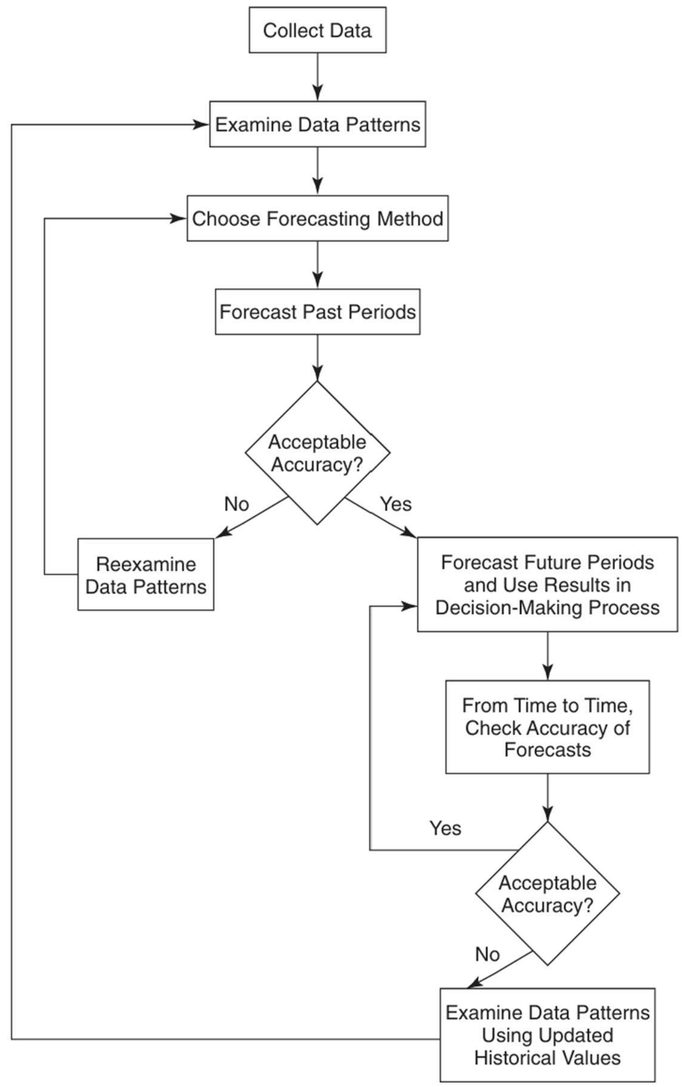


FIGURE 1 **The Operational Phase of the Forecasting Process**

Collect Data

1. Data should be reliable and accurate. Proper care must be taken that data are col lected from a reliable source with proper attention given to accuracy.
2. Data should be relevant. The data must be representative of the circumstances for which they are being used.
3. Data should be consistent. When definitions concerning data collection change, adjustments need to be made to retain consistency in historical patterns. This can be a problem, for example, when government agencies change the mix or “market basket” used in determining a cost-of-living index. Years ago personal computers were not part of the mix of products being purchased by consumers; now they are.
4. Data should be timely. Data collected, summarized, and published on a timely basis will be of greatest value to the forecaster. There can be too little data (not enough history on which to base future outcomes) or too much data (data from irrelevant historical periods far in the past).

Cross-sectional	data are observations collected at a single point in time.
Time series	consists of data that are collected, recorded, or observed over successive increments of time.

Dependent variable (Y) → the thing you are trying to explain or predict.
Independent variables (X) → the factors you think explain changes in Y.

Examine Data Patterns

Data Patterns – Time Series

four general types of patterns

1. Horizontal

   When data collected over time fluctuate around a constant level or mean, a Horizontal pattern exists.

   - stationary

2. Trend

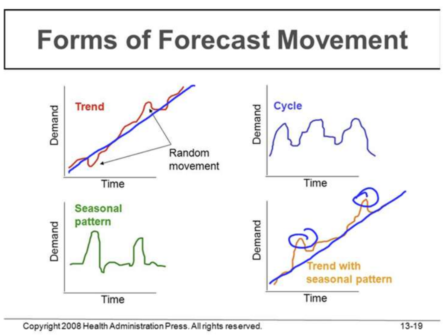
   The trend is the long-term component that represents the growth or decline in the time series over an extended period of time.

3. Seasonal

   The seasonal component is a pattern of change that repeats itself year after year.

4. Cyclical

   The cyclical component is the wavelike fluctuation around the trend.



Stationarity basics

A stationary time series has:

- Constant mean over time
- Constant variance over time
- Constant autocorrelation structure over time

So — the statistical properties of the series do not change as time passes.

Seasonality and stationarity

- Seasonal data can be stationary, but only if the seasonal pattern is constant (i.e., repeats with the same amplitude and frequency over time).
- However, most real seasonal series are not stationary, because their mean changes periodically (e.g., sales are higher every December).

Example of stationary seasonal data:

A sine wave that repeats perfectly with the same amplitude and period forever.

Example of non-stationary seasonal data:

Sales that increase every year but also have seasonal peaks (trend + seasonality).

Trend and stationarity

- If a series has a trend (upward or downward), it cannot be stationary, because the mean changes over time.
- A trending series violates the constant-mean condition.

Trend → Non-stationary.

After differencing (removing the trend), it can become stationary.

Process for time Series Forecasting

Convert data to a time series

ts(data, start = c(year, period), frequency = n)

data — your numeric vector (e.g., sales, temperature, etc.)

start — (e.g., c(2020, 1)

This tells R

Data starts in January 2020

This becomes important when:

- Plotting (x-axis will show years properly)
- Using time-based models (like SARIMA, STL decomposition, forecasting)
- Doing seasonal differencing (diff(y, lag = 12))

End — end is optional and rarely needed — R calculates it automatically from the length of your data and the start/frequency.

