Selected Topics in Mathematics of Learning

High-Dimensional Statistics

Lecturer: Marius Yamakou

Winter Semester 2024/25
Department of Data Science, FAU

January 21, 2025

Part VI

Sparse vector autoregressive models

Motivation

- Traditional assumptions:
  - Data are multivariate Gaussian
  - Observations are independent and identically distributed (i.i.d.)

Part VI

Sparse vector autoregressive models

Motivation

- Traditional assumptions:
  - Data are multivariate Gaussian
  - Observations are independent and identically distributed (i.i.d.)
- Real-world data often deviate from these assumptions:
  - Examples: fMRI data, ETFs
  - Temporal correlations are common

Part VI

Sparse vector autoregressive models

Motivation

- Traditional assumptions:
  - Data are multivariate Gaussian
  - Observations are independent and identically distributed (i.i.d.)
- Real-world data often deviate from these assumptions:
  - Examples: fMRI data, ETFs
  - Temporal correlations are common

Key Idea: Vector Autoregressive Models (VARs) are powerful tools for capturing temporal correlations in high-dimensional data.

Outline

- Preliminaries of VAR
- Estimation of Sparse VAR using LASSO
- LASSO: Estimation
- LASSO: Properties

1. Preliminaries of VAR

A time series $\{X_t\}_{t\in\mathbb{Z}} = \{(X_{j,t})_{j=1,\ldots,d}\}_{t\in\mathbb{Z}}$ follows a **VAR(p)** model if:

$$X_t = \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} + \epsilon_t, \quad t \in \mathbb{Z},$$

1. Preliminaries of VAR

A time series $\{X_t\}_{t\in\mathbb{Z}} = \{(X_{j,t})_{j=1,\ldots,d}\}_{t\in\mathbb{Z}}$ follows a **VAR(p)** model if:

$$X_t = \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} + \epsilon_t, \quad t \in \mathbb{Z},$$

where:

- $\Phi_1, \ldots, \Phi_p$ are $d \times d$ matrices.
- $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a white noise series with the following statistics:

$$\mathbb{E}[\epsilon_t] = 0, \quad \mathbb{E}[\epsilon_t \epsilon_t'] = \Sigma_\epsilon, \quad \mathbb{E}[\epsilon_s \epsilon_t'] = 0 \quad \text{for } s \neq t.$$

1. Preliminaries of VAR

A time series $\{X_t\}_{t\in\mathbb{Z}} = \{(X_{j,t})_{j=1,\ldots,d}\}_{t\in\mathbb{Z}}$ follows a **VAR(p)** model if:

$$X_t = \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} + \epsilon_t, \quad t \in \mathbb{Z},$$

where:

- $\Phi_1, \ldots, \Phi_p$ are $d \times d$ matrices.
- $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a white noise series with the following statistics:

$$\mathbb{E}[\epsilon_t] = 0, \quad \mathbb{E}[\epsilon_t \epsilon_t'] = \Sigma_\epsilon, \quad \mathbb{E}[\epsilon_s \epsilon_t'] = 0 \quad \text{for } s \neq t.$$

Example: VAR(1) with $d = 2$

$$\Phi_1 = (\Phi_{1,jk})_{j,k=1,2} = \begin{bmatrix} \Phi_{1,11} & \Phi_{1,12} \\ \Phi_{1,21} & \Phi_{1,22} \end{bmatrix}$$

1. Preliminaries of VAR

A time series $\{X_t\}_{t\in\mathbb{Z}} = \{(X_{j,t})_{j=1,\dots,d}\}_{t\in\mathbb{Z}}$ follows a **VAR(p)** model if:

$$X_t = \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} + \epsilon_t, \quad t \in \mathbb{Z},$$

where:

- $\Phi_1, \ldots, \Phi_p$ are $d \times d$ matrices.
- $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a white noise series with the following statistics:

$$\mathbb{E}[\epsilon_t] = 0, \quad \mathbb{E}[\epsilon_t \epsilon_t'] = \Sigma_\epsilon, \quad \mathbb{E}[\epsilon_s \epsilon_t'] = 0 \quad \text{for } s \neq t.$$

Example: VAR(1) with $d = 2$

$$\Phi_1 = (\Phi_{1,jk})_{j,k=1,2} = \begin{bmatrix} \Phi_{1,11} & \Phi_{1,12} \\ \Phi_{1,21} & \Phi_{1,22} \end{bmatrix}$$

