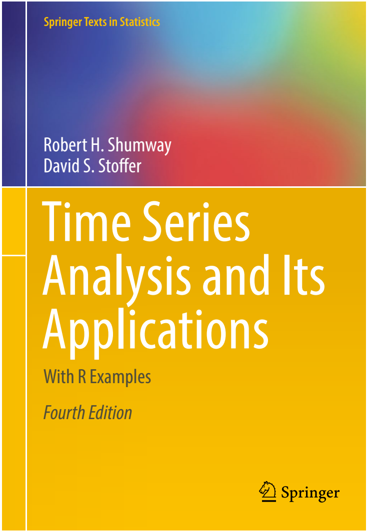
Intro to Forecasting and Predictive Analytics
Module 2: Characteristics of Time Series

Winter 2024 Edition

Textbook


Springer Texts in Statistics
Robert H. Shumway
David S. Stoffer
Time Series Analysis and Its Applications
With R Examples
Fourth Edition
Springer

Topics Covered in These Slides

1. **Time Series Statistical Models**
   - Introduction
   - White Noise Series
2. **Measures of Dependence**
   - Mean Function
   - Autocorrelation Function
3. **Stationary Time Series**
   - Definition
   - Characteristics

Topics Covered in These Slides

1. **Time Series Statistical Models**
   - Introduction
   - White Noise Series
2. **Measures of Dependence**
   - Mean Function
   - Autocorrelation Function
3. **Stationary Time Series**
   - Definition
   - Characteristics

Topics Covered in These Slides

1. **Time Series Statistical Models**
   - Introduction
   - White Noise Series
2. **Measures of Dependence**
   - Mean Function
   - Autocorrelation Function
3. **Stationary Time Series**
   - Definition
   - Characteristics

Outline

Time Series Analysis

Definition (Time Series)

A **time series** is a sequence of values, $X_0, X_1, X_2, \ldots$, recorded at different points in **time**, where the **random variable** X_t represents the **value** at a specific **time** t.

Time Series Analysis

Time series analysis studies data collected over time, facing **unique challenges** due to the natural **ordering** of the data points.

Conventional Statistical Methods vs. Time Series Analysis

Unlike standard **statistical methods** that assume **independent** data, time series data points are often strongly **correlated**, making traditional approaches difficult to apply.

Time Series Analysis

Definition (Time Series)

A **time series** is a sequence of values, $X_0, X_1, X_2, \ldots$, recorded at different points in **time**, where the **random variable** X_t represents the **value** at a specific **time** t.

Time Series Analysis

Time series analysis studies data collected over time, facing **unique challenges** due to the natural **ordering** of the data points.

Conventional Statistical Methods vs. Time Series Analysis

Unlike standard **statistical methods** that assume **independent** data, time series data points are often strongly **correlated**, making traditional approaches difficult to apply.

Time Series Analysis

Definition (Time Series)

A **time series** is a sequence of values, $X_0, X_1, X_2, \ldots$, recorded at different points in **time**, where the **random variable** X_t represents the **value** at a specific **time** t.

Time Series Analysis

Time series analysis studies data collected over time, facing **unique challenges** due to the natural **ordering** of the data points.

Conventional Statistical Methods vs. Time Series Analysis

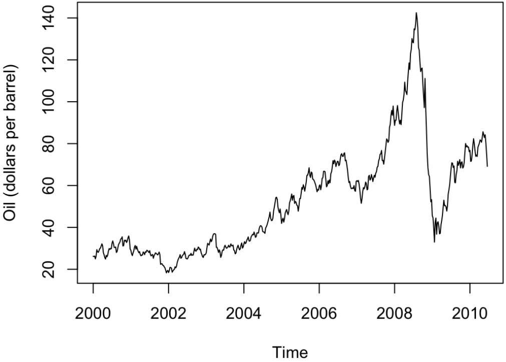
Unlike standard **statistical methods** that assume **independent** data, time series data points are often strongly **correlated**, making traditional approaches difficult to apply.

