential Smoothing

$F_{t+1} = F_t + \alpha(Y_t - F_t)$

$F_{t+1} = (1-\alpha)F_t + \alpha Y_t$

Where:

F_{t+1} is the forecasting of the period t+1

F_t: forecasting of the period t

Y_t: real observation of period t

α: *smoothing constant* in the range [0,1]

Holt's Trend Corrected Exponential Smoothing

- If a time series is increasing or decreasing approximately at a fixed rate, then it may be described by the LINEAR TREND model

  $$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

  If the values of the parameters β_0 and β_1 are slowly changing over time, Holt's trend corrected exponential smoothing method can be applied to the time series observations.

  <u>Note</u>: When neither β_0 nor β_1 is changing over time, regression can be used to forecast future values of y_t.

- Level (or mean) at time T: $\beta_0 + \beta_1 T$

  Growth rate (or trend): β_1

Holt's Trend Corrected Exponential Smoothing

- Holt's model Augments SES by capturing a trend component
- Series has Levl (l_t)
- Trend b_t
- Noise : Unpredictable
- Forecast = estimated level + trend at most recent time point
  - $F_{t+k} = l_t + kb_t$

Updating the Level and trend

- Level estimate

$$\ell_T = \alpha y_T + (1-\alpha)(\ell_{T-1} + b_{T-1})$$

- Trend estimate

$$b_T = \gamma(\ell_T - \ell_{T-1}) + (1-\gamma)b_{T-1}$$

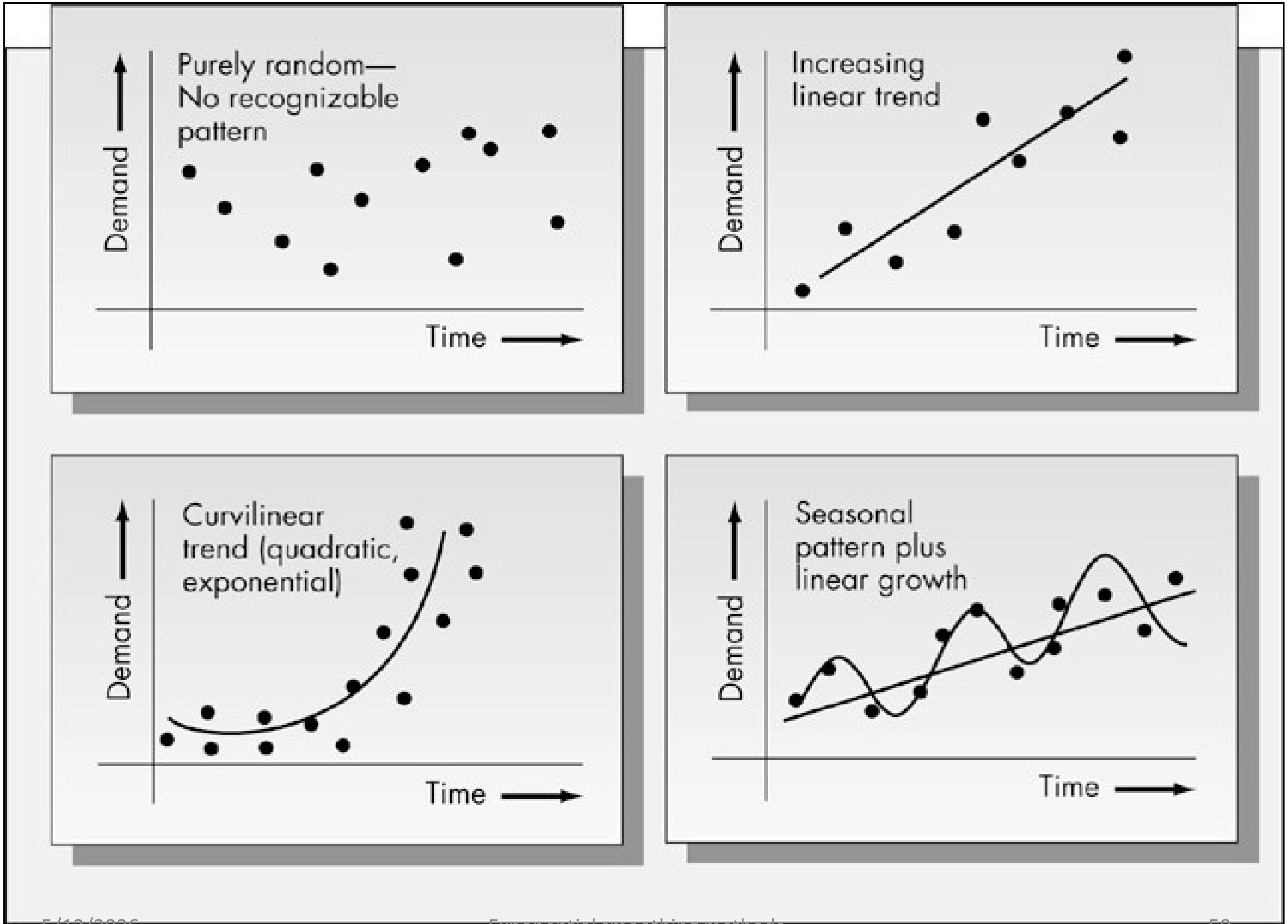
where α = smoothing constant for the level ($0 \le \alpha \le 1$)

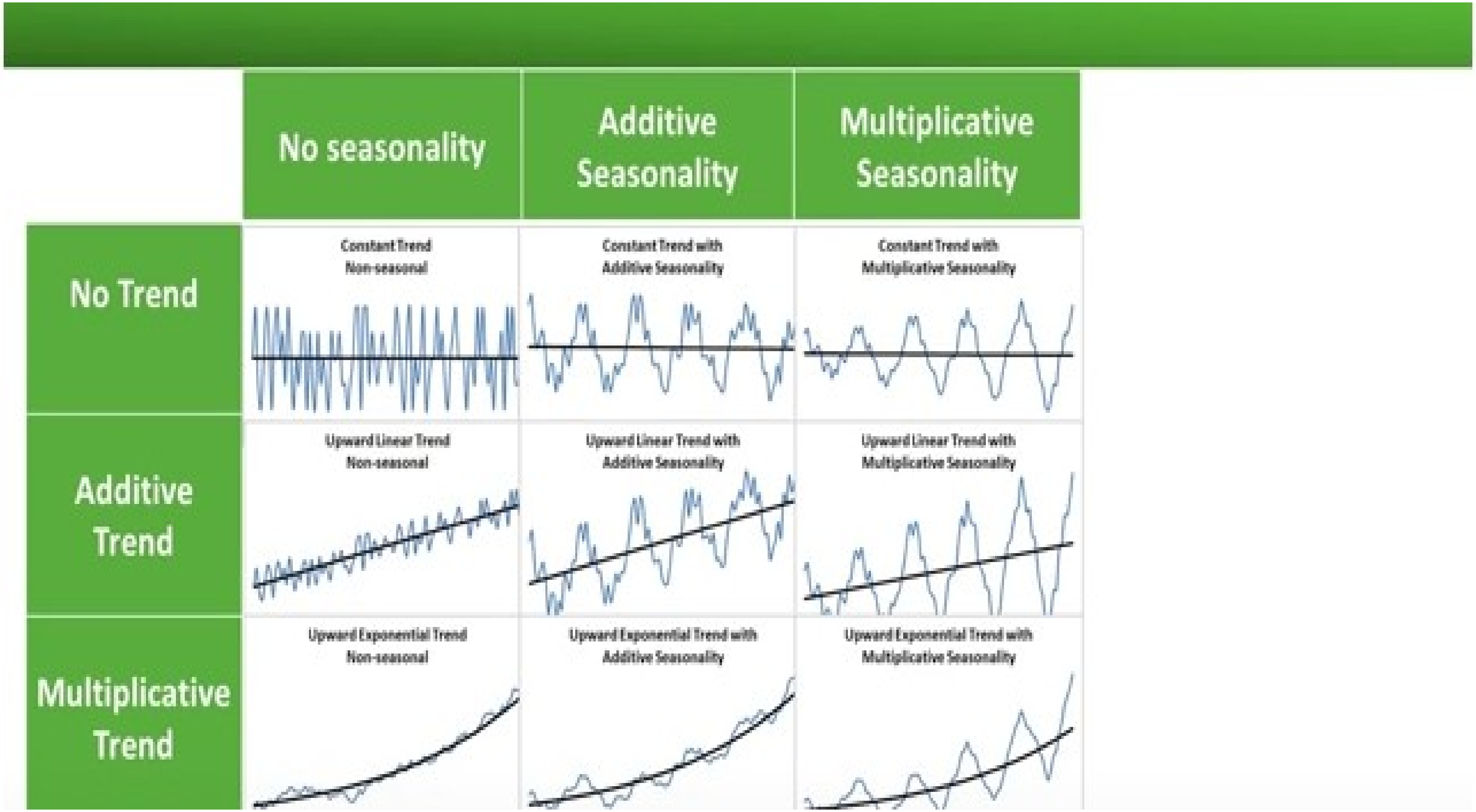
γ = smoothing constant for the trend ($0 \le \gamma \le 1$)