You'd only specify end if:

- Your data doesn't cover a full year or period
- You want to set a specific final date manually

End — (e.g., c(2023, 6))

this defines data to June 2023, even if the number of observations might suggest otherwise.

Frequency — Frequency = 1

Use when: No seasonality (annual data).

Examples:

- Yearly population
- Annual GDP
- Annual rainfall totals

One data point per cycle → no repeating pattern.

frequency = 4

Use when: Quarterly data with yearly seasonality.

Examples:

- Quarterly GDP
- Quarterly revenue
- Quarterly expenses

4 quarters per year.

frequency = 12

Use when: Monthly data with yearly seasonality.

Examples:

- Monthly sales
- Monthly temperature
- Monthly expenses

12 months per year.

frequency = 52

Use when: Weekly data with yearly seasonality.

Examples:

- Weekly shipping volume
- Weekly demand
- Weekly energy generation

52 weeks per year.

frequency = 7

Use when: Daily data showing weekly seasonality.

Examples:

- Daily website visits (weekday vs weekend)
- Daily sales with a weekly cycle
- Daily foot traffic to stores

Weekly cycle = 7 days.

frequency = 365 (or 365.25)

Use when: Daily data showing yearly seasonality.

Examples:

- Temperature
- Rainfall
- Daylight hours
- Annual biological/medical cycles

Yearly cycle = 365 days.

frequency = 5

Use when: Daily data recorded only on weekdays (business-day data).

Examples:

- Stock market prices
- Stock trading volume
- Business-day call center volume
- Manufacturing output (Mon–Fri)
- Operational metrics without weekends

Weekly business cycle = 5 days.

make time series.R

Find out what the data is:

Autocorrelation Analysis

„Autocorrelation is the correlation between a variable lagged one or more time periods and itself."

a) Plotting the Data

ts.plot(data)

Look for:

- Trend → general upward/downward movement
- Seasonality → repeating cycles (e.g., yearly peaks)
- Stationarity → fluctuations around a constant mean/variance

b) ACF and PACF

acf(data)

pacf(data)

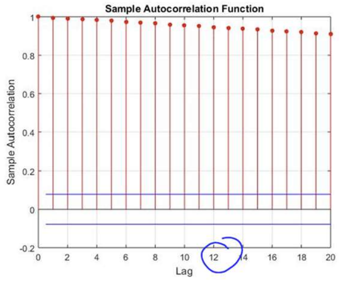
or together

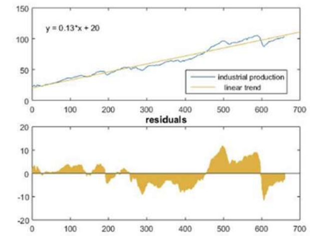
library(ggplot2)

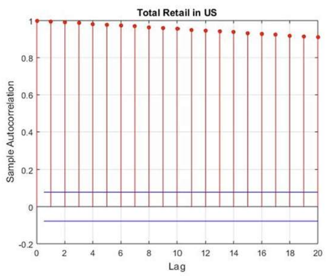
tsdisplay(x)

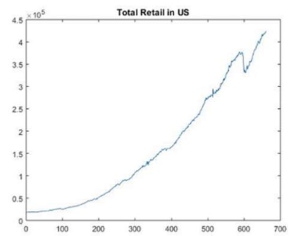
- Stationary Series
    - ACF behavior: Autocorrelations drop quickly toward zero as lag increases.
    - Interpretation: The series has no long-term trend; shocks are temporary.
    - Visual cue: Bars in the correlogram fall within the significance bounds after a few lags.
    - Example: White noise or detrended series.
- Trending (Nonstationary) Series
    - ACF behavior: Autocorrelations decay slowly or remain very high even at large lags.
    - Interpretation: The series has a trend or unit root; shocks have lasting effects.
    - Visual cue: Bars stay above the significance bounds for many lags.
    - Example: Raw GDP or cumulative sales data.
- Quick Rule of Thumb
    - Fast drop → stationary.
    - Slow decay / persistent high autocorrelation → trending / nonstationary.

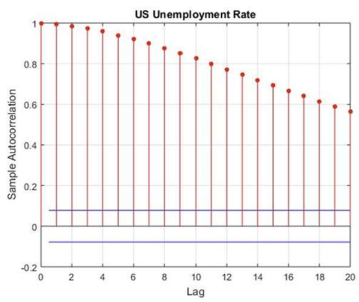
Trending Series

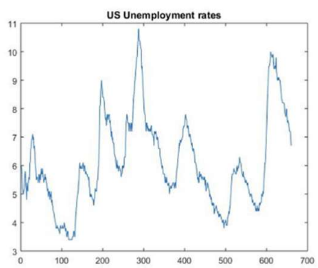




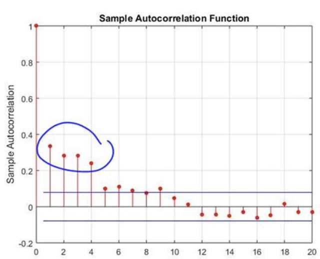


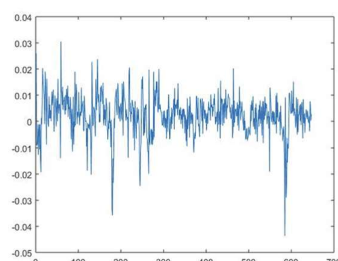


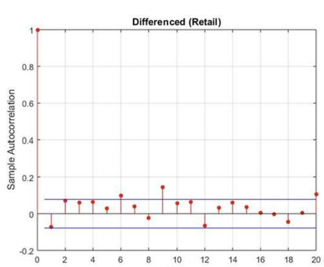


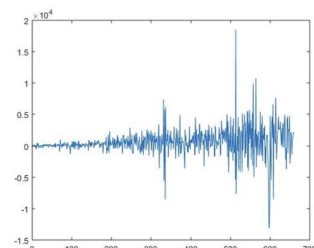


Stationary Series









c) ADF Test

Hypothesis	Meaning
H_0 (null)	Series **has a unit root** → non-stationary
H_1 (alternative)	Series **does not have a unit root** → stationary