The equations are:

$$X_{1,t} = \Phi_{1,11} X_{1,t-1} + \Phi_{1,12} X_{2,t-1} + \epsilon_{1,t},$$
$$X_{2,t} = \Phi_{1,21} X_{1,t-1} + \Phi_{1,22} X_{2,t-1} + \epsilon_{2,t}.$$

1. Preliminaries of VAR

Please note the following:

- $\mathbb{E}[\epsilon_t] = 0$: This indicates that the white noise has a mean (or expected value) of 0 for all time points t. Essentially, the noise fluctuates around zero on average.
- $\mathbb{E}[\epsilon_t \epsilon_t'] = \Sigma_\epsilon$: Here, Σ_ϵ is the covariance matrix of ϵ_t, which characterizes the variance (for scalar ϵ_t) or the relationships between components (for vector-valued ϵ_t). For scalar white noise, this reduces to $\mathbb{E}[\epsilon_t^2] = \sigma_\epsilon^2$, where σ_ϵ^2 is the variance.
- $\mathbb{E}[\epsilon_s \epsilon_t'] = 0$ for $s \neq t$: This indicates that ϵ_t values at different time points t are uncorrelated. That is, the noise at time t does not depend on or influence the noise at any other time s. For vector-valued ϵ_t, this implies that the cross-covariance between vectors at different times is 0.

1. Preliminaries of VAR

- If the **VAR(1)** model has $d = 1$ (i.e., a univariate time series) and $p = 1$, the equation simplifies significantly because there is only one time series and only one lag, making it one of the simplest forms of a VAR model.

1. Preliminaries of VAR

- If the **VAR(1)** model has $d = 1$ (i.e., a univariate time series) and $p = 1$, the equation simplifies significantly because there is only one time series and only one lag, making it one of the simplest forms of a VAR model.
- The univariate time series $\{X_t\}_{t\in\mathbb{Z}}$, a **VAR(1)** model with $d = 1$ and $p = 1$ simplifies to:

$$X_t = \Phi_1 X_{t-1} + \epsilon_t,$$

1. Preliminaries of VAR

- If the **VAR(1)** model has $d = 1$ (i.e., a univariate time series) and $p = 1$, the equation simplifies significantly because there is only one time series and only one lag, making it one of the simplest forms of a VAR model.
- The univariate time series $\{X_t\}_{t\in\mathbb{Z}}$, a **VAR(1)** model with $d = 1$ and $p = 1$ simplifies to:

$$X_t = \Phi_1 X_{t-1} + \epsilon_t,$$

where:

  - Φ_1 is a scalar (instead of a matrix since $d = 1$).
  - ϵ_t is white noise:

$$\mathbb{E}[\epsilon_t] = 0, \quad \mathbb{E}[\epsilon_t^2] = \sigma_\epsilon^2, \quad \mathbb{E}[\epsilon_s \epsilon_t] = 0 \text{ for } s \neq t.$$

1.1 Stability

- A VAR(1) model is:
  - **Strictly stable** if:

$$\det(I_d - \Phi_1 z) \neq 0, \quad \text{for } |z| < 1,$$

    which is equivalent to requiring that all eigenvalues of Φ_1 satisfy $|\lambda| < 1$.

1.1 Stability

- A VAR(1) model is:
  - **Strictly stable** if:

$$\det(I_d - \Phi_1 z) \neq 0, \quad \text{for } |z| < 1,$$

    which is equivalent to requiring that all eigenvalues of Φ_1 satisfy $|\lambda| < 1$.
  - **Marginally stable** if:

$$\det(I_d - \Phi_1 z) = 0 \quad \text{for } |z| = 1,$$

    but the system does not diverge over time. This typically occurs when some eigenvalues lie exactly on the unit circle (e.g., $|\lambda| = 1$).

1.1 Stability

- A VAR(1) model is:
  - **Strictly stable** if:

$$\det(I_d - \Phi_1 z) \neq 0, \quad \text{for } |z| < 1,$$

    which is equivalent to requiring that all eigenvalues of Φ_1 satisfy $|\lambda| < 1$.
  - **Marginally stable** if:

$$\det(I_d - \Phi_1 z) = 0 \quad \text{for } |z| = 1,$$

    but the system does not diverge over time. This typically occurs when some eigenvalues lie exactly on the unit circle (e.g., $|\lambda| = 1$).
- For $d = 1$, the model $X_t = \Phi_1 X_{t-1} + \epsilon_t$, is:
  - **Strictly stable** if $|\Phi_1| < 1$,