Plotting Time Series Data

- The **first step** in **analyzing** time series data is to carefully examine a **plot** of the data points over **time**.
- By studying the **time series plot**, we can gain insights into which **methods** might be useful for effectively **modeling** the data.

Plotting Time Series Data

- The **first step** in **analyzing** time series data is to carefully examine a **plot** of the data points over **time**.
- By studying the **time series plot**, we can gain insights into which **methods** might be useful for effectively **modeling** the data.



White Noise

Definition (White Noise)

A sequence of **uncorrelated** random variables $W_0, W_1, W_2, \ldots$, all with a mean of 0 and a finite variance σ_W^2, is known as **white noise**, denoted by $W_t \sim \mathtt{WN}[0, \sigma_W^2]$.

- The designation **white** originates from the analogy with **white light** which has the **same** intensity at every **frequency**.
- A particularly **useful** type of white noise series is **Gaussian white noise**, wherein the W_t are independent **Normal** random variables, $W_t \overset{\text{iid}}{\sim} \mathtt{N}[0, \sigma_W^2]$

White Noise

Definition (White Noise)

A sequence of **uncorrelated** random variables $W_0, W_1, W_2, \ldots$, all with a mean of 0 and a finite variance σ_W^2, is known as **white noise**, denoted by $W_t \sim \mathtt{WN}[0, \sigma_W^2]$.

- The designation **white** originates from the analogy with **white light** which has the **same** intensity at every **frequency**.
- A particularly **useful** type of white noise series is **Gaussian white noise**, wherein the W_t are independent **Normal** random variables, $W_t \overset{\text{iid}}{\sim} \mathtt{N}[0, \sigma_W^2]$

White Noise

Definition (White Noise)

A sequence of **uncorrelated** random variables $W_0, W_1, W_2, \ldots$, all with a mean of 0 and a finite variance σ_W^2, is known as **white noise**, denoted by $W_t \sim \mathtt{WN}[0, \sigma_W^2]$.

- The designation **white** originates from the analogy with **white light** which has the **same** intensity at every **frequency**.
- A particularly **useful** type of white noise series is **Gaussian white noise**, wherein the W_t are independent **Normal** random variables, $W_t \overset{\text{iid}}{\sim} \mathtt{N}[0, \sigma_W^2]$

Outline

Mean Function

Definition (Mean Function)

Let $X_0, X_1, X_2, \ldots$ be a **time series**. The **mean function** of this time series at time t is denoted by $\mu_X(t)$ and defined as

$$\mu_X(t) := \mathbb{E}[X_t] = \int_{-\infty}^{\infty} x f_{X_t}(x)dx$$

- Throughout, if **no confusion** arises regarding which time series is being discussed, we will **simplify** our notation by **removing** the subscript X across all instances, such as simplifying $\mu_X(t)$ to $\mu(t)$

Mean Function

Definition (Mean Function)

Let $X_0, X_1, X_2, \ldots$ be a **time series**. The **mean function** of this time series at time t is denoted by $\mu_X(t)$ and defined as

$$\mu_X(t) := \mathbb{E}[X_t] = \int_{-\infty}^{\infty} x f_{X_t}(x) dx$$

- Throughout, if **no confusion** arises regarding which time series is being discussed, we will **simplify** our notation by **removing** the subscript X across all instances, such as simplifying $\mu_X(t)$ to $\mu(t)$

Example

Example (Simple Ransom Walk)

A simple **random walk** model can explain **stock price** dynamics, where today's price equals yesterday's price plus random noise.

$$\begin{cases} X_t = X_{t-1} + W_t & \text{for } t = 1, 2, \dots \\ X_0 = 0 & \text{for } t = 0 \end{cases}$$

where $W_t \overset{\text{iid}}{\sim} \mathtt{N}[0, \sigma_W^2]$

1. **Show** that $X_t = \sum_{j=1}^{t} W_j$ for $t = 1, 2, \dots$
2. Find the **mean** function at time t