Hot’s model forecasting

- Additive Trend
- $F_{t+k} = l_t + k\, b_t$

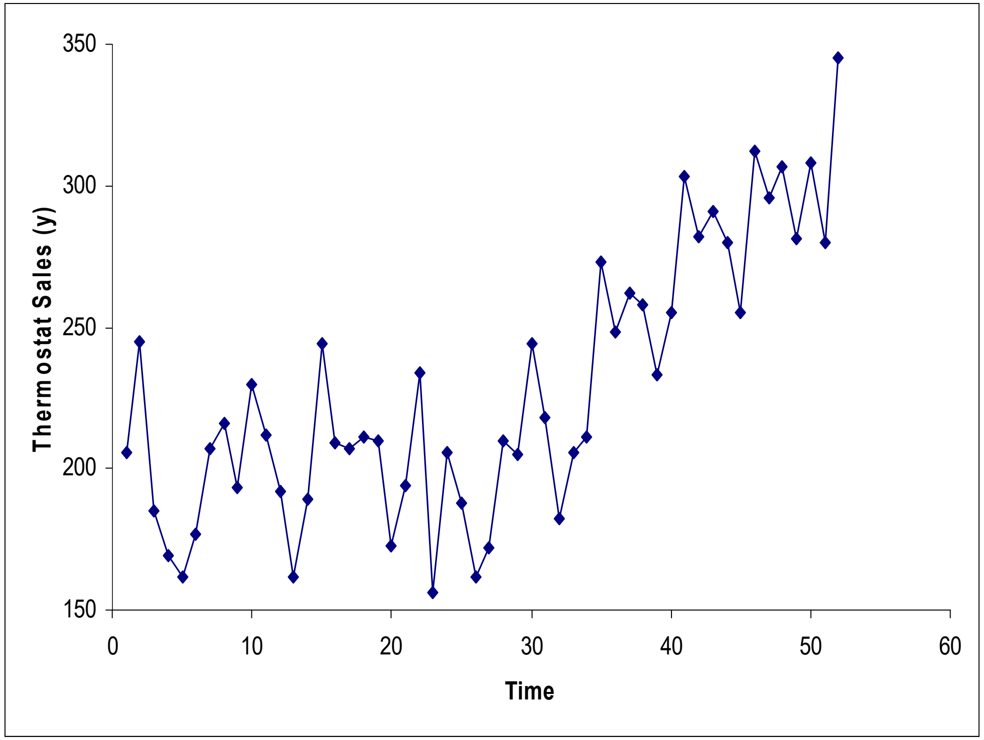
- Multiplicative model:
- $F_{t+k} = l_t * (b_t)^k$
- Initialization:
  - l_0= intercept of the linear regression (others values can be considered)
  - b_0= trend of the linear regression (others values can be considered)


Purely random—
No recognizable
pattern
Demand
Time
Increasing
linear trend
Demand
Time
Curvilinear
trend (quadratic,
exponential)
Demand
Time
Seasonal
pattern plus
linear growth
Demand
Time


No seasonality
Additive Seasonality
Multiplicative Seasonality
No Trend
Additive Trend
Multiplicative Trend
Constant Trend Non-seasonal
Constant Trend with Additive Seasonality
Constant Trend with Multiplicative Seasonality
Upward Linear Trend Non-seasonal
Upward Linear Trend with Additive Seasonality
Upward Linear Trend with Multiplicative Seasonality
Upward Exponential Trend Non-seasonal
Upward Exponential Trend with Additive Seasonality
Upward Exponential Trend with Multiplicative Seasonality

Holt-Winter's model (Triple exponential smoothing)

- Augment holt's method by capturing the seasonal component



Holt-Winters Methods

- Two Holt-Winters methods are designed for time series that exhibit linear trend
  - Additive Holt-Winters method: used for time series with constant (additive) seasonal variations
  - Multiplicative Holt-Winters method: used for time series with increasing (multiplicative) seasonal variations
- Holt-Winters method is an exponential smoothing approach for handling SEASONAL data.
- The multiplicative Holt-Winters method is the better known of the two methods.

Multiplicative Holt-Winters Method

- It is generally considered to be best suited to forecasting time series that can be described by the equation:

$$y_t = (\beta_0 + \beta_1 t) \times SN_t \times IR_t$$

  - SN_t: seasonal pattern
  - IR_t: irregular component
- This method is appropriate when a time series has a linear trend with a multiplicative seasonal pattern for which the level ($\beta_0 + \beta_1 t$), growth rate (β_1), and the seasonal pattern (SN_t) may be slowly changing over time.

Multiplicative Holt-Winters Method

- Estimate of the level

$$\ell_T = \alpha(y_T / sn_{T-L}) + (1-\alpha)(\ell_{T-1} + b_{T-1})$$

- Estimate of the growth rate (or trend)

$$b_T = \gamma(\ell_T - \ell_{T-1}) + (1-\gamma)b_{T-1}$$

- Estimate of the seasonal factor

$$sn_T = \delta(y_T / \ell_T) + (1-\delta)sn_{T-L}$$

where α, γ, and δ are smoothing constants between 0 and 1,

L = number of seasons in a year (*L* = 12 for monthly data, and *L* = 4 for quarterly data)

Multiplicative Holt-Winters Method

- Point forecast made at time *T* for y_{T+p}

$$\hat{y}_{T+p}(T) = (\ell_T + pb_T)sn_{T+p-L} \qquad (p = 1, 2, 3, \ldots)$$

- MSE and the standard errors at time *T*

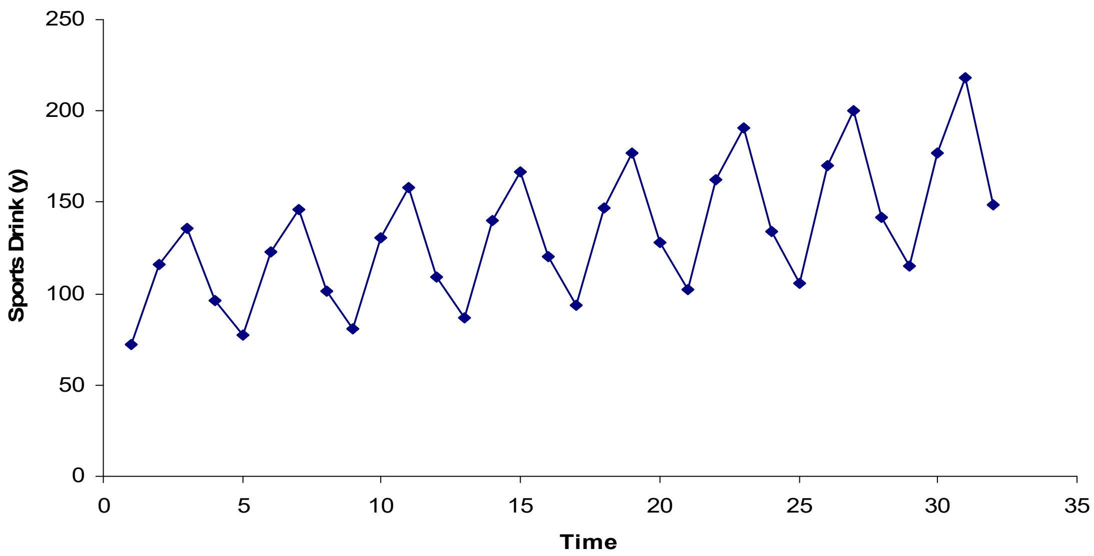
$$SSE = \sum_{t=1}^{T} [y_t - \hat{y}_t(t-1)]^2$$

$$MSE = \frac{SSE}{T-3}, \qquad s = \sqrt{MSE}$$

Procedures of Multiplicative Holt-Winters Method

- Use the Sports Drink example as an illustration

Quarterly sales of Tiger Sports Drink								
	Year							
Quarter	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	72	77	81	87	94	102	106	115
2	116	123	131	140	147	162	170	177
3	136	146	158	167	177	191	200	218
4	96	101	109	120	128	134	142	149