1.1 Stability

- A VAR(1) model is:
  - **Strictly stable** if:

$$\det(I_d - \Phi_1 z) \neq 0, \quad \text{for } |z| < 1,$$

    which is equivalent to requiring that all eigenvalues of Φ_1 satisfy $|\lambda| < 1$.
  - **Marginally stable** if:

$$\det(I_d - \Phi_1 z) = 0 \quad \text{for } |z| = 1,$$

    but the system does not diverge over time. This typically occurs when some eigenvalues lie exactly on the unit circle (e.g., $|\lambda| = 1$).
- For $d = 1$, the model $X_t = \Phi_1 X_{t-1} + \epsilon_t$, is:
  - **Strictly stable** if $|\Phi_1| < 1$,
  - **Marginally stable** if $|\Phi_1| = 1$,

1.1 Stability

- A VAR(1) model is:
  - **Strictly stable** if:
$$\det(I_d - \Phi_1 z) \neq 0, \quad \text{for } |z| < 1,$$
    which is equivalent to requiring that all eigenvalues of Φ_1 satisfy $|\lambda| < 1$.
  - **Marginally stable** if:
$$\det(I_d - \Phi_1 z) = 0 \quad \text{for } |z| = 1,$$
    but the system does not diverge over time. This typically occurs when some eigenvalues lie exactly on the unit circle (e.g., $|\lambda| = 1$).
- For $d = 1$, the model $X_t = \Phi_1 X_{t-1} + \epsilon_t$, is:
  - **Strictly stable** if $|\Phi_1| < 1$,
  - **Marginally stable** if $|\Phi_1| = 1$,
  - **Unstable** if $|\Phi_1| > 1$.

1.1 Stability

General Stability Condition: For a VAR(p) model, stability requires that its characteristic polynomial satisfies:

$$\det(\Phi(z)) \neq 0, \quad \text{for } |z| \leq 1,\ z \in \mathbb{C},$$

1.1 Stability

General Stability Condition: For a VAR(p) model, stability requires that its characteristic polynomial satisfies:

$$\det(\Phi(z)) \neq 0, \quad \text{for } |z| \leq 1,\ z \in \mathbb{C},$$

where:

$$\Phi(z) = I_d - \Phi_1 z - \Phi_2 z^2 - \ldots - \Phi_p z^p.$$

1.1 Stability

General Stability Condition: For a VAR(p) model, stability requires that its characteristic polynomial satisfies:

$$\det(\Phi(z)) \neq 0, \quad \text{for } |z| \leq 1, \ z \in \mathbb{C},$$

where:

$$\Phi(z) = I_d - \Phi_1 z - \Phi_2 z^2 - \ldots - \Phi_p z^p.$$

Assumption: In this part of the course, we assume all VAR models are strictly stable unless otherwise stated.

1.2 Stationarity

A VAR model is (second-order) **Stationarity** if:

1. The mean of the process is constant:

$$\mathbb{E}(X_t) = 0 \quad \text{(or constant over time).}$$

1.2 Stationarity

A VAR model is (second-order) **Stationarity** if:

1. The mean of the process is constant:

$$\mathbb{E}(X_t) = 0 \quad \text{(or constant over time).}$$

2. The **autocovariance function (ACVF)** depends only on the lag ℓ and not on t, that is:

$$\Gamma_X(\ell) = \mathbb{E}(X_{t+\ell}X_t') = \mathbb{E}(X_\ell X_0').$$

1.2 Stationarity

A VAR model is (second-order) **Stationarity** if:

1. The mean of the process is constant:

$$\mathbb{E}(X_t) = 0 \quad \text{(or constant over time).}$$

2. The **autocovariance function (ACVF)** depends only on the lag ℓ and not on t, that is:

$$\Gamma_X(\ell) = \mathbb{E}(X_{t+\ell}X_t') = \mathbb{E}(X_\ell X_0').$$

Remark: When $\ell = 0$, we get the **covariance**: $\Gamma_X(0) = \Sigma_X = \mathbb{E}(X_t X_t')$

1.2 Stationarity

A VAR model is (second-order) **Stationarity** if:

1. The mean of the process is constant:

$$\mathbb{E}(X_t) = 0 \quad \text{(or constant over time).}$$

2. The **autocovariance function (ACVF)** depends only on the lag ℓ and not on t, that is:

$$\Gamma_X(\ell) = \mathbb{E}(X_{t+\ell}X_t') = \mathbb{E}(X_\ell X_0').$$

Remark: When $\ell = 0$, we get the **covariance**: $\Gamma_X(0) = \Sigma_X = \mathbb{E}(X_t X_t')$

Implications of stationarity:

- The statistical properties of the time series do not change over time.