- p-value < 0.05 → reject H_0 → series is stationary
- p-value ≥ 0.05 → fail to reject H_0 → series is non-stationary

Install and load package

install.packages("tseries")

library(tseries)

Run ADF test

adf_result <- *adf.test(data_ts)*

View result

adf_result

d) KPSS Test (Kwiatkowski–Phillips–Schmidt–Shin)

H_0 (null)	The series is **stationary**
H_1 (alternative)	The series is **non-stationary** (has a unit root)

- p < 0.05 → the series is NOT stationary
- p ≥ 0.05 → the series is stationary

Install and load package

install.packages("urca")

library(urca)

Run KPSS test

kpss_result <- *ur.kpss(data_ts)*

Get summary

summary(kpss_result)

e) Decomposition

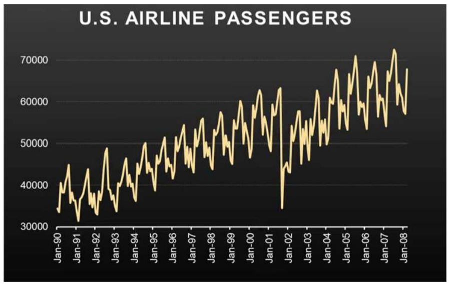
Method	When to use	Pros / Cons
Classical (`decompose()`)	Regular series, **fixed seasonal pattern**, additive or multiplicative	Simple, easy to visualize; cannot handle changing seasonal patterns
STL (`stl()`)	Flexible, can handle **changing seasonal amplitude**, robust trend	Handles outliers, seasonal component can vary over time; more flexible than classical
X-11 / SEATS / TBATS	Complex series: multiple seasonalities (daily + weekly + yearly), or long series with irregular cycles	TBATS can model multiple seasonal periods, non-integer frequencies; SEATS/X-11 used for official economic statistics

Classical

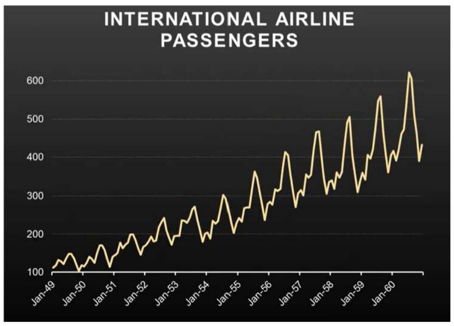
decompose(data_ts, type = c("additive", "multiplicative"), filter = NULL)

- Series has stable and regular seasonality.
- decompose() requires the time series to have a frequency > 1. Otherwise, seasonality cannot be computed.
- Data is not noisy and trend/seasonal patterns are simple.
- Quick and simple method for understanding data structure.

- Additive
    - Use when **seasonal fluctuations are roughly constant**, regardless of level



- Multiplicative
    - Use when seasonal fluctuations grow with the level of the series



STL decomposition

library(forecast)

stl_decomp <- stl(data_ts, s.window="periodic")

Fixed seasonal pattern

stl_fixed <- stl(data_ts, s.window="periodic")

Flexible seasonal pattern

stl_flexible <- stl(data_ts, s.window=13)

- smooth over 13 months

plot(stl_decomp)

- More flexible, can handle changing seasonality
- Allows robust trend estimation
    - "periodic" → if the seasonal pattern does not change over time
    - Numeric → if the seasonal effect changes gradually (e.g., increasing holiday sales over years)

SEATS decomposition

install.packages("seasonal")

library(seasonal)

Fit the X-11/SEATS model

x11_fit <- seas(data_ts, x11 = "")

- seas() is the main function.
- x11 = "" → tells the function to use the **X-11 decomposition method**.

View the results

summary(x11_fit)

plot(x11_fit)

- Plots the original series, trend, seasonal component, and remainder.

Split the data into Train/Test

- Test set must be same length as forecast Horizon to evaluate later!!!
- Common rules of thumb:
  - 70–90% for training, 10–30% for testing
  - For example, if you have 2010–2025 data:
    - Training: 2010–2022
    - Testing: 2023–2025
- window() ensures the time continuity is preserved
- No random sampling!

Example: 2010–2025 monthly

sales_ts <- ts(sales, start=c(2010,1), frequency=12)

Train: 2010-2022

train_ts <- window(sales_ts, end=c(2022,12))

Test: 2023-2025

test_ts <- window(sales_ts, start=c(2023,1))

Example Yearly data

- Frequency = 1 (one observation per year)
- Example: 1995–2005

  # yearly time series

  yearly_ts <- ts(data$Sales, start=1995, frequency=1)

  # Split

  train_ts <- window(yearly_ts, end=2002)

  test_ts <- window(yearly_ts, start=2003)
- end and start are just years, no months/quarters.