Procedures of Multiplicative Holt-Winters Method

- Observations:
  - Linear upward trend over the 8-year period
  - Magnitude of the seasonal span increases as the level of the time series increases

  $\Rightarrow$ Multiplicative Holt-Winters method can be applied to forecast future sales

Procedures of Multiplicative Holt-Winters Method

- **Step 1**: Obtain initial values for the level ℓ_0, the growth rate b_0, and the seasonal factors sn_{-3}, sn_{-2}, sn_{-1}, and sn_0, by fitting a least squares trend line to at least four or five years of the historical data.
  - y-intercept = ℓ_0; slope = b_0

Procedures of Multiplicative Holt-Winters Method

- Example
  - Fit a least squares trend line to the first 16 observations
  - Trend line

$$\hat{y}_t = 95.2500 + 2.4706t$$

  - $\ell_0 = 95.2500$; $b_0 = 2.4706$

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.403809754
R Square	0.163062318
Adjusted R Square	0.103281055
Standard Error	27.58325823
Observations	16
ANOVA	
	df
Regression	1
Residual	14
Total	15
	Coefficients
Intercept	95.25
X Variable 1	2.470588235

Procedures of Multiplicative Holt-Winters Method

- **Step 2**: Find the initial seasonal factors
  1. Compute $\hat{y}_t$ for the in-sample observations used for fitting the regression. In this example, *t* = 1, 2, …, 16.

$$\hat{y}_1 = 95.2500 + 2.4706(1) = 97.7206$$

$$\hat{y}_2 = 95.2500 + 2.4706(2) = 100.1912$$

$$\ldots\ldots$$

$$\hat{y}_{16} = 95.2500 + 2.4706(16) = 134.7794$$

Procedures of Multiplicative Holt-Winters Method

- **Step 2**: Find the initial seasonal factors
  2. Detrend the data by computing $S_t = y_t / \hat{y}_t$ for each time period that is used in finding the least squares regression equation. In this example, *t* = 1, 2, ..., 16.

$$S_1 = y_1 / \hat{y}_1 = 72 / 97.7206 = 0.7368$$
$$S_2 = y_2 / \hat{y}_2 = 116 / 100.1912 = 1.1578$$
$$\ldots\ldots$$
$$S_{16} = y_{16} / \hat{y}_{16} = 120 / 134.7794 = 0.8903$$

Procedures of Multiplicative Holt-Winters Method

- **Step 2**: Find the initial seasonal factors
  3. Compute the average seasonal values for each of the *L* seasons. The *L* averages are found by computing the average of the detrended values for the corresponding season. For example, for quarter 1,

$$\bar{S}_{[1]} = \frac{S_1 + S_5 + S_9 + S_{13}}{4}$$

$$= \frac{0.7368 + 0.7156 + 0.6894 + 0.6831}{4} = 0.7062$$

Procedures of Multiplicative Holt-Winters Method

- **Step 2**: Find the initial seasonal factors
  4. Multiply the average seasonal values by the normalizing constant

$$CF = \frac{L}{\sum_{i=1}^{L} \overline{S}_{[i]}}$$

such that the average of the seasonal factors is 1. The initial seasonal factors are

$$sn_{i-L} = \overline{S}_{[i]}(CF) \qquad (i = 1, 2, \ldots, L)$$

Procedures of Multiplicative Holt-Winters Method

- **Step 2**: Find the initial seasonal factors
  4. Multiply the average seasonal values by the normalizing constant such that the average of the seasonal factors is 1.
     - Example

       CF = 4/3.9999 = 1.0000

$$sn_{-3} = sn_{1-4} = \bar{S}_{[1]}(CF) = 0.7062(1) = 0.7062$$

$$sn_{-2} = sn_{2-4} = \bar{S}_{[2]}(CF) = 1.1114(1) = 1.1114$$

$$sn_{-1} = sn_{3-4} = \bar{S}_{[3]}(CF) = 1.2937(1) = 1.2937$$

$$sn_{0} = sn_{4-4} = \bar{S}_{[1]}(CF) = 0.8886(1) = 0.8886$$

Procedures of Multiplicative Holt-Winters Method

- **Step 3**: Calculate a point forecast of y_1 from time 0 using the initial values

$$\hat{y}_{T+p}(T) = (\ell_T + pb_T)sn_{T+p-L} \qquad (T = 0,\ p = 1)$$

$$\hat{y}_1(0) = (\ell_0 + b_0)sn_{1-4} = (\ell_0 + b_0)sn_{-3}$$

$$= (95.2500 + 2.4706)(0.7062)$$

$$= 69.0103$$

Procedures of Multiplicative Holt-Winters Method

- **Step 4**: Update the estimates ***ℓ**$_T$*, *b$_T$*, and *sn$_T$* by using some predetermined values of smoothing constants.
- Example: let $\alpha = 0.2$, $\gamma = 0.1$, and $\delta = 0.1$

$$\ell_1 = \alpha(y_1 / sn_{1-4}) + (1-\alpha)(\ell_0 + b_0)$$
$$= 0.2(72/0.7062) + 0.8(95.2500 + 2.4706) = 98.5673$$

$$b_1 = \gamma(\ell_1 - \ell_0) + (1-\gamma)b_0$$
$$= 0.1(98.5673 - 95.2500) + 0.9(2.4706) = 2.5553$$

$$sn_1 = \delta(y_1 / \ell_1) + (1-\delta)sn_{1-4}$$
$$= 0.1(72/98.5673) + 0.9(0.7062) = 0.7086$$

$$\hat{y}_2(1) = (\ell_1 + b_1)sn_{2-4}$$
$$= (98.5673 + 2.5553)(1.1114) = 112.3876$$

$$
\begin{aligned}
\ell_2 &= \alpha\left(y_2 / sn_{2-4}\right) + (1-\alpha)(\ell_1 + b_1) \\
&= 0.2(116/1.1114) + 0.8(98.5673 + 2.5553) \\
&= 101.7727 \\
b_2 &= \gamma(\ell_2 - \ell_1) + (1-\gamma)b_1 \\
&= 0.1(101.7727 - 98.5673) + 0.9(2.5553) \\
&= 2.62031 \\
sn_2 &= \delta\left(y_2 / \ell_2\right) + (1-\delta)sn_{2-4} \\
&= 0.1(116/101.7727) + 0.9(1.1114) \\
&= 1.114239 \\
\hat{y}_3(2) &= (\ell_2 + b_2)sn_{3-4} \\
&= (101.7727 + 2.62031)(1.2937) \\
&= 135.053
\end{aligned}
$$

$$
\begin{aligned}
\ell_4 &= \alpha\left(y_4/sn_{4-4}\right) + (1-\alpha)(\ell_3 + b_3) \\
&= 0.2(96/0.8886) + 0.8(104.5393 + 2.6349) \\
&= 107.3464 \\
b_4 &= \gamma(\ell_4 - \ell_3) + (1-\gamma)b_3 \\
&= 0.1(107.3464 - 104.5393) + 0.9(2.6349) \\
&= 2.65212 \\
sn_4 &= \delta\left(y_4/\ell_4\right) + (1-\delta)sn_{4-4} \\
&= 0.1(96/107.3464) + 0.9(0.8886) \\
&= 0.889170 \\
\hat{y}_5(4) &= (\ell_4 + b_4)sn_{5-4} \\
&= (107.3464 + 2.65212)(0.7086) \\
&= 77.945
\end{aligned}
$$