1.2 Stationarity

A VAR model is (second-order) **Stationarity** if:

1. The mean of the process is constant:

$$\mathbb{E}(X_t) = 0 \quad \text{(or constant over time)}.$$

2. The **autocovariance function (ACVF)** depends only on the lag ℓ and not on t, that is:

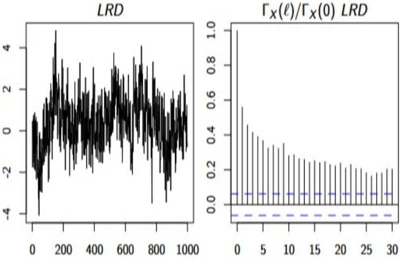
$$\Gamma_X(\ell) = \mathbb{E}(X_{t+\ell} X_t') = \mathbb{E}(X_\ell X_0').$$

Remark: When $\ell = 0$, we get the **covariance**: $\Gamma_X(0) = \Sigma_X = \mathbb{E}(X_t X_t')$

Implications of stationarity:

- The statistical properties of the time series do not change over time.
- Stationarity ensures that the VAR model's behavior is predictable and consistent over time, making it suitable for modeling and forecasting.

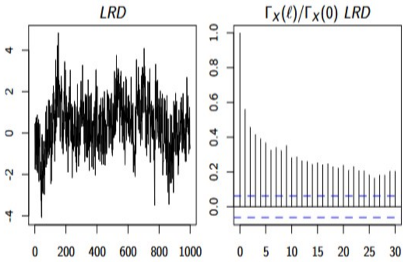
1.2 Stationarity: Long-range dependence (LRD)



Left: Time Series:

- The time series shows a pattern with persistent, slowly decaying behavior over time, indicating that observations are strongly correlated even at long time lags.

1.2 Stationarity: Long-range dependence (LRD)



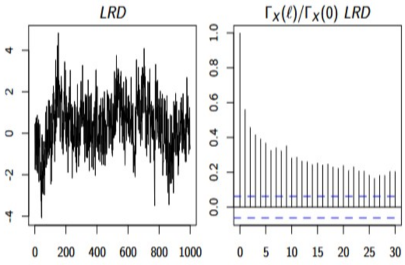
Right: Autocorrelation Function (ACF):

- The autocorrelation function (ACF) $\Gamma_X(\ell)/\Gamma_X(0)$ decreases very slowly as the lag ℓ increases.

Left: Time Series:

- The time series shows a pattern with persistent, slowly decaying behavior over time, indicating that observations are strongly correlated even at long time lags.

1.2 Stationarity: Long-range dependence (LRD)



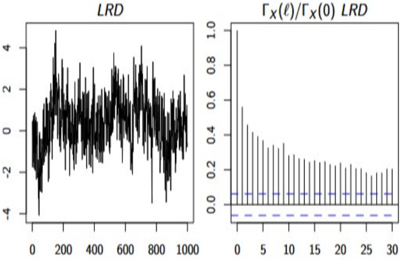
Right: Autocorrelation Function (ACF):

- The autocorrelation function (ACF) $\Gamma_X(\ell)/\Gamma_X(0)$ decreases very slowly as the lag ℓ increases.
- This behavior reflects **long-range dependence (LRD)**, where correlations persist over long time periods.

Left: Time Series:

- The time series shows a pattern with persistent, slowly decaying behavior over time, indicating that observations are strongly correlated even at long time lags.

1.2 Stationarity: Long-range dependence (LRD)



Left: Time Series:

- The time series shows a pattern with persistent, slowly decaying behavior over time, indicating that observations are strongly correlated even at long time lags.

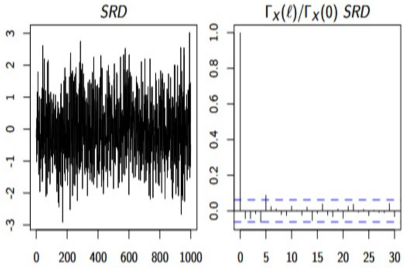
Right: Autocorrelation Function (ACF):

- The autocorrelation function (ACF) $\Gamma_X(\ell)/\Gamma_X(0)$ decreases very slowly as the lag ℓ increases.
- This behavior reflects **long-range dependence (LRD)**, where correlations persist over long time periods.