Example Quarterly data

- Frequency = 4 (4 quarters per year)
- Example: Q1 1995 → Q4 2000

  # quarterly time series

  quarterly_ts <- ts(data$Sales, start=c(1995,1), frequency=4)

  # Split

  train_ts <- window(quarterly_ts, end=c(2000,3)) # Q3 2000

  test_ts <- window(quarterly_ts, start=c(2000,4)) # Q4 2000

Example Monthly Data

- Frequency = 12 (12 months per year)
- Example: Jan 1995 → Dec 2000

  # monthly time series

  monthly_ts <- ts(data$Sales, start=c(1995,1), frequency=12)

  # Split

train_ts <- window(monthly_ts, end=c(2000,10)) # Oct 2000

test_ts <- window(monthly_ts, start=c(2000,11)) # Nov 2000

Example Daily data

- Frequency = 365 (or 7 for weekly if needed)
- Example: Jan 1, 1995 → Dec 31, 1997

  # daily time series

  daily_ts <- ts(data$Sales, start=c(1995,1), frequency=365)

  # Split

  train_ts <- window(daily_ts, end=c(1997,120)) # day 120 of 1997

  test_ts <- window(daily_ts, start=c(1997,121)) # day 121 of 1997
- Daily data uses c(year, day_number) → tricky, because window() counts periods in the year.
- For daily data, often it's easier to use row indices or tsibble/zoo objects.

First steps find out stationary or not and

Choose Forecasting Method

TABLE 6 Choosing a Forecasting Technique

Method	*Pattern of Data*	*Time Horizon*	*Type of Model*	*Minimal Data Requirements*	
				Nonseasonal	*Seasonal*
Naive	ST, T, S	S	TS	1	
Simple averages	ST	S	TS	30	
Moving averages	ST	S	TS	4–20	
Exponential smoothing	ST	S	TS	2	
Linear exponential smoothing	T	S	TS	3	
Quadratic exponential smoothing	T	S	TS	4	
Seasonal exponential smoothing	S	S	TS		$2 \times s$
Adaptive filtering	S	S	TS		$5 \times s$
Simple regression	T	I	C	10	
Multiple regression	C, S	I	C	$10 \times V$	
Classical decomposition	S	S	TS		$5 \times s$
Exponential trend models	T	I, L	TS	10	
S-curve fitting	T	I, L	TS	10	
Gompertz models	T	I, L	TS	10	
Growth curves	T	I, L	TS	10	
Census *X*-12	S	S	TS		$6 \times s$
Box-Jenkins	ST, T, C, S	S	TS	24	$3 \times s$
Leading indicators	C	S	C	24	
Econometric models	C	S	C	30	
Time series multiple regression	T, S	I, L	C		$6 \times s$

Pattern of the data: ST, stationary; T, trending; S, seasonal; C, cyclical
Time horizon: S, short term (less than three months); I, intermediate term; L, long term
Type of model: TS, time series; C, causal
Seasonal: s, length of seasonality
Variable: V, number of variables

Table: A Guide to Selecting an Appropriate Forecasting Method *

Forecasting Method	Data Pattern	Quantity of Historical Data (Number of Observations)	Forecast Horizon
Naïve	Stationary	1 or 2	Very short
Moving averages	Stationary	Number equal to the periods in the moving average	Very short
Exponential smoothing			
Simple	Stationary	5 to 10	Short
Adaptive response	Stationary	10 to 15	Short
Holt's	Linear trend	10 to 15	Short to medium
Winters'	Trend and seasonality	At least 4 or 5 per season	Short to medium
Bass model	S-curve	Small, 3 to 10	Medium to long
Regression-based			
Trend	Linear and non-linear trend with ot without seasonality	Minimum of 10 with 4 or 5 per season if seasonality is included	Short to medium
Causal	Can handle nearly all data patterns	Minimum of 10 per independent variable	Short, medium & long
TS decomposition	Can handle trend, seasonal and cyclical patterns	Enough to see two peaks and troughs in the cycle	Short, medium & long
ARIMA	Stationary or transformed to stationary	Minimum 50	Short, medium & long

1. Time-Series Forecasting (Univariate)

Method	Data Pattern	Time Horizon	Minimum Historical Data	Notes
Naive	Stationary, trending, seasonal	Short (<3 months)	1 point	Last value = forecast.
Simple Averages	Stationary	Short	5–10 points	Mean of last N observations.
Moving Averages	Stationary	Short	Window length	Smooths noise.
Simple Exponential Smoothing (SES)	Stationary	Short	5–10 points	Weighted decaying average.
Adaptive Response ES	Stationary	Short	10–15 points	Dynamic smoothing.
Adaptive Filtering	Seasonal	Short	2 × s (24 months for monthly)	Seasonal smoothing.
Holt's Linear	Trend	Short–Medium	10–15 points	Level + trend.
Holt–Winters (Winters)	Trend + Seasonal	Short–Medium	4–5 per season	Additive or multiplicative seasonality.
Quadratic Exponential Smoothing	Quadratic trend	Short	≥4 points	Captures curvature.
Seasonal ES (Holt-Winters variant)	Seasonal	Short	2 × s	Seasonal exponential smoothing.
Bass Diffusion	S-curve	Medium–Long	3–10 points	Product adoption.
Classical Decomposition	Seasonal	Short	5 × s (60 monthly)	Requires long stable seasonality.
TS Decomposition (e.g., STL)	Trend + Seasonal + Cyclical	Short–Long	Enough for 2 peaks/troughs	Flexible decomposition.
Exponential Trend Models	Trend	Medium–Long	~10 points	Exponential growth patterns.
S-Curve Fitting / Gompertz / Growth Curves	Trend	Medium–Long	~10 points	Logistic/S-curve growth.
Box–Jenkins (ARIMA Framework)	Stationary, trend (via differencing), seasonal	Short	≥24 points	Identification-oriented.
ARMA	Stationary with autocorrelation	Short–Medium	≥50 points	Autoregressive + moving average.
ARIMA	Stationary (or transformed)	Short–Long	≥50 points	Trend handled via differencing.
SARIMA	Trend + Seasonality + autocorrelation	Short–Medium	≥50 points	Seasonal ARIMA.