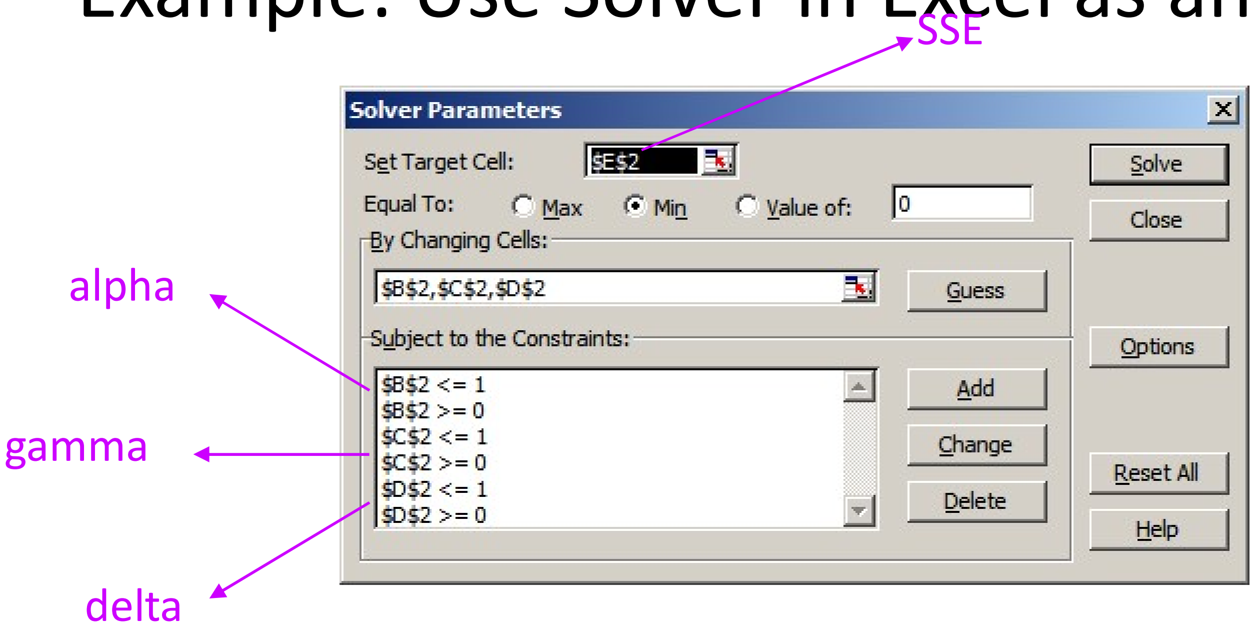
1	n	alpha	gamma	delta	SSE	MSE	s	
2	32	0.2	0.1	0.1	177.3223	6.1146	2.4728	
3								
4								
5						Forecast		Squared
6				Growth	Seasonal	Made Last	Forecast	Forecast
7	Time	y	Level	Rate	Factor	Period	Error	Error
8	-3				0.7062			
9	-2				1.1114			
10	-1				1.2937			
11	0		95.25	2.4706	0.8886			
12	1	72	98.56729	2.5553	0.7086	69.0103	2.9897	8.9384
13	2	116	101.7726	2.6203	1.1142	112.3876	3.6124	13.0494
14	3	136	104.5393	2.6349	1.2944	135.0531	0.9469	0.8967
15	4	96	107.3464	2.6521	0.8892	95.2350	0.7650	0.5853
16	5	77	109.731	2.6254	0.7079	77.9478	-0.9478	0.8984
17	6	123	111.9629	2.5860	1.1127	125.1919	-2.1919	4.8043
18	7	146	114.1974	2.5509	1.2928	148.2750	-2.2750	5.1755
19	8	101	116.1165	2.4877	0.8872	103.8091	-2.8091	7.8911
20	9	81	117.7668	2.4040	0.7059	83.9641	-2.9641	8.7858
21	10	131	119.6835	2.3552	1.1109	133.7108	-2.7108	7.3482
22	11	158	122.0734	2.3587	1.2930	157.7754	0.2246	0.0504
23	12	109	124.1164	2.3271	0.8863	110.4005	-1.4005	1.9615
24	13	87	125.8035	2.2631	0.7045	89.2593	-2.2593	5.1044
25	14	140	127.6589	2.2224	1.1094	142.2642	-2.2642	5.1268
26	15	167	129.7369	2.2079	1.2924	167.9337	-0.9337	0.8718

…… ……

38	27	200	156.1396	2.1752	1.2903	202.0396	-2.0396	4.1601
39	28	142	158.5505	2.1988	0.8908	140.9508	1.0492	1.1008
40	29	115	161.2803	2.2519	0.7047	113.1314	1.8686	3.4918
41	30	177	162.8178	2.1804	1.1046	180.9529	-3.9529	15.6252
42	31	218	165.7889	2.2595	1.2928	212.8988	5.1012	26.0220
43	32	149	167.8899	2.2437	0.8905	149.7057	-0.7057	0.4981

Procedures of Multiplicative Holt-Winters Method

- **Step 5**: Find the most suitable combination of α, γ, and δ that minimizes SSE (or MSE)
- Example: Use Solver in Excel as an illustration



1	n	alpha	gamma	delta	SSE	MSE	s	
2	32	0.3356	0.0455	0.1342	168.4747	5.8095	2.4103	
3								
4								
5						Forecast		Squared
6				Growth	Seasonal	Made Last	Forecast	Forecast
7	Time	y	Level	Rate	Factor	Period	Error	Error
8	-3				0.7062			
9	-2				1.1114			
10	-1				1.2937			
11	0		95.25	2.4706	0.8886			
12	1	72	99.14144	2.5353	0.7089	69.0103	2.9897	8.9384
13	2	116	102.5816	2.5765	1.1140	113.0035	2.9965	8.9789
14	3	136	105.1469	2.5760	1.2937	136.0431	-0.0431	0.0019
15	4	96	107.8277	2.5808	0.8888	95.7226	0.2774	0.0769
16	5	77	109.8084	2.5534	0.7079	78.2674	-1.2674	1.6064
17	6	123	111.7076	2.5236	1.1123	125.1717	-2.1717	4.7164
18	7	146	113.7703	2.5027	1.2923	147.7768	-1.7768	3.1569
19	8	101	115.3868	2.4623	0.8870	103.3468	-2.3468	5.5075
20	9	81	116.7014	2.4100	0.7060	83.4207	-2.4207	5.8597
	......	......						
38	27	200	155.9042	2.2691	1.2906	202.1107	-2.1107	4.4552
39	28	142	158.5811	2.2876	0.8915	140.9173	1.0827	1.1721
40	29	115	161.7496	2.3278	0.7044	113.1540	1.8460	3.4078
41	30	177	162.7095	2.2655	1.1038	181.5085	-4.5085	20.3262
42	31	218	166.2957	2.3256	1.2934	212.9210	5.0790	25.7957
43	32	149	168.1213	2.3028	0.8908	150.3283	-1.3283	1.7643

Multiplicative Holt-Winters Method

- *p*-step-ahead forecast made at time *T*

$$\hat{y}_{T+p}(T) = (\ell_T + pb_T)sn_{T+p-L} \qquad (p = 1, 2, 3, \ldots)$$

- Example

$$\hat{y}_{33}(32) = (\ell_{32} + b_{32})sn_{33-4} = (168.1213 + 2.3028)(0.7044) = 120.0467$$