Example

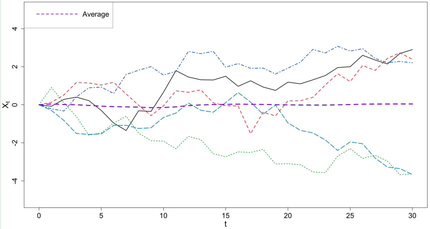
Example (Simple Ransom Walk)

A simple **random walk** model can explain **stock price** dynamics, where today's price equals yesterday's price plus random noise.

$$\begin{cases} X_t = X_{t-1} + W_t & \text{for } t = 1, 2, \dots \\ X_0 = 0 & \text{for } t = 0 \end{cases}$$

where $W_t \overset{\text{iid}}{\sim} \mathtt{N}[0, \sigma_W^2]$

1. **Show** that $X_t = \sum_{j=1}^{t} W_j$ for $t = 1, 2, \dots$
2. Find the **mean** function at time t



Autocovariance Function

Definition (Autocovariance Function)

Let $X_0, X_1, X_2, \ldots$ be a **time series**. The **autocovariance function** of this time series at times s and t is denoted by $\gamma_X(s,t)$ and defined as

$$\gamma_X(s,t) := \mathsf{Cov}(X_s, X_t) = \mathbb{E}[(X_s - \mu_X(s))\,(X_t - \mu_X(t))]$$

- It can be easily **observed** that $\gamma_X(t,t) = \mathsf{Var}(X_t)$

Autocovariance Function

Definition (Autocovariance Function)

Let $X_0, X_1, X_2, \ldots$ be a **time series**. The **autocovariance function** of this time series at times s and t is denoted by $\gamma_X(s,t)$ and defined as

$$\gamma_X(s,t) := \mathrm{Cov}(X_s, X_t) = \mathbb{E}[(X_s - \mu_X(s))(X_t - \mu_X(t))]$$

- It can be easily **observed** that $\gamma_X(t,t) = \mathrm{Var}(X_t)$

Autocorrelation Function

Definition (Autocorrelation Function)

Let $X_0, X_1, X_2, \ldots$ be a **time series**. The **autocorrelation function** (ACF) of this time series at times s and t is denoted by $\rho_X(s,t)$ and defined as

$$\rho_X(s,t) := \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}}$$

- It is easy to **notice** that $\rho_X(t,t) = 1$, meaning each time point is **perfectly correlated** with itself.
- The **autocorrelation** function measures how the **same variable** relates to itself at different **time points**, unlike regular **correlation** between variables.
- If the **autocorrelation** value, $\rho_X(s,t)$, is close to one for some time points $s < t$, it means that the value at time t can be **predicted** from the value at time s using a **linear regression model** $X_t = \beta_0 + \beta_1 X_s + W_t$, where W_t a Gaussian white noise series.

Autocorrelation Function

Definition (Autocorrelation Function)

Let $X_0, X_1, X_2, \ldots$ be a **time series**. The **autocorrelation function** (ACF) of this time series at times s and t is denoted by $\rho_X(s,t)$ and defined as

$$\rho_X(s,t) := \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}}$$

- It is easy to **notice** that $\rho_X(t,t) = 1$, meaning each time point is **perfectly correlated** with itself.
- The **autocorrelation** function measures how the **same variable** relates to itself at different **time points**, unlike regular **correlation** between variables.
- If the **autocorrelation** value, $\rho_X(s,t)$, is close to one for some time points $s < t$, it means that the value at time t can be **predicted** from the value at time s using a **linear regression model** $X_t = \beta_0 + \beta_1 X_s + W_t$, where W_t a Gaussian white noise series.