Classic decomposition and TS decomposition not for forecasting

2. Dynamic Time-Series Models (TS + Explanatory Variables)

Method	Data Pattern	Time Horizon	Minimum Historical Data	Notes
ADL (Autoregressive Distributed Lag)	Stationary (or differenced)	Short–Medium	≥30 observations	Lags of Y and X included.
ARDL	Mix of I(0)/I(1); cointegration	Short–Medium	30–50 points	Short-run + long-run effects.
VAR	Multivariate stationary	Short–Medium	50–100+ points	All variables depend on each other's lags.
IRF	Same as VAR	Diagnostics	Same as VAR	Shock-response tool, not a forecasting model.

---

3. Regression-Based Forecasting (Causal + Time-Series Regression)

(This section is now fully integrated to reflect that regression can be used directly for time-series forecasting.)

Regression Method	Data Pattern	Time Horizon	Min Historical Data	Time-Series Suitability
Simple Regression	Trend	Short–Long	~10 points	Often used to detrend or model structural trend.
Multiple Regression	Seasonal + Cyclical	Short–Long	10 × number of variables	Supports seasonality via dummies; AIC/BIC selection.
Time-Series Multiple Regression	Trend + Seasonal	Medium–Long	6 × s (e.g., 72 months monthly)	Deterministic seasonality + structural drivers.
Trend-Based Regression	Linear/nonlinear trend, with or without seasonality	Short–Medium	≥10 + 4–5 per season	Flexible trend modeling.
Causal Regression (General)	Nearly all types	Short–Long	≥10 per variable	Broad causal/structural modeling framework.
Regression with Lags (ADL form)	Trends, seasonality, autocorrelation	Short–Medium	≥30 points	Bridge between regression and time-series (dynamic regression).
Regression with ARIMA Errors	Trend + Seasonal + stochastic residuals	Short–Long	≥50 points	Handles remaining autocorrelation after regression.
Leading Indicator Models	Cyclical predictors	Short–Long	≥24 points	Example: composite leading indicators (CLI).
Econometric Structural Models	Cyclical, structural	Short–Long	≥30 points	Theory-driven forecasts.

Time series forecasting 1 Variable:

Naive

- Stationary, Trending, Seasonal
- Time Horizon forecast: Short, less than 3 months
- Number of past observation: You only need one past data point (e.g., last month's value).
- Time Series

Naive Forecast.R

Simple Averages

- Stationary
- Time Horizon forecast: Short, less than 3 months
- Time Series
- Number of past observation: 5 - 10 Observations (Example: Average of the last 6 months of sales used as forecast)

simple moving average in excel.xlsx

Moving Averages

- Stationary
- Time Horizon forecast: Short, less than 3 months
- Time Series
- Number of past observation: equal to the periods in the moving average (Example: 3-month moving average → use last 3 months' values)
- How to know which Moving Average to use?

Data Frequency	Example Data	Typical Seasonal Cycle	Suggested MA Windows	Forecast Horizon (h)	Explanation
Daily	Website visits (3 months)	Weekly (7 days)	3, 5, 7	1–7 days	MA3 → reacts quickly; MA5 → smooths ~workweek; MA7 → full weekly cycle
Weekly	Retail store weekly sales	Monthly (~4 weeks)	2, 4, 6	1–4 weeks	MA2 → fast reaction; MA4 → smooths 1 month; MA6 → slightly longer trend
Monthly	Electricity consumption (3 years)	Yearly (12 months)	3, 6, 12	1–12 months	MA3 → 3-month smoothing; MA6 → half-year; MA12 → full year
Quarterly	GDP over 5 years	Yearly (4 quarters)	2, 4, 6	1–4 quarters	MA2 → half-year; MA4 → full-year; MA6 → 1.5 years
Yearly	Population growth over 50 years	Multi-year	2, 3, 5	1–5 years	MA2 → 2-year smoothing; MA3 → medium; MA5 → long-term trend

Moving Averages.R

Exponential smoothing

Simple

- Stationary
- Time Horizon forecast: Short, less than 3 months
- Number of past observation: 5 - 10 historical Observations (example monthly data: Forecast for October uses weighted average of last 5–10 months)
- Time Series
- Beta true or False:

Parameter	Meaning	When to use
`beta = FALSE`	No trend is included. Only the **level** of the series is smoothed.	Use when your time series is **stationary** (no clear upward or downward trend). Example: monthly sales that fluctuate around a constant mean.
`beta = TRUE`	Include a **trend component.** The model smooths both the level and the trend.	Use when your series shows a **linear or approximately linear trend** over time. Example: monthly revenue steadily increasing each month.

Simple Exponential smoothing.R

Adaptive Response

- Stationary
- Time Horizon forecast: Short, less than 3 months
- Number of past observation: 10 - 15 historical Observations (example monthly data: Forecast for October uses weighted last 12 months)
- Time Series
- Which model ANN; AAN; AAA?