Stationarity Implications:

- The slow decay in the ACF suggests that the time series is **not stationary**, as stationarity requires the ACF to decay rapidly (e.g., exponentially or geometrically–decreases by a constant ratio in each time step.

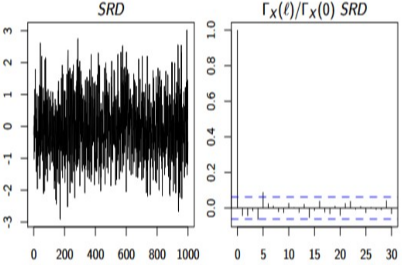
1.2 Stationarity: Short-range dependence (SRD)



Left: Time Series:

- The time series exhibits more erratic behavior compared to LRD. While there are small dependencies between consecutive observations, they dissipate quickly over time.

1.2 Stationarity: Short-range dependence (SRD)



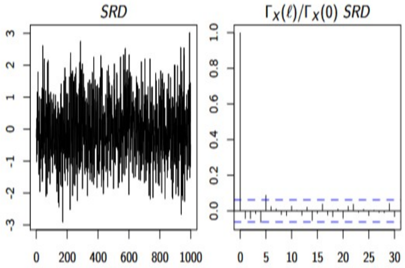
Right: Autocorrelation Function (ACF):

- The ACF decreases rapidly as ℓ increases and eventually becomes close to zero, indicating weak or no correlation at large lags. This is characteristic of **short-range dependence (SRD)**.

Left: Time Series:

- The time series exhibits more erratic behavior compared to LRD. While there are small dependencies between consecutive observations, they dissipate quickly over time.

1.2 Stationarity: Short-range dependence (SRD)



Left: Time Series:

- The time series exhibits more erratic behavior compared to LRD. While there are small dependencies between consecutive observations, they dissipate quickly over time.

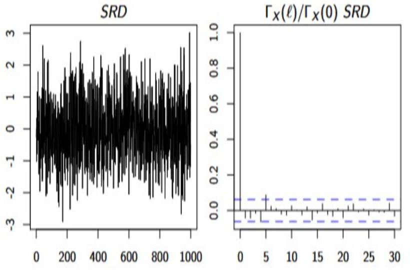
Right: Autocorrelation Function (ACF):

- The ACF decreases rapidly as ℓ increases and eventually becomes close to zero, indicating weak or no correlation at large lags. This is characteristic of **short-range dependence (SRD)**.
- The rapid decay indicates that the time series is likely stationary, as the ACF does not persist indefinitely.

Stationarity Implications:

- The rapid decay in the ACF suggests that the time series is likely **stationary**.

1.2 Stationarity: Short-range dependence (SRD)



Left: Time Series:

- The time series exhibits more erratic behavior compared to LRD. While there are small dependencies between consecutive observations, they dissipate quickly over time.

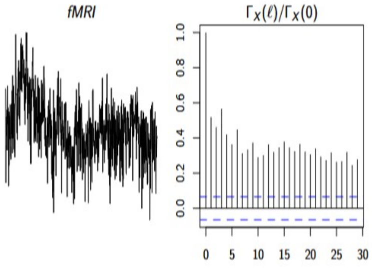
Right: Autocorrelation Function (ACF):

- The ACF decreases rapidly as ℓ increases and eventually becomes close to zero, indicating weak or no correlation at large lags. This is characteristic of **short-range dependence (SRD)**.
- The rapid decay indicates that the time series is likely stationary, as the ACF does not persist indefinitely.

Stationarity Implications:

- The rapid decay in the ACF suggests that the time series is likely **stationary**.
- Stationary time series exhibit consistent statistical properties over time.

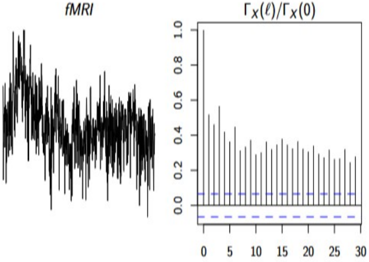
1.2 Stationarity: In Functional Magnetic Resonance Imaging (fMRI).



Left: Time Series:

- The fMRI signal shows fluctuations over time, typical of neuro-physiological processes, with moderate temporal correlations.