Autocorrelation Function

Definition (Autocorrelation Function)

Let $X_0, X_1, X_2, \ldots$ be a **time series**. The **autocorrelation function** (ACF) of this time series at times s and t is denoted by $\rho_X(s,t)$ and defined as

$$\rho_X(s,t) := \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}}$$

- It is easy to **notice** that $\rho_X(t,t) = 1$, meaning each time point is **perfectly correlated** with itself.
- The **autocorrelation** function measures how the **same variable** relates to itself at different **time points**, unlike regular **correlation** between variables.
- If the **autocorrelation** value, $\rho_X(s,t)$, is close to one for some time points $s < t$, it means that the value at time t can be **predicted** from the value at time s using a **linear regression model** $X_t = \beta_0 + \beta_1 X_s + W_t$, where W_t a Gaussian white noise series.

Autocorrelation Function

Definition (Autocorrelation Function)

Let $X_0, X_1, X_2, \ldots$ be a **time series**. The **autocorrelation function** (ACF) of this time series at times s and t is denoted by $\rho_X(s,t)$ and defined as

$$\rho_X(s,t) := \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}}$$

- It is easy to **notice** that $\rho_X(t,t) = 1$, meaning each time point is **perfectly correlated** with itself.
- The **autocorrelation** function measures how the **same variable** relates to itself at different **time points**, unlike regular **correlation** between variables.
- If the **autocorrelation** value, $\rho_X(s,t)$, is close to one for some time points $s < t$, it means that the value at time t can be **predicted** from the value at time s using a **linear regression model** $X_t = \beta_0 + \beta_1 X_s + W_t$, where W_t a Gaussian white noise series.

Cross-correlation Function

Definition (Cross-correlation Function)

Let $X_0, X_1, X_2, \ldots$ and $Y_0, Y_1, Y_2, \ldots$ be two **time series**. The **cross-correlation function** (CCF) of these two time series at times s and t is denoted by $\rho_{X,Y}(s,t)$ and defined as

$$\rho_{X,Y}(s,t) := \frac{\gamma_{X,Y}(s,t)}{\sqrt{\gamma_X(s,s)\gamma_Y(t,t)}}$$

where $\gamma_{X,Y}(s,t) := \mathrm{Cov}(X_s, Y_t)$ is the **cross-covariance function** between the two time series at times s and t.

- If the **magnitude** of $\rho_{X,Y}(s,t)$ is nearly one for some time points $s < t$, then Y_t can be **predicted** from X_s through a **linear regression model** $Y_t = \beta_0 + \beta_1 X_s + W_t$, where W_t is a Gaussian white noise series.

Cross-correlation Function

Definition (Cross-correlation Function)

Let $X_0, X_1, X_2, \ldots$ and $Y_0, Y_1, Y_2, \ldots$ be two **time series**. The **cross-correlation function** (CCF) of these two time series at times s and t is denoted by $\rho_{X,Y}(s,t)$ and defined as

$$\rho_{X,Y}(s,t) := \frac{\gamma_{X,Y}(s,t)}{\sqrt{\gamma_X(s,s)\gamma_Y(t,t)}}$$

where $\gamma_{X,Y}(s,t) := \mathrm{Cov}(X_s, Y_t)$ is the **cross-covariance function** between the two time series at times s and t.

- If the **magnitude** of $\rho_{X,Y}(s,t)$ is nearly one for some time points $s < t$, then Y_t can be **predicted** from X_s through a **linear regression model** $Y_t = \beta_0 + \beta_1 X_s + W_t$, where W_t is a Gaussian white noise series.

Outline

Stationary Time Series

Definition (Stationary Time Series)

A time series $X_0, X_1, X_2, \ldots$ is called **stationary** if:

1. Its **mean** value, $\mu_X(t)$, is **constant** and does **not change** over time for any time t.
2. Its **autocovariance** function, $\gamma_X(s,t)$, depends only on the **difference** between the time points $|s-t|$ (called the **lag**), **not** on the specific times s or t themselves.

- In some textbooks, including your **textbook**, this type of stationarity is also called **weakly stationary**.