Model	Level	Trend	Seasonality	Notes
ANN	Additive	None	None	Only smooths the level of the series. Equivalent to **simple exponential smoothing (SES).** Use when there is **no trend or seasonality.**
AAN	Additive	Additive	None	Smooths level and trend. Equivalent to **Holt's linear method.** Use when series shows a **linear trend** but no seasonality.
AAA	Additive	Additive	Additive	Smooths level, trend, and seasonal component. Equivalent to **Holt-Winters additive.** Use when series shows both **trend and additive seasonality.**

SES adaptive response.R

Adaptive filtering

- Seasonality
- Time Horizon forecast: Short, less than 3 months
- Number of past observation: 2 × s Example if monthly (s = 12) - 2 × 12 = 24 months (2 years)
- Time Series

SES adaptive filtering.R

Holts (linear exponential smoothing)

- Linear and Trend
- Time Horizon forecast: Short, less than 3 months to medium
- Number of past observation: 10 - 15 historical Observations (example monthly data: Forecast linear trend using last 12 months)
- Time Series

Holt's Linear Trend.R

Quadratic exponential smoothing (Holt extended to quadratic trend)

- Trend
- Time Horizon forecast: Short, less than 3 months
- Number of past observation: At least 4 points to capture curvature. (example monthly data: Last 4 months enough to capture curvature)
- Time Series

Quadratic Exponential Smoothir

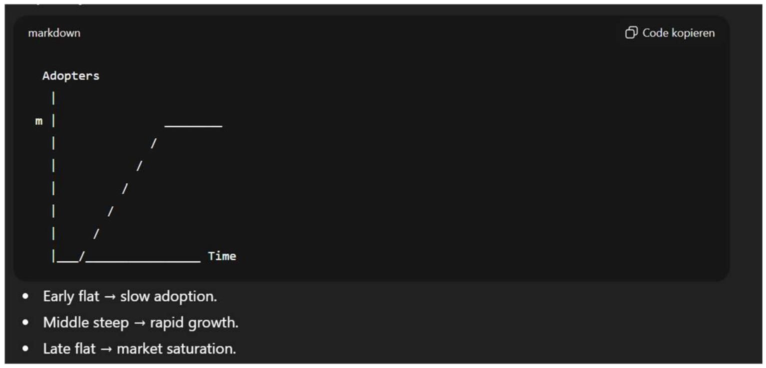
Seasonal exponential smoothing (Holt-Winters)

- Seasonal
- Time Horizon forecast: Short, less than 3 months
- Number of past observation: 2 × s Example if monthly (s = 12) - 2 × 12 = 24 months (2 years)
- Time Series

Seasonal exponential smoothing (Holt-Wint

Bass

- The Bass model is a specific type of S-curve designed for new product adoption.
- S curve
    - An **S-curve** is a **cumulative growth pattern** often seen in:
        - Product adoption
        - Technology diffusion
        - Market penetration
    - **Characteristics:**
        - **Slow start:** Early adopters → innovators begin adoption.
        - **Rapid growth:** Imitators accelerate adoption as more people adopt.
        - **Saturation:** Growth slows as market approaches maximum potential (m in Bass model).



- Time Horizon forecast: Medium to long
- Number of past observation: 3 - 10 historical Observations (example monthly data: Use first 6 months of product adoption to forecast next month)
- Time Series

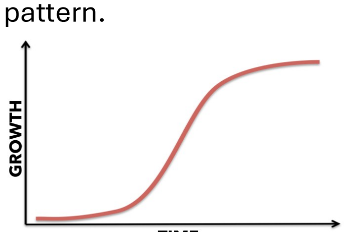
Exponential Trend models

- Trend
- time series forecasting techniques that account for data that is growing or decaying at an accelerating rate
- Assumes the data grows/decays multiplicatively
  Example: 10, 12, 14.4, 17.28 → each period grows by **20% of the current value**.
- Time Horizon forecast: Medium to long
- Number of past observation: 10
- Time Series

Exponential Trend Forecast.R

S-curve fitting

- Trend
- Time Horizon forecast: Medium to long
- An S-curve is a general term for any curve with a sigmoid (S-shaped) growth pattern.

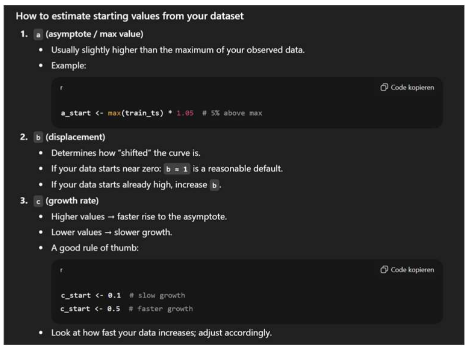


- Number of past observation: 10
- Time Series

S-Curve - Logistic Growth Forecast.R

Gompertz Models

- Trend
- It works well only for data that has an S-curve pattern.
- Time Horizon forecast: Medium to long
- Number of past observation: 10
- Time Series



How to estimate starting values from your dataset

1. `a` (asymptote / max value)
   - Usually slightly higher than the maximum of your observed data.
   - Example:

```r
a_start <- max(train_ts) * 1.05  # 5% above max
```

2. `b` (displacement)
   - Determines how "shifted" the curve is.
   - If your data starts near zero: `b = 1` is a reasonable default.
   - If your data starts already high, increase `b`.
3. `c` (growth rate)
   - Higher values → faster rise to the asymptote.
   - Lower values → slower growth.
   - A good rule of thumb:

```r
c_start <- 0.1  # slow growth
c_start <- 0.5  # faster growth
```

   - Look at how fast your data increases; adjust accordingly.