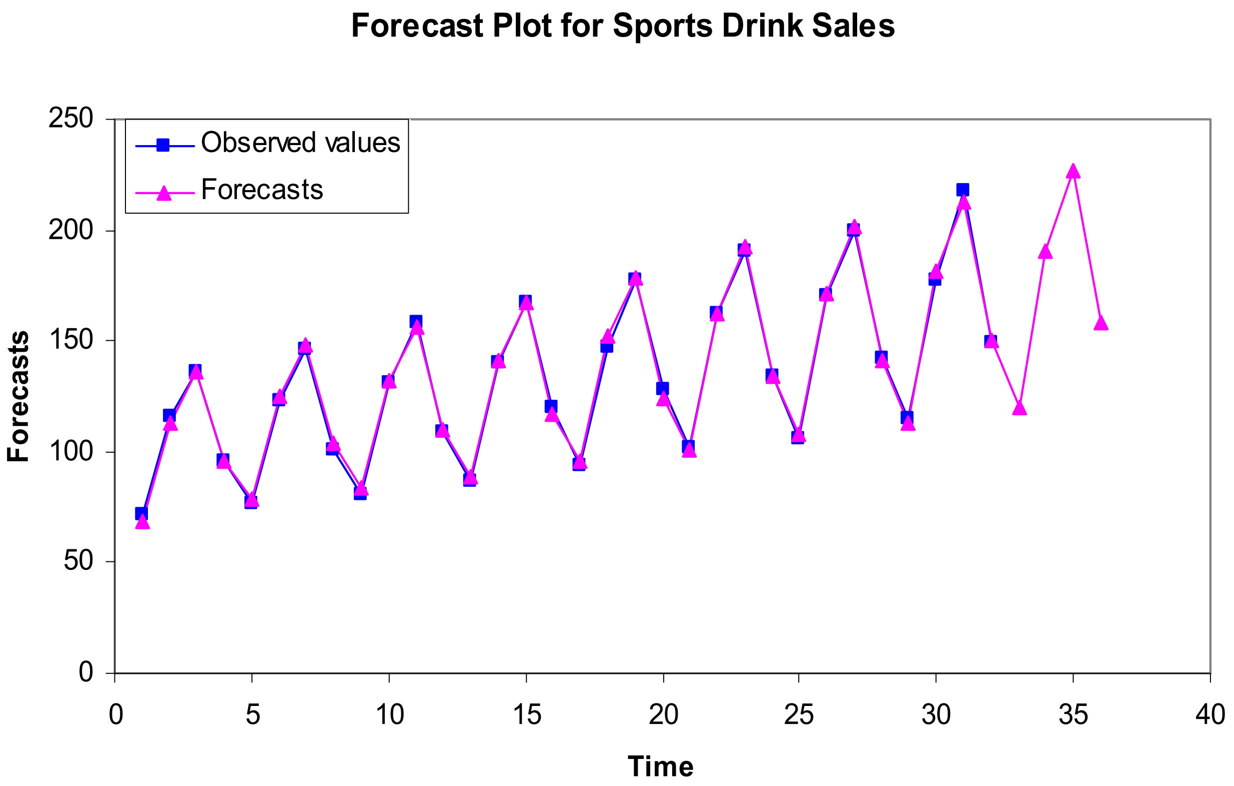
$$\hat{y}_{34}(32) = (\ell_{32} + 2b_{32})sn_{34-4} = [168.1213 + 2(2.3028)](1.1038) = 190.6560$$

$$\hat{y}_{35}(32) = (\ell_{32} + 3b_{32})sn_{35-4} = [(168.1213 + 3(2.3028)](1.2934) = 226.3834$$

$$\hat{y}_{36}(32) = (\ell_{32} + 4b_{32})sn_{36-4} = [(168.1213 + 4(2.3028)](0.8908) = 157.9678$$

Multiplicative Holt-Winters Method

- Example



Additive Holt-Winters Method

- It is generally considered to be best suited to forecasting a time series that can be described by the equation:

$$y_t = (\beta_0 + \beta_1 t) + SN_t + IR_t$$

  - SN_t: seasonal pattern
  - IR_t: irregular component
- This method is appropriate when a time series has a linear trend with a constant (additive) seasonal pattern such that the level ($\beta_0 + \beta_1 t$), growth rate (β_1), and the seasonal pattern (SN_t) may be slowly changing over time.

Additive Holt-Winters Method

- Estimate of the level

$$\ell_T = \alpha(y_T - sn_{T-L}) + (1-\alpha)(\ell_{T-1} + b_{T-1})$$

- Estimate of the growth rate (or trend)

$$b_T = \gamma(\ell_T - \ell_{T-1}) + (1-\gamma)b_{T-1}$$

- Estimate of the seasonal factor

$$sn_T = \delta(y_T - \ell_T) + (1-\delta)sn_{T-L}$$

where α, γ, and δ are smoothing constants between 0 and 1,

L = number of seasons in a year (L = 12 for monthly data, and L = 4 for quarterly data)

Additive Holt-Winters Method

- Point forecast made at time *T* for y_{T+p}

$$\hat{y}_{T+p}(T) = \ell_T + pb_T + sn_{T+p-L} \qquad (p = 1, 2, 3, \ldots)$$

- MSE and the standard error *s* at time *T*

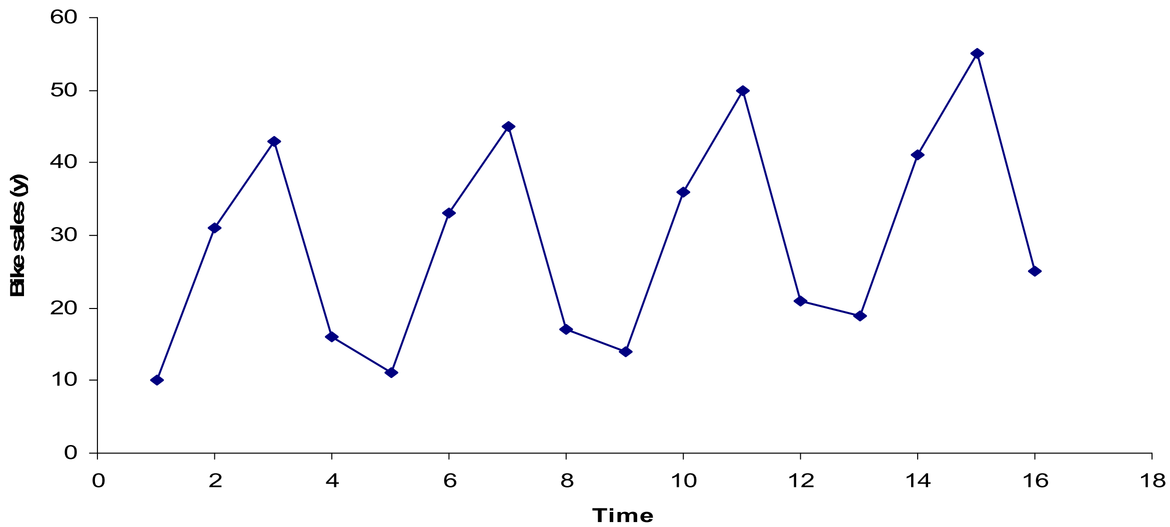
$$SSE = \sum_{t=1}^{T} [y_t - \hat{y}_t(t-1)]^2$$

$$MSE = \frac{SSE}{T-3}, \qquad s = \sqrt{MSE}$$

Procedures of Additive Holt-Winters Method

- Consider the Mountain Bike example,

Quarterly sales of the TRK-50 Mountain Bike				
	Year			
Quarter	1	2	3	4
1	10	11	14	19
2	31	33	36	41
3	43	45	50	55
4	16	17	21	25



Procedures of Additive Holt-Winters Method

- Observations:
  - Linear upward trend over the 4-year period
  - Magnitude of seasonal span is almost constant as the level of the time series increases

  $\Rightarrow$ Additive Holt-Winters method can be applied to forecast future sales

Procedures of Additive Holt-Winters Method

- **Step 1**: Obtain initial values for the level ℓ_0, the growth rate b_0, and the seasonal factors sn_{-3}, sn_{-2}, sn_{-1}, and sn_0, by fitting a least squares trend line to at least four or five years of the historical data.
  - y-intercept = ℓ_0; slope = b_0