Stationary Time Series

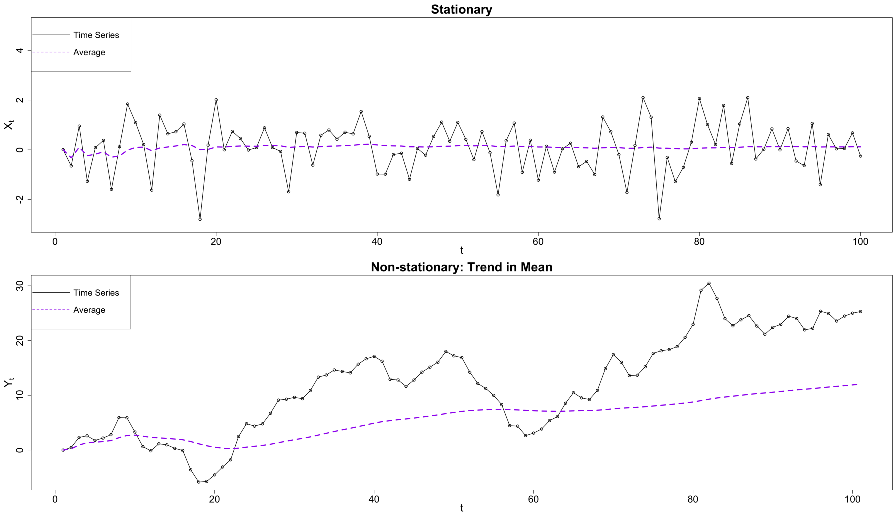
Definition (Stationary Time Series)

A time series $X_0, X_1, X_2, \ldots$ is called **stationary** if:

1. Its **mean** value, $\mu_X(t)$, is **constant** and does **not change** over time for any time t.
2. Its **autocovariance** function, $\gamma_X(s, t)$, depends only on the **difference** between the time points $|s - t|$ (called the **lag**), **not** on the specific times s or t themselves.

- In some textbooks, including your **textbook**, this type of stationarity is also called **weakly stationary**.

Stationary vs. Non-stationary Time Series



ACF of a Stationary Time Series

Definition (Autocovariance, Variance, and Autocorrelation)

Let $X_0, X_1, X_2, \ldots$ be a **stationary** time series. Then:

- The **autocovariance function** at lag $h = 0, 1, 2, \ldots$ is defined as

$$\gamma_X(h) := \mathrm{Cov}(X_{t+h}, X_t) \quad \text{for any time } t$$

- The **variance** of the time series equals

$$\mathrm{Var}(X_t) = \gamma_X(0)$$

- The **autocorrelation function** (ACF) at lag $h = 0, 1, 2, \ldots$ is defined as

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)}$$

  and $\rho_X(h) = \rho_X(-h)$ for $h = -1, -2, \ldots$

ACF of a Stationary Time Series

Definition (Autocovariance, Variance, and Autocorrelation)

Let $X_0, X_1, X_2, \ldots$ be a **stationary** time series. Then:

- The **autocovariance function** at lag $h = 0, 1, 2, \ldots$ is defined as

$$\gamma_X(h) := \mathrm{Cov}(X_{t+h}, X_t) \quad \text{for any time } t$$

- The **variance** of the time series equals

$$\mathrm{Var}(X_t) = \gamma_X(0)$$

- The **autocorrelation function** (ACF) at lag $h = 0, 1, 2, \ldots$ is defined as

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)}$$

and $\rho_X(h) = \rho_X(-h)$ for $h = -1, -2, \ldots$

ACF of a Stationary Time Series

Definition (Autocovariance, Variance, and Autocorrelation)

Let $X_0, X_1, X_2, \ldots$ be a **stationary** time series. Then:

- The **autocovariance function** at lag $h = 0, 1, 2, \ldots$ is defined as

$$\gamma_X(h) := \text{Cov}(X_{t+h}, X_t) \quad \text{for any time } t$$

- The **variance** of the time series equals

$$\text{Var}(X_t) = \gamma_X(0)$$

- The **autocorrelation function** (ACF) at lag $h = 0, 1, 2, \ldots$ is defined as

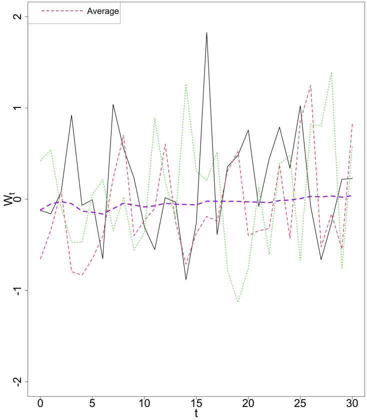
$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)}$$

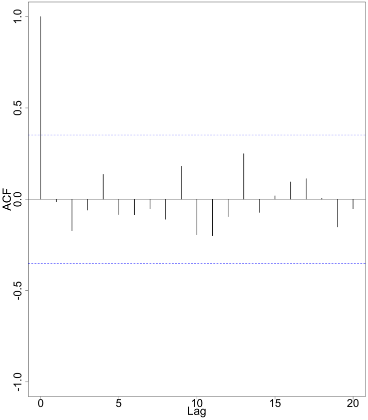
  and $\rho_X(h) = \rho_X(-h)$ for $h = -1, -2, \ldots$

Example

Example (Gaussian White Noise Series)

Is the **Gaussian white noise** time series $W_0, W_1, W_2, \dots$ **stationary**?



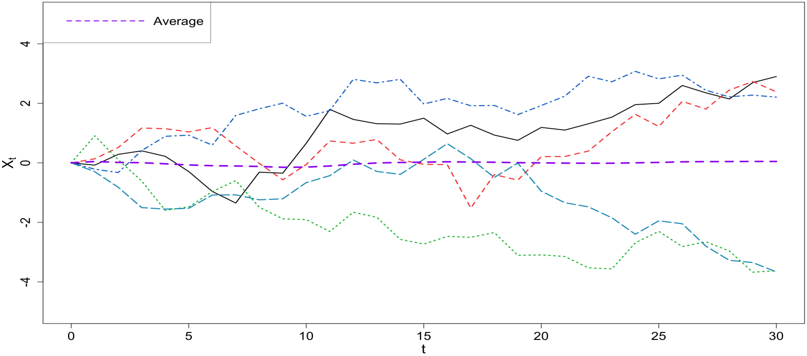


Example

Example (Simple Random Walk)

Is the simple **random walk** model, where $W_t \overset{\text{iid}}{\sim} \mathtt{N}[0, \sigma_W^2]$, **stationary**?

$$\begin{cases} X_t = X_{t-1} + W_t & \text{for } t = 1, 2, \ldots \\ X_0 = 0 & \text{for } t = 0 \end{cases}$$



Bonus Assignment

Question

Is the following time series model **stationary**? Justify your answer.

$$\begin{cases} X_t = X_{t-1} + W_t & \text{for } t = 1, 2, \dots \\ X_0 = W_0 & \text{for } t = 0 \end{cases}$$

where $W_0, W_1, W_2, \dots$ is a **Gaussian white noise** series with mean zero and variance one.

Homoskedastic vs. Heteroskedastic Time Series

Definition (Homoskedasticity and Heteroskedasticity)

A time series is **homoskedastic** if its variance is **consistent** over time, and **heteroskedastic** if its variance **changes** over time.

- Any **stationary** time series is homoskedastic, while a heteroskedastic time series is generally **non-stationary**.
- Knowing if a time series is heteroskedastic or homoskedastic is **essential** for selecting the appropriate **model** and analysis **method**.
- Heteroskedasticity is common in **financial** data, where price fluctuations often lead to **changing** levels of volatility.