1.2 Stationarity: In Functional Magnetic Resonance Imaging (fMRI).



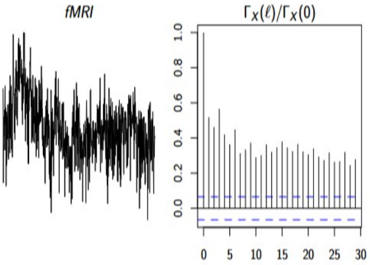
Right: Autocorrelation Function (ACF):

- The ACF decreases gradually but not as rapidly as in SRD. The slower decay suggests some degree of correlation over time, though it is weaker than in LRD.

Left: Time Series:

- The fMRI signal shows fluctuations over time, typical of neuro-physiological processes, with moderate temporal correlations.

1.2 Stationarity: In Functional Magnetic Resonance Imaging (fMRI).



Left: Time Series:

- The fMRI signal shows fluctuations over time, typical of neuro-physiological processes, with moderate temporal correlations.

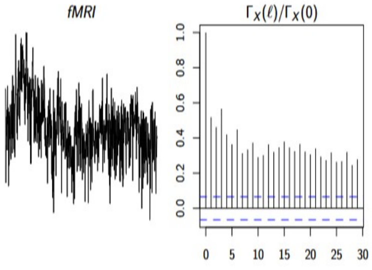
Right: Autocorrelation Function (ACF):

- The ACF decreases gradually but not as rapidly as in SRD. The slower decay suggests some degree of correlation over time, though it is weaker than in LRD.

Stationarity Implications:

- The fMRI data may exhibit **weak stationarity**, as the mean and variance appear constant over time.

1.2 Stationarity: In Functional Magnetic Resonance Imaging (fMRI).



Left: Time Series:

- The fMRI signal shows fluctuations over time, typical of neuro-physiological processes, with moderate temporal correlations.

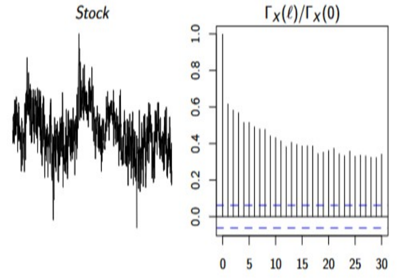
Right: Autocorrelation Function (ACF):

- The ACF decreases gradually but not as rapidly as in SRD. The slower decay suggests some degree of correlation over time, though it is weaker than in LRD.

Stationarity Implications:

- The fMRI data may exhibit **weak stationarity**, as the mean and variance appear constant over time.
- However, correlations persist over moderate lags, requiring further investigation to confirm stationarity.

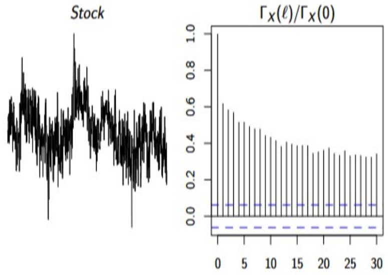
1.2 Stationarity: In Stock Prices



Left: Time Series:

- The stock price time series appears highly volatile, with no clear trend or periodicity.

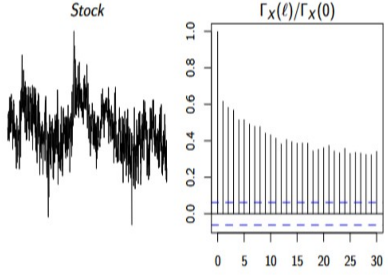
1.2 Stationarity: In Stock Prices



Left: Time Series:

- The stock price time series appears highly volatile, with no clear trend or periodicity.
- The variations are large and erratic, consistent with typical financial time series.

1.2 Stationarity: In Stock Prices



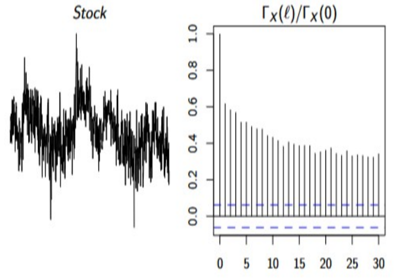
Right: Autocorrelation Function (ACF):

- The autocorrelation function (ACF) $\Gamma_X(\ell)/\Gamma_X(0)$ does not decay as rapidly as expected for short-range dependence.

Left: Time Series:

- The stock price time series appears highly volatile, with no clear trend or periodicity.
- The variations are large and erratic, consistent with typical financial time series.