Procedures of Additive Holt-Winters Method

- Example
  - Fit a least squares trend line to all 16 observations
  - Trend line

$$\hat{y}_t = 20.85 + 0.980882\,t$$

  - $\ell_0 = 20.85$; $b_0 = 0.9809$

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.320508842
R Square	0.102725918
Adjusted R Square	0.038634912
Standard Error	14.28614022
Observations	16
ANOVA	
	df
Regression	1
Residual	14
Total	15
	Coefficients
Intercept	20.85
Time	0.980882353

Procedures of Additive Holt-Winters Method

- **Step 2**: Find the initial seasonal factors
  1. Compute $\hat{y}_t$ for each time period that is used in finding the least squares regression equation. In this example, *t* = 1, 2, …, 16.

$$\hat{y}_1 = 20.85 + 0.980882(1) = 21.8309$$

$$\hat{y}_2 = 20.85 + 0.980882(2) = 22.8118$$

$$\ldots\ldots$$

$$\hat{y}_{16} = 20.85 + 0.980882(16) = 36.5441$$

Procedures of Additive Holt-Winters Method

- **Step 2**: Find the initial seasonal factors
  2. Detrend the data by computing $S_t = y_t - \hat{y}_t$ for each observation used in the least squares fit. In this example, *t* = 1, 2, …, 16.

$$S_1 = y_1 - \hat{y}_1 = 10 - 21.8309 = -11.8309$$

$$S_2 = y_2 - \hat{y}_2 = 31 - 22.8112 = 8.1882$$

……

$$S_{16} = y_{16} - \hat{y}_{16} = 25 - 36.5441 = -11.5441$$

Procedures of Additive Holt-Winters Method

- **Step 2**: Find the initial seasonal factors
  3. Compute the average seasonal values for each of the *L* seasons. The *L* averages are found by computing the average of the detrended values for the corresponding season. For example, for quarter 1,

$$\overline{S}_{[1]} = \frac{S_1 + S_5 + S_9 + S_{13}}{4}$$

$$= \frac{(-11.8309) + (-14.7544) + (-15.6779) + (-14.6015)}{4} = -14.2162$$

Procedures of Additive Holt-Winters Method

- **Step 2**: Find the initial seasonal factors
  4. Compute the average of the *L* seasonal factors. The average should be 0.

Procedures of Additive Holt-Winters Method

- **Step 3**: Calculate a point forecast of y_1 from time 0 using the initial values

$$\hat{y}_{T+p}(T) = \ell_T + pb_T + sn_{T+p-L} \qquad (T = 0, p = 1)$$

$$\hat{y}_1(0) = \ell_0 + b_0 + sn_{1-4} = \ell_0 + b_0 + sn_{-3}$$

$$= 20.85 + 0.9809 + (-14.2162) = 7.6147$$

Procedures of Additive Holt-Winters Method

- **Step 4**: Update the estimates ℓ_T, b_T, and sn_T by using some predetermined values of smoothing constants.
- Example: let $\alpha = 0.2$, $\gamma = 0.1$, and $\delta = 0.1$

$$\ell_1 = \alpha(y_1 - sn_{1-4}) + (1-\alpha)(\ell_0 + b_0)$$
$$= 0.2(10 - (-14.2162)) + 0.8(20.85 + 0.9808) = 22.3079$$

$$b_1 = \gamma(\ell_1 - \ell_0) + (1-\gamma)b_0$$
$$= 0.1(22.3079 - 20.85) + 0.9(0.9809) = 1.0286$$

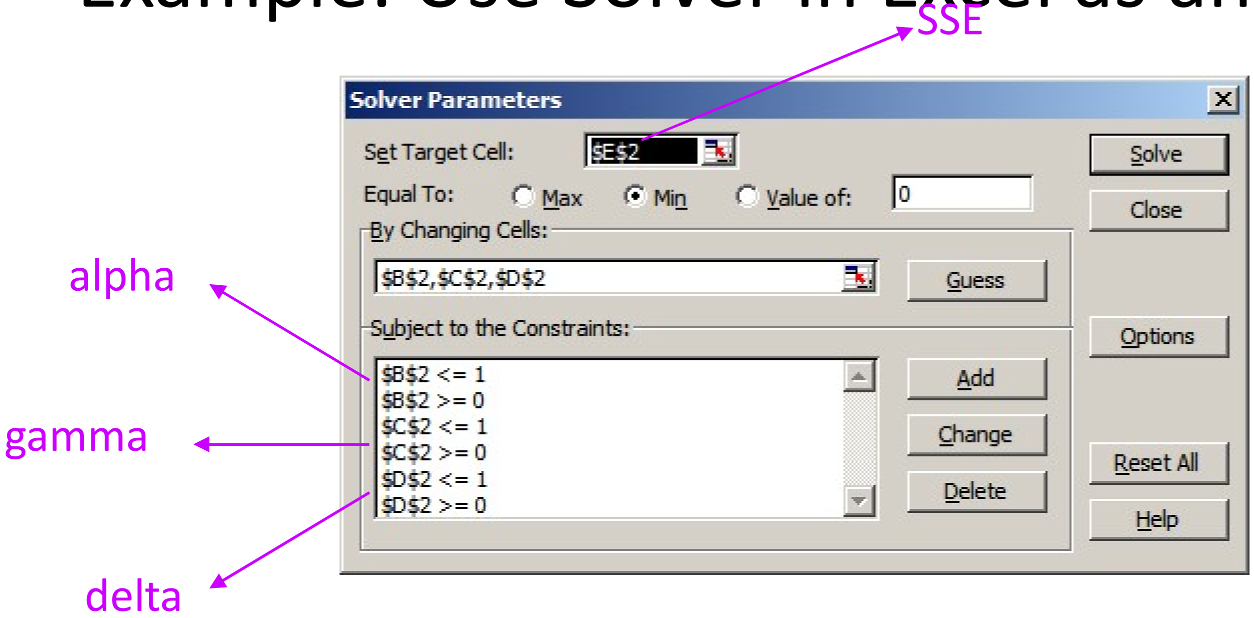
$$sn_1 = \delta(y_1 - \ell_1) + (1-\delta)sn_{1-4}$$
$$= 0.1(10 - 22.3079) + 0.9(-14.2162) = -14.0254$$

$$\hat{y}_2(1) = \ell_1 + b_1 + sn_{2-4} = \ell_1 + b_1 + sn_{-2}$$
$$= 22.3079 + 1.0286 + 6.5529 = 29.8895$$

1	n	alpha	gamma	delta	SSE	MSE	s	
2	16	0.2000	0.1000	0.1000	25.2166	1.9397	1.3927	
3								
4								
5						Forecast		Squared
6				Growth	Seasonal	Made Last	Forecast	Forecast
7	Time	y	Level	Rate	Factor	Period	Error	Error
8	-3				-14.2162			
9	-2				6.5529			
10	-1				18.5721			
11	0		20.85	0.9809	-10.9088			
12	1	10	22.30794	1.0286	-14.0254	7.6147	2.3853	5.6896
13	2	31	23.55864	1.0508	6.6418	29.8895	1.1105	1.2333
14	3	43	24.57314	1.0472	18.5575	43.1815	-0.1815	0.0329
15	4	16	25.87801	1.0729	-10.8057	14.7115	1.2885	1.6603
16	5	11	26.56583	1.0344	-14.1794	12.9256	-1.9256	3.7079
17	6	33	27.35185	1.0096	6.5424	34.2420	-1.2420	1.5427
18	7	45	27.97764	0.9712	18.4040	46.9190	-1.9190	3.6825
19	8	17	28.72023	0.9483	-10.8972	18.1431	-1.1431	1.3067
20	9	14	29.37074	0.9186	-14.2985	15.4892	-1.4892	2.2176
21	10	36	30.12295	0.9019	6.4759	36.8317	-0.8317	0.6918
22	11	50	31.1391	0.9133	18.4497	49.4289	0.5711	0.3262
23	12	21	32.0214	0.9102	-10.9096	21.1553	-0.1553	0.0241
24	13	19	33.00502	0.9176	-14.2692	18.6331	0.3669	0.1346
25	14	41	34.04291	0.9296	6.5240	40.3985	0.6015	0.3618
26	15	55	35.28807	0.9612	18.5759	53.4222	1.5778	2.4894
27	16	25	36.18131	0.9544	-10.9368	25.3396	-0.3396	0.1153