Homoskedastic vs. Heteroskedastic Time Series

Definition (Homoskedasticity and Heteroskedasticity)

A time series is **homoskedastic** if its variance is **consistent** over time, and **heteroskedastic** if its variance **changes** over time.

- Any **stationary** time series is homoskedastic, while a heteroskedastic time series is generally **non-stationary**.
- Knowing if a time series is heteroskedastic or homoskedastic is **essential** for selecting the appropriate **model** and analysis **method**.
- Heteroskedasticity is common in **financial** data, where price fluctuations often lead to **changing** levels of volatility.

Homoskedastic vs. Heteroskedastic Time Series

Definition (Homoskedasticity and Heteroskedasticity)

A time series is **homoskedastic** if its variance is **consistent** over time, and **heteroskedastic** if its variance **changes** over time.

- Any **stationary** time series is homoskedastic, while a heteroskedastic time series is generally **non-stationary**.
- Knowing if a time series is heteroskedastic or homoskedastic is **essential** for selecting the appropriate **model** and analysis **method**.
- Heteroskedasticity is common in **financial** data, where price fluctuations often lead to **changing** levels of volatility.

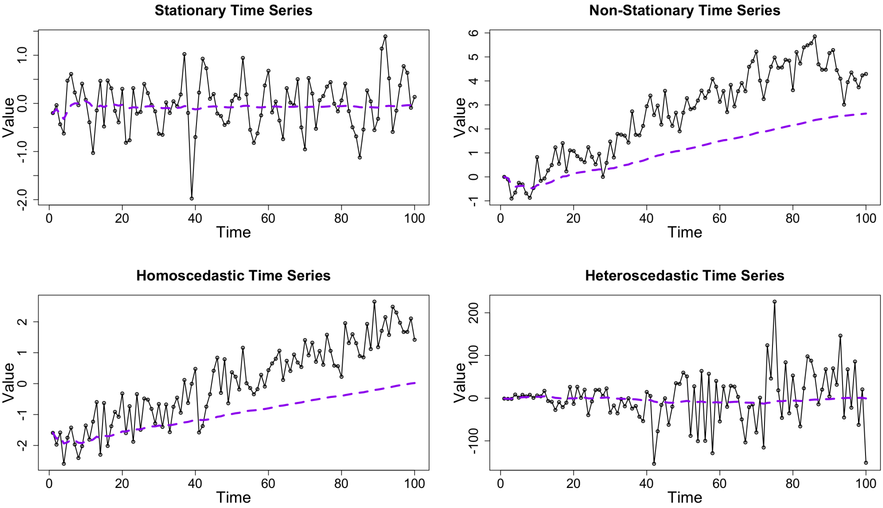
Homoskedastic vs. Heteroskedastic Time Series

Definition (Homoskedasticity and Heteroskedasticity)

A time series is **homoskedastic** if its variance is **consistent** over time, and **heteroskedastic** if its variance **changes** over time.

- Any **stationary** time series is homoskedastic, while a heteroskedastic time series is generally **non-stationary**.
- Knowing if a time series is heteroskedastic or homoskedastic is **essential** for selecting the appropriate **model** and analysis **method**.
- Heteroskedasticity is common in **financial** data, where price fluctuations often lead to **changing** levels of volatility.

Trends in Mean and Variance



Statistical Hypothesis Test

- **Plotting** the time series data offers **insightful** information about the data's **nature**.
- However, making **inferences** should rely on conducting an appropriate **statistical test**.
- **KPSS** (Kwiatkowski-Phillips-Schmidt-Shin) test for checking **stationarity**:

$$\begin{cases} H_0 : \text{The time series is } \textbf{stationary} \\ H_a : \text{The time series is } \textbf{not} \text{ stationary} \end{cases}$$