1.2 Stationarity: In Stock Prices



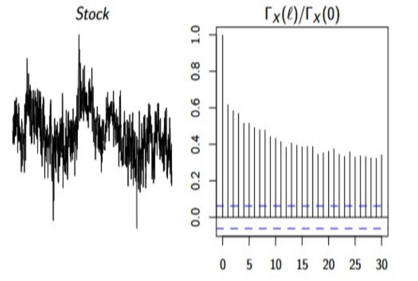
Right: Autocorrelation Function (ACF):

- The autocorrelation function (ACF) $\Gamma_X(\ell)/\Gamma_X(0)$ does not decay as rapidly as expected for short-range dependence.
- This behavior suggests that correlations persist over time, indicating potential long-range dependence, weak or even non-stationarity.

Left: Time Series:

- The stock price time series appears highly volatile, with no clear trend or periodicity.
- The variations are large and erratic, consistent with typical financial time series.

1.2 Stationarity: In Stock Prices



Left: Time Series:

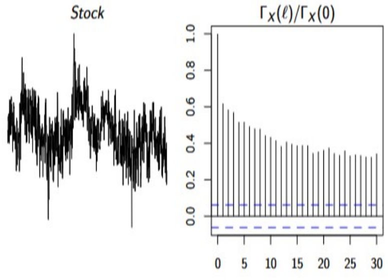
- The stock price time series appears highly volatile, with no clear trend or periodicity.
- The variations are large and erratic, consistent with typical financial time series.

Right: Autocorrelation Function (ACF):

- The autocorrelation function (ACF) $\Gamma_X(\ell)/\Gamma_X(0)$ does not decay as rapidly as expected for short-range dependence.
- This behavior suggests that correlations persist over time, indicating potential long-range dependence, weak or even non-stationarity.

Stationarity Implications:

- The slow decay in the ACF indicates that the time series may be **weakly stationary**.



Left: Time Series:

- The stock price time series appears highly volatile, with no clear trend or periodicity.
- The variations are large and erratic, consistent with typical financial time series.

Right: Autocorrelation Function (ACF):

- The autocorrelation function (ACF) $\Gamma_X(\ell)/\Gamma_X(0)$ does not decay as rapidly as expected for short-range dependence.
- This behavior suggests that correlations persist over time, indicating potential long-range dependence, weak or even non-stationarity.

Stationarity Implications:

- The slow decay in the ACF indicates that the time series may be **weakly stationary**.
- Financial data can exhibit behaviors such as long-range dependence or structural breaks that violate stationarity assumptions.

1.2. Key Points on Stationarity and ACF

1. Stationarity:

- A time series is stationary if its mean and variance are constant over time, and the autocovariance $\Gamma_X(\ell)$ depends only on the lag ℓ, not on time t.

1.2. Key Points on Stationarity and ACF

1. Stationarity:

- A time series is stationary if its mean and variance are constant over time, and the autocovariance $\Gamma_X(\ell)$ depends only on the lag ℓ, not on time t.
- In the figures:
  - SRD, fMRI, and stock data are examples of stationary (or weakly stationary) time series.

1.2. Key Points on Stationarity and ACF

1. Stationarity:

- A time series is stationary if its mean and variance are constant over time, and the autocovariance $\Gamma_X(\ell)$ depends only on the lag ℓ, not on time t.
- In the figures:
  - SRD, fMRI, and stock data are examples of stationary (or weakly stationary) time series.
  - LRD may not be stationary due to the slow decay of the ACF, which suggests that correlations persist across large lags.

2. Autocorrelation Function (ACF):

- The ACF quantifies the dependence of the time series on its past values at various lags ℓ.

1.2. Key Points on Stationarity and ACF

1. **Stationarity:**
   - A time series is stationary if its mean and variance are constant over time, and the autocovariance $\Gamma_X(\ell)$ depends only on the lag ℓ, not on time t.
   - In the figures:
     - SRD, fMRI, and stock data are examples of stationary (or weakly stationary) time series.
     - LRD may not be stationary due to the slow decay of the ACF, which suggests that correlations persist across large lags.
2. **Autocorrelation Function (ACF):**
   - The ACF quantifies the dependence of the time series on its past values at various lags ℓ.
   - Stationary time series typically exhibit ACFs that decay rapidly (e.g., exponentially or geometrically) as ℓ increases.

1.2. Key Points on Stationarity and ACF

1. **Stationarity:**
   - A time series is stationary if its mean and variance are constant over time, and the autocovariance $\Gamma_X(\ell)$ depends only on the lag ℓ, not on time t.
   - In the figures:
     - SRD, fMRI, and stock data are examples of stationary (or weakly stationary) time series.
     - LRD may not be stationary due to the slow decay of the ACF, which suggests that correlations persist across large lags.