Procedures of Additive Holt-Winters Method

- **Step 5**: Find the most suitable combination of α, γ, and δ that minimizes SSE (or MSE)
- Example: Use Solver in Excel as an illustration



1	n	alpha	gamma	delta	SSE	MSE	s	
2	16	0.5606	0.0000	0.0000	18.7975	1.4460	1.2025	
3								
4								
5						Forecast		Squared
6				Growth	Seasonal	Made Last	Forecast	Forecast
7	Time	y	Level	Rate	Factor	Period	Error	Error
8	-3				-14.2162			
9	-2				6.5529			
10	-1				18.5721			
11	0		20.85	0.9809	-10.9088			
12	1	10	23.16818	0.9809	-14.2162	7.6147	2.3853	5.6896
13	2	31	24.31613	0.9809	6.5529	30.7020	0.2980	0.0888
14	3	43	24.80977	0.9809	18.5721	43.8691	-0.8691	0.7553
15	4	16	26.41755	0.9809	-10.9088	14.8818	1.1182	1.2503
16	5	11	26.17496	0.9809	-14.2162	13.1823	-2.1823	4.7622
17	6	33	26.75847	0.9809	6.5529	33.7088	-0.7088	0.5024
18	7	45	27.00412	0.9809	18.5721	46.3114	-1.3114	1.7198
19	8	17	27.94229	0.9809	-10.9088	17.0762	-0.0762	0.0058
20	9	14	28.5268	0.9809	-14.2162	14.7070	-0.7070	0.4998
21	10	36	29.47369	0.9809	6.5529	36.0606	-0.0606	0.0037
22	11	50	31.00029	0.9809	18.5721	49.0266	0.9734	0.9474
23	12	21	31.94061	0.9809	-10.9088	21.0723	-0.0723	0.0052
24	13	19	33.0867	0.9809	-14.2162	18.7053	0.2947	0.0868
25	14	41	34.28034	0.9809	6.5529	40.6205	0.3795	0.1440
26	15	55	35.91533	0.9809	18.5721	53.8333	1.1667	1.3612
27	16	25	36.34264	0.9809	-10.9088	25.9874	-0.9874	0.9749

Additive Holt-Winters Method

- *p*-step-ahead forecast made at time *T*

$$\hat{y}_{T+p}(T) = \ell_T + pb_T + sn_{T+p-L} \qquad (p = 1, 2, 3, \ldots)$$

- Example

$$\hat{y}_{17}(16) = \ell_{16} + b_{16} + sn_{17-4} = 36.3426 + 0.9809 - 14.2162 = 23.1073$$

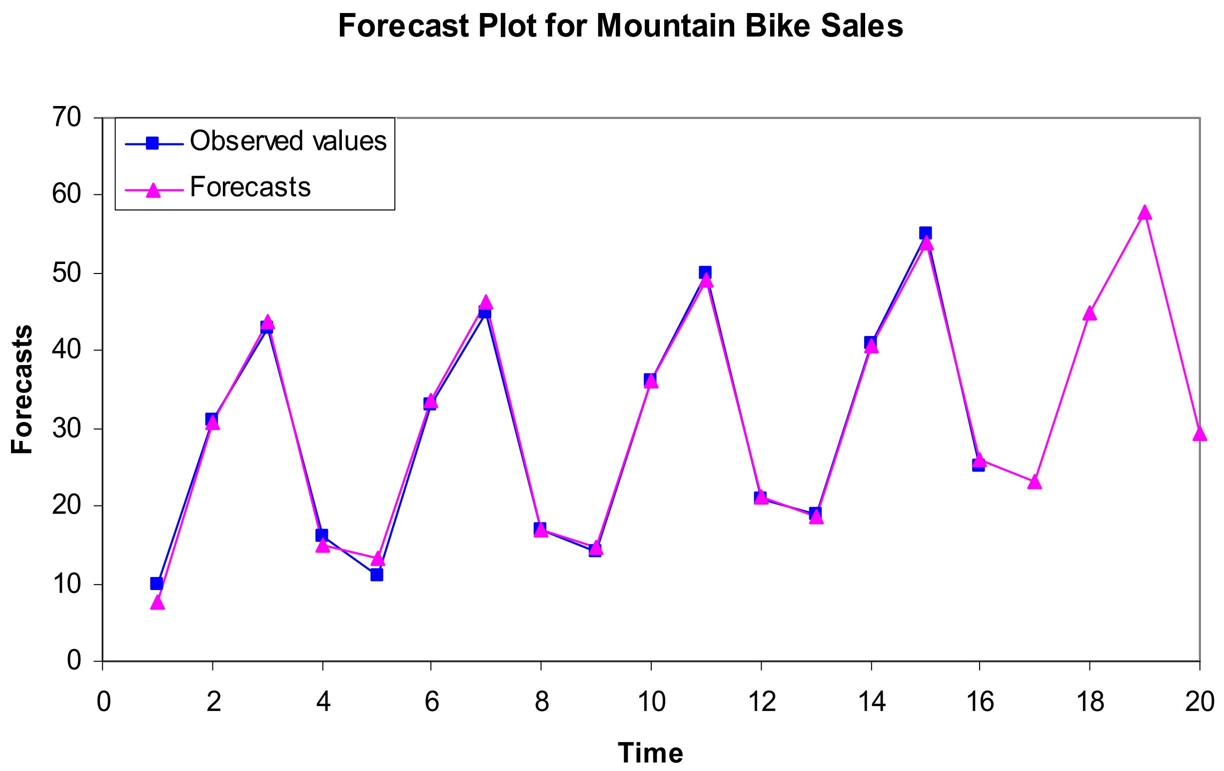
$$\hat{y}_{18}(16) = \ell_{16} + 2b_{16} + sn_{18-4} = 36.3426 + 2(0.9809) + 6.5529 = 44.8573$$

$$\hat{y}_{19}(16) = \ell_{16} + 3b_{16} + sn_{19-4} = 36.3426 + 3(0.9809) + 18.5721 = 57.8573$$

$$\hat{y}_{20}(16) = \ell_{16} + 4b_{16} + sn_{20-4} = 36.3426 + 4(0.9809) - 10.9088 = 29.3573$$

Additive Holt-Winters Method

- Example



Chapter Summary

- Simple Exponential Smoothing
  - No trend, no seasonal pattern
- Holt's Trend Corrected Exponential Smoothing
  - Trend, no seasonal pattern
- Holt-Winters Methods
  - Both trend and seasonal pattern
    - Multiplicative Holt-Winters method
    - Additive Holt-Winters Method

Regression Models

Abdellah El Fallahi
SCM1

Introduction

- Definition of linear regression
  - Linear regression is a statistical technique used to determine the relationship between variables. In an ML context, linear regression is used to find the relationship between characteristics and a label.

Why it's important in data science and statistics

1. Foundation for Advanced Models

- It is one of the simplest and most fundamental ML algorithms.
- Many advanced models (e.g., logistic regression, neural networks) build upon similar mathematical concepts.

2. Easy to Interpret

- Unlike complex models, linear regression provides clear relationships between variables.
- The coefficients explain how much each independent variable affects the dependent variable.

3. Used for Predictive Analysis

- It helps predict continuous values, such as house prices, stock prices, and sales forecasting.

4. Efficient and Fast

- Computationally inexpensive and works well on small to medium datasets.
- Requires less processing power compared to deep learning models.

Why it's important in data science and statistics

5. Feature Selection & Insights

- Helps in identifying important features (variables) that have the most impact on predictions.
- Useful in understanding trends and correlations in data.

6. Good Baseline Model

- Often used as a benchmark before implementing more complex models.
- Helps in evaluating whether more sophisticated techniques are necessary.

7. Works Well with Linearly Separable Data

- If data follows a linear pattern, this algorithm provides excellent performance.