Statistical Hypothesis Test

- **Plotting** the time series data offers **insightful** information about the data's **nature**.
- However, making **inferences** should rely on conducting an appropriate **statistical test**.
- **KPSS** (Kwiatkowski-Phillips-Schmidt-Shin) test for checking **stationarity**:

$$\begin{cases} H_0 : \text{The time series is } \textbf{stationary} \\ H_a : \text{The time series is } \textbf{not} \text{ stationary} \end{cases}$$

Statistical Hypothesis Test

- **Plotting** the time series data offers **insightful** information about the data's **nature**.
- However, making **inferences** should rely on conducting an appropriate **statistical test**.
- **KPSS** (Kwiatkowski-Phillips-Schmidt-Shin) test for checking **stationarity**:

$$\begin{cases} H_0 : \text{The time series is } \textbf{stationary} \\ H_a : \text{The time series is } \textbf{not} \text{ stationary} \end{cases}$$

Jointly Stationary Time Series

Definition (Jointly Stationary)

Two time series, like $X_0, X_1, X_2, \ldots$ and $Y_0, Y_1, Y_2, \ldots$, are called **jointly stationary** if:

1. Each time series is **stationary** on its own, meaning their patterns do not change over time.
2. The **relationship** between the two time series (measured by **covariance**) depends only on the **difference** between them, not on when we look at it.

Definition (Cross-correlation Function)

The **cross-correlation function** (CCF) of **jointly stationary** time series X_t and Y_t at lag $h = 0, 1, 2, \ldots$ is defined as

$$\rho_{X,Y}(h) := \frac{\text{Cov}(X_{t+h}, Y_t)}{\sqrt{\text{Var}(X_{t+h})\text{Var}(Y_t)}} = \frac{\gamma_{X,Y}(h)}{\sqrt{\gamma_X(0)\gamma_Y(0)}}$$

and $\rho_{X,Y}(h) = \rho_{Y,X}(-h)$ for $h = -1, -2, \ldots$

Jointly Stationary Time Series

Definition (Jointly Stationary)

Two time series, like $X_0, X_1, X_2, \ldots$ and $Y_0, Y_1, Y_2, \ldots$, are called **jointly stationary** if:

1. Each time series is **stationary** on its own, meaning their patterns do not change over time.
2. The **relationship** between the two time series (measured by **covariance**) depends only on the **difference** between them, not on when we look at it.

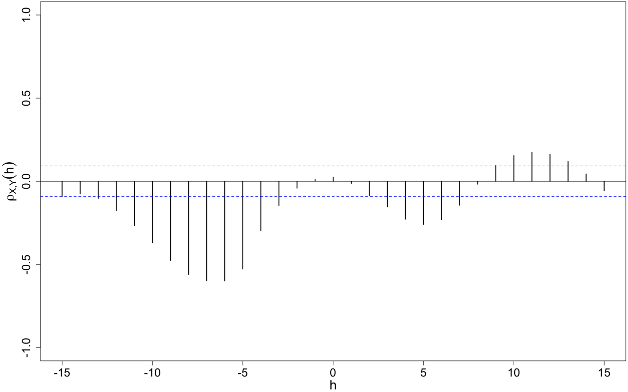
Definition (Cross-correlation Function)

The **cross-correlation function** (CCF) of **jointly stationary** time series X_t and Y_t at lag $h = 0, 1, 2, \ldots$ is defined as

$$\rho_{X,Y}(h) := \frac{\text{Cov}(X_{t+h}, Y_t)}{\sqrt{\text{Var}(X_{t+h})\text{Var}(Y_t)}} = \frac{\gamma_{X,Y}(h)}{\sqrt{\gamma_X(0)\gamma_Y(0)}}$$

and $\rho_{X,Y}(h) = \rho_{Y,X}(-h)$ for $h = -1, -2, \ldots$

Case Study: SOI (X_t) vs. Recruitment (Y_t) Cross-correlation Analysis



References

1. R.H. Shumway and D.S. Stoffer, *Time Series Analysis and Its Applications With R Examples*, Springer, New York, 2017.