Gompertz Forecast.R

Box Jenkinson

- Stationary, Trend, Seasonality
- Time Horizon forecast: Short, less than 3 months
- Number of past observation: 24
- Time Series

Manual ARIMA or Box-Jenkins ARIMA.R

ARMA

- Stationary, autocorrelation (AR + MA)
- If series is non-stationary, you must difference it (then it becomes ARIMA instead of pure ARMA)
- Time Horizon forecast: Short to medium
- Number of past observation: min 50.
- Only for stationary series; non-stationary → difference → ARIMA
- Time Series

ARMA.R

ARIMA

- Stationary or transformed to stationary

- Time Horizon forecast: short, medium, long
- Number of past observation: min 50.
- Time Series

✅ When to use which

Scenario	stepwise	approximation	Notes
Small/medium dataset (<~500 observations)	FALSE	FALSE	Full, accurate search.
Large dataset (>~1000 observations)	TRUE	TRUE	Faster search, may sacrifice slight accuracy.
Large dataset, want more accurate model	FALSE	TRUE	Full search with approximate likelihoods.

Auto ARIMA.R

SARIMA

- SARIMA only works if your ts() object has a seasonal frequency (frequency > 1)
- Trend, Seasonality, Autocorrelation (AR + MA + seasonal)
- Time Horizon forecast: Short to medium
- Number of past observation: min 50.
- Extends ARIMA to include seasonal effects; series may need differencing
- Time Series

✅ When to use which

Scenario	stepwise	approximation	Notes
Small/medium dataset (<~500 observations)	FALSE	FALSE	Full, accurate search.
Large dataset (>~1000 observations)	TRUE	TRUE	Faster search, may sacrifice slight accuracy.
Large dataset, want more accurate model	FALSE	TRUE	Full search with approximate likelihoods.

SARIMA.R

2 Variables:

First check if both variables are stationary or not

If BOTH variables are stationary (I(0)).

This is the BEST case because ALL models are allowed.

if ONE variable is stationary (I(0)) and the other is non-stationary

You MUST use ARDL/ADL models (BEST CHOICE)

ARDL was created exactly for situations where variables have mixed integration orders.

BOTH variables non-stationary (I(1))

Test if variables are cointegrated

Phillips–Ouliaris Test (Engle–Granger)

Or Johansen Test (for >1 x)

Allowed forecasting models when NOT cointegrated

Difference both, then use ADL/VAR/Regression

COINTEGRATED

ECM

VECM

Use Static Regression on levels (Run a normal regression using the raw (non-differenced) values of Y and X.)

two variables, check if cointegrated and de

Causal/ Static Regression Models

These models focus on the relationships between variables at a single point in time and do not require data to be a time series

Simple Regression

- Trend
- can use it to detrend or deseasonalise the data.
- For model selection, we minimize one of the following: Akaike information criterion (AIC), Bayesian information criterion (BIC)
- Number of past observation: 10

Simple Regression.R

Multiple Regression

- Cyclical and seasonal
- can use it to detrend or deseasonalise the data.
- The multiple regression setup works for time series data.
- For model selection, we minimize one of the following: Akaike information criterion (AIC), Bayesian information criterion (BIC)
- Number of past observation: 10 x V

Multiple Regression.R

Dynamic Forecasting

Past: test importance → statistical evidence

ADL

- All variables must be stationary or made stationary

- Lagged dependent and independent variables (dynamic effects)
- Dependent variable: Usually one main variable that you are trying to forecast or explain.
    - Example: Monthly sales.
- Independent variables: Can be one or more predictors.
    - Example: Advertising spend, temperature, promotions.
- When to use it:
    - You have one main variable to forecast (univariate) or a few predictors (small multivariate).
    - You want to capture short-term dynamics — e.g., how past sales or marketing spend affects current sales.
    - Significant test

      The outcome of the ADL model:

      ```
      Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
      (Intercept) 24.93273    7.67834   3.247 0.001234 **
      L(pass, 1)   0.62624    0.03565  17.567  < 2e-16 ***
      L(pass, 2)   0.08686    0.03479   2.497 0.012807 *
      L(pass, 3)   0.09824    0.03505   2.803 0.005233 **
      L(pass, 4)   0.12612    0.03466   3.639 0.000298 ***
      L(pass, 5)  -0.12286    0.03443  -3.568 0.000390 ***
      L(pass, 6)  -0.03465    0.03492  -0.992 0.321439
      L(pass, 7)   0.64832    0.03499  18.529  < 2e-16 ***
      L(pass, 8)  -0.49252    0.03513 -14.019  < 2e-16 ***
      si          -0.59551    0.43060  -1.383 0.167218
      L(si, 1)     0.72669    0.62080   1.171 0.242264
      L(si, 2)    -0.60396    0.61762  -0.978 0.328544
      L(si, 3)    -0.86173    0.61055  -1.411 0.158675
      L(si, 4)     0.61996    0.61124   1.014 0.310886
      L(si, 5)    -1.14089    0.60250  -1.894 0.058784 .
      L(si, 6)     0.87456    0.59709   1.465 0.143551
      L(si, 7)     0.30129    0.59590   0.506 0.613328
      L(si, 8)     0.18926    0.59504   0.318 0.750553
      L(si, 9)    -0.92374    0.59504  -1.552 0.121124
      L(si, 10)    1.08877    0.43092   2.527 0.011787 *
      ---
      Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
      ```

      If X has **no significant lags**, then X does not help explain Y, in that case do not use ADL
- Preferably stationary series (or you difference first).
- Time horizon: Short to medium.
- Data needed: At least 30 observations to estimate the lags reliably.
- Example: Forecasting monthly electricity demand using past electricity demand and past temperature.

  ADL.R

ARDL

- If stationary: ARDL works like ADL.
- What it is: A generalization of ADL. Can handle variables that are a mix of stationary (I(0)) and non-stationary (I(1)).