Real-World Applications

- Predicting house prices
- Forecasting sales revenue
- Medical predictions (e.g., disease progression)
- Predict a car's fuel efficiency in miles per gallon as a function of its weight

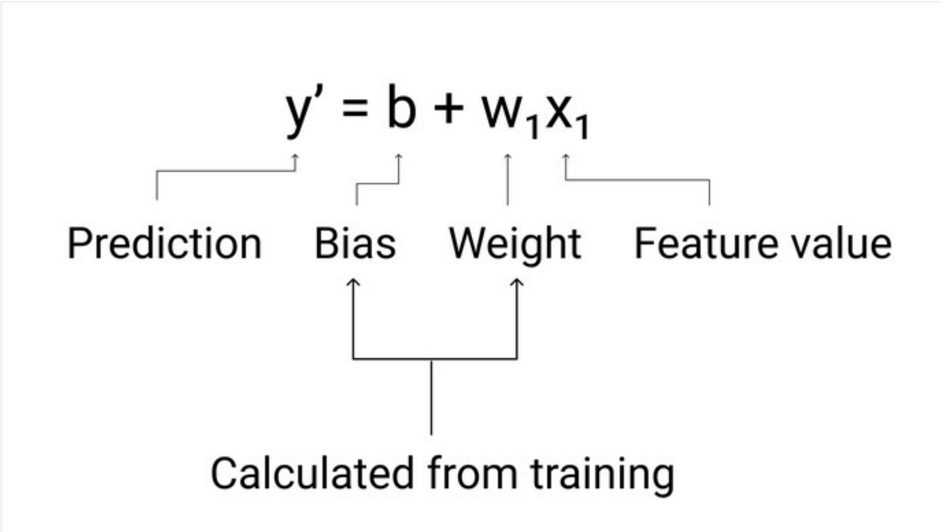
Livres par milliers (fonctionnalité)	Miles par gallon (libellé)
3.5	18
3,69	15
3.44	18
3,43	16
4,34	15
4,42	14
2,37	24

Types of Linear Regression

- Simple Linear Regression (one independent variable)
- Multiple Linear Regression (multiple independent variables)

The Linear Regression Equation

- Formula:



- Explanation of slope (w1), intercept (b), and dependent variable (y)
- Y’ the value estimated by the model

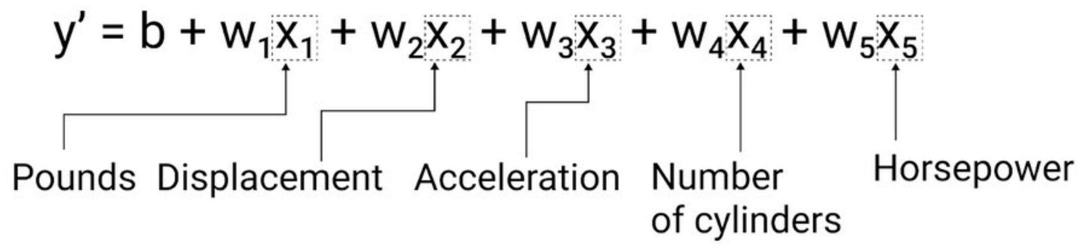
The Linear Regression Equation

- Although the example in this section uses just one characteristic, a more sophisticated model may be based on several characteristics, each with its own weight.

$$y' = b + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5$$

The Linear Regression Equation

- For example, a model that predicts fuel consumption can also use features such as the following:
  - Pounds
  - Displacement
  - Acceleration
  - Number of cylinders
  - Horsepower



Assumptions of Linear Regression

- Linearity
- Independence of errors
- Homoscedasticity
- Normal distribution of residuals

Data Requirements

- Types of data suitable for linear regression
- Preprocessing steps (handling missing values, scaling)

Visualizing Linear Regression

- Scatter plot with best-fit line
- Example dataset with regression line

How to Find the Best Fit Line

- Ordinary Least Squares (OLS) method
- Minimizing the sum of squared errors

General Idea of p-values

- For a coefficient βi:
  - **Null hypothesis**: βi=0
  - **Alternative hypothesis**: βi≠0

- Typical interpretation:
  - p<0.05p : significant
  - p<0.01: highly significant
  - p>0.05: coefficient may not contribute meaningfully

Analysis of P-values

- If a coefficient has a **non-significant p-value** (for example $p>0.05$), it means that the variable may not have a statistically detectable effect in the model. There are several things you can do depending on the context.
- **1. Remove the Variable**
- If the variable is not theoretically important, you may remove it from the model.
- Example:
  - $y=\beta 0+\beta 1x1+\beta 2x2$ If x2 has a high p-value, try fitting:
- $Y=\beta 0+\beta 1x1$ Then compare the models using:
  - Adjusted R2R^2R2
  - AIC
  - RMSE
- Compare Models

  anova(model_reduced, model_full)

Analysis of P-values

- **2. Check Multicollinearity**
  - Sometimes a variable becomes non-significant because predictors are strongly correlated.
- Use:
  - correlation matrix
  - VIF (Variance Inflation Factor)
- In R:
  - library(car)
    vif(model)
- Large VIF values (usually > 5 or > 10) indicate multicollinearity.

Analysis of P-values

- **3. Increase Sample Size**
  - With small datasets, important variables may appear non-significant because statistical power is weak.
  - More observations often reduce standard errors and improve significance.

Evaluating Model Performance

- **Goodness-of-fit measures**
- R2
- Adjusted R2
- RMSE
- MAE
- AIC / BIC
- **Residual analysis**
  - Check whether residuals are:
  - randomly distributed
  - homoscedastic
  - approximately normal

Overfitting vs. Underfitting

- Definition and examples
- How to avoid overfitting (regularization techniques)

Implementing Linear Regression in Python

- Brief code snippet using scikit-learn
- Step-by-step explanation

Common Challenges

- • Multicollinearity
- • Outliers
- • Non-linearity in data

Case Study Example

- A simple case study on predicting salaries based on experience

Conclusion

- Summary of key points
- Importance of linear regression in machine learning

Exo 1

- The quality of a mattress is measured by its foam density in kg/m^3. The foam density of a mattress (Y) is generally affected by three factors: polyols (X1), isocyanates (X2), and blowing agents (X3). Table 1 presents the results of 30 tests carried out by varying the values of the three factors X1, X2, and X3. The objective is to highlight the individual impact of each factor on the response variable. In addition, we seek to analyze the effects of interaction terms and quadratic terms on foam density.

Data

X1	X2	X3	Y	Z
2.39	1.18	2.34	15.14	1
1.57	1.87	2.17	11.37	0
1.45	1.86	2.25	10.04	0
2.10	1.99	2.35	15.11	1
2.44	1.85	2.68	16.12	1
1.85	1.62	1.17	14.88	0
2.96	1.85	2.53	21.28	1
2.37	2.79	1.49	21.16	1
1.96	2.89	1.39	18.94	1
1.78	2.00	2.14	13.48	0
1.69	2.25	1.19	15.73	1
2.46	1.23	2.77	14.03	0
1.88	1.63	2.25	12.73	0
1.12	1.83	2.45	6.03	0
1.80	2.73	1.03	17.63	1
2.48	1.50	2.19	17.17	1
1.36	1.97	2.11	10.33	0
1.35	2.97	1.32	15.55	1
2.06	2.04	1.31	17.18	1
2.06	2.23	2.39	15.29	1
2.27	1.24	1.64	15.94	1
2.70	2.65	2.38	21.94	1
2.45	2.21	2.11	19.26	1
2.22	2.09	1.78	17.93	1
2.44	1.69	2.85	14.86	0
1.65	1.61	2.68	8.38	0
1.72	1.83	1.71	14.03	0
1.46	2.36	1.09	14.88	0
1.59	2.75	1.61	15.84	1
2.26	2.02	1.80	17.97	1

Required work:

- Analyze the individual effects of the three factors (X1, X2, X3) on the foam density of a mattress (Y).
- Provide the mathematical model obtained using only the simple terms.
- Analyze the individual effects, interaction effects, and quadratic terms on foam density.
- Comment on the variation of the effects of the simple terms between the models obtained in Questions 1 and 3.
- Determine the density of a mattress for (X1=4.5, X2=2.5, X3=3)(X1 = 4.5,\; X2 = 2.5,\; X3 = 3)(X1=4.5,X2=2.5,X3=3) using the two models from Questions 1 and 3.