1 ARDL Stationarity Requirements

- When people say:

> "ARDL: Can be I(0) or I(1), Cannot be I(2)"

This is about **integration order:**

Term	Meaning
I(0)	Stationary at level (no differencing needed)
I(1)	Becomes stationary after first differencing
I(2)	Becomes stationary only after second differencing

- When to use it:
    - You want to model both short-run and long-run relationships.
    - Some variables are non-stationary but you still want to use levels rather than differencing everything.
- Often used in economics for cointegration analysis.
- Time horizon: Short to medium.
- Data needed: 30–50 observations minimum.
- Example: Modeling GDP growth with interest rates and investment, where some variables are trending over time

ARDL.R

VAR

All variables must be stationary

- A multivariate time series model where every variable depends on its own past and the past of all other variables.
- When to use it:
- You have multiple interdependent variables.
- You want to capture how shocks or changes in one variable affect the others.
- Series should be stationary (or differenced if non-stationary).
- Time horizon: Short to medium.
- Data needed: ≥50–100 observations (more if more variables or lags).
- Example: Forecasting GDP, inflation, and interest rates together; each depends on its own past and the past of the other two.

VAR.R

For dynamic forecating and regression models use this Test

⭐ 1. Breusch–Pagan Test (BP test)

📌 What it tests

Whether residuals have **constant variance (homoscedasticity).**

Null hypothesis:

- **H_0:** Residuals are homoscedastic (constant variance)
- **H_1:** They are heteroscedastic (variance changes)

📌 Interpretation

- **$p > 0.05$ → GOOD**
  Residuals have constant variance → OK
- **$p < 0.05$ → BAD**
  Heteroscedasticity → not ideal but **NOT fatal** for forecasting.

📌 When to proceed?

➡️ You can **still use the model** even if BP is significant.
Forecasting is robust to heteroscedasticity.

📌 When to stop?

➡️ Only if variance changes are **extreme** or if you do structural analysis, not forecasting.

Conclusion:
BP test failing is **not a model killer** in ADL forecasting.

⭐ 2. Durbin–Watson Test / Breusch–Godfrey (Autocorrelation)

📌 What it tests

Whether residuals are autocorrelated.
Null hypothesis:

- **H_0:** No autocorrelation
- **H_1:** Autocorrelation exists

📌 Interpretation

- **$p > 0.05$ → GOOD**
  No autocorrelation → OK
- **$p < 0.05$ → VERY BAD**
  There is autocorrelation → model is **invalid**.

📌 Why is autocorrelation so bad?

Because it means:

- the model **misses lags**
- the dynamic structure is wrong
- you did not include enough AR lags
- or X is missing important dynamics

This destroys the validity of ADL.

📌 When to proceed?

➡️ Only when **residual autocorrelation is NOT significant**.

📌 When to stop?

❌ If autocorrelation exists → **STOP. Reject ADL.**
Try:

- increase p_range
- include more Y-lags
- or switch to ARIMA or VAR

Conclusion:
Autocorrelation is the **most important diagnostic**.
If it fails → ADL is not acceptable.

⭐ 3. Jarque–Bera Normality Test

📌 What it tests

Whether residuals are normally distributed.

Null hypothesis:

- **H_0:** Residuals are normal
- **H_1:** Residuals are non-normal

📌 Interpretation

- **$p > 0.05$ → GOOD**
  Residuals normal → fine
- **$p < 0.05$ → NOT IDEAL**
  Residuals non-normal → BUT forecasting is still allowed.

📌 When to proceed?

➡️ You **can still use the model** for forecasting.
Normality is NOT required for time series forecasting.
(Only required for confidence intervals and hypothesis testing.)

📌 When to stop?

➡️ You do **NOT** stop because of JB test alone.

Conclusion:
JB failing is **not a reason** to reject the model for forecasting.

Forecast Combination:

combining forecast models.R

Gliederung:

1. Introdruction (Purpose of study, scope of study, sources and methods)
2. Analysis and findings
    - 2.1 -> is the data y variable or not, seasonal or trending explain how this was found out
    - 2.2 Decomposition -> y variable and explain what is seen
    - 2.3 Model 1 -> time series model explain what was done and how
    - 2.4 Model 2 -> time series model explain how it was done and outcome (ARIMA)
        - 2.4.1 Manuell ARIMA
        - 2.4.2 Auto ARIMA
    - 2.5 Model 3 -> Exponantial smoothing Method (Simple, HOlt, Holt winters etc.)
    - 2.6 explain variable x is stationary or not? transformation?

    Have a look at variable x

    if we have two variables.

    First check if both variables are stationary or not

    - If BOTH variables are stationary (I(0)).
        - This is the BEST case because ALL models are allowed.
    - if ONE variable is stationary (I(0)) and the other is non-stationary
        - You MUST use ARDL/ADL models (BEST CHOICE)
        - ARDL was created exactly for situations where variables have mixed integration orders.
    - BOTH variables non-stationary (I(1))
        - Test if variables are cointegrated
            - Phillips–Ouliaris Test (Engle–Granger)
            - Or Johansen Test (for >1 x)
        - Allowed forecasting models when NOT cointegrated
            - Difference both, then use ADL/VAR/Regression
        - COINTEGRATED
            - ECM
            - VECM
            - Use Static Regression on levels-> Test Breusch-Pagan, Durbin-Watson and Jarque-Bera tests -> concluded that there are big problems with this regression and that the results of it are not trustworthy
    - 2.7 Model 4 -> Causal/ Static Regression Models ->
    - 2.8 Model 5 -> Dynamic Forecasting (VAR and ADL)

2.8 Comparasion of the models and combination models -> compare all model RMSE etc. -> combination of models equal weights and Granger-Ramathan method

-> compare RMSE etc.

3. Conclusions -> show final model and plot
4. Recommendations