2. **Autocorrelation Function (ACF):**
   - The ACF quantifies the dependence of the time series on its past values at various lags ℓ.
   - Stationary time series typically exhibit ACFs that decay rapidly (e.g., exponentially or geometrically) as ℓ increases.
   - Non-stationary time series may show very slow or no decay in the ACF.

1. **Autocovariance Function (ACVF):**
   - ACVF measures the covariance of a time series with its lagged version.

1. **Autocovariance Function (ACVF):**

   - ACVF measures the covariance of a time series with its lagged version.
   - For a time series X_t, the ACVF at lag ℓ is defined as:

$$\Gamma_X(\ell) = \mathsf{Cov}(X_t, X_{t+\ell}) = \mathbb{E}[(X_t - \mu)(X_{t+\ell} - \mu)],$$

   where $\mu = \mathbb{E}[X_t]$ is the mean of the time series (assuming stationarity).

1. **Autocovariance Function (ACVF):**
   - ACVF measures the covariance of a time series with its lagged version.
   - For a time series X_t, the ACVF at lag ℓ is defined as:

     $$\Gamma_X(\ell) = \mathsf{Cov}(X_t, X_{t+\ell}) = \mathbb{E}[(X_t - \mu)(X_{t+\ell} - \mu)],$$

     where $\mu = \mathbb{E}[X_t]$ is the mean of the time series (assuming stationarity).
   - **Units:** The autocovariance is measured in the square of the units of X_t, so it depends on the scale of the time series.

2 **Autocorrelation Function (ACF):**

2. **Autocorrelation Function (ACF):**

   - The ACF is the normalized version of the ACVF.

2 Autocorrelation Function (ACF):

- The ACF is the normalized version of the ACVF.
- It measures the linear dependence between X_t and X_{t+h}, scaled to lie between -1 and 1. It is defined as:

$$\rho_X(\ell) = \frac{\Gamma_X(\ell)}{\Gamma_X(0)},$$

where $\Gamma_X(0)$ is the variance of the time series ($\Gamma_X(0) = \mathrm{Var}(X_t)$).

2 Autocorrelation Function (ACF):

- The ACF is the normalized version of the ACVF.
- It measures the linear dependence between X_t and X_{t+h}, scaled to lie between -1 and 1. It is defined as:

$$\rho_X(\ell) = \frac{\Gamma_X(\ell)}{\Gamma_X(0)},$$

  where $\Gamma_X(0)$ is the variance of the time series ($\Gamma_X(0) = \mathrm{Var}(X_t)$).
- **Units:** The autocorrelation is dimensionless and does not depend on the scale of the time series.

1.2 Note 2: Covariance and Variance Relationship at lag ℓ

The covariance, $\text{Cov}(X_t, X_{t+\ell}) = \mathbb{E}[(X_t - \mu)(X_{t+\ell} - \mu)]$ will be equal to the variance of the time series, $\text{Var}(X_t) = \mathbb{E}[(X_t - \mu)^2]$, when the lag $\ell = 0$.

1.2 Note 2: Covariance and Variance Relationship at lag ℓ

The covariance, $\mathrm{Cov}(X_t, X_{t+\ell}) = \mathbb{E}[(X_t - \mu)(X_{t+\ell} - \mu)]$ will be equal to the variance of the time series, $\mathrm{Var}(X_t) = \mathbb{E}[(X_t - \mu)^2]$, when the lag $\ell = 0$.

Explanation:

- **Variance:** The variance of a random variable X_t is a special case of the covariance where the two random variables are identical, i.e., the lag is $\ell = 0$:

$$\mathrm{Var}(X_t) = \mathrm{Cov}(X_t, X_t) = \mathbb{E}[(X_t - \mu)(X_t - \mu)].$$

1.2 Note 2: Covariance and Variance Relationship at lag ℓ

The covariance, $\text{Cov}(X_t, X_{t+\ell}) = \mathbb{E}[(X_t - \mu)(X_{t+\ell} - \mu)]$ will be equal to the variance of the time series, $\text{Var}(X_t) = \mathbb{E}[(X_t - \mu)^2]$, when the lag $\ell = 0$.

Explanation:

- **Variance:** The variance of a random variable X_t is a special case of the covariance where the two random variables are identical, i.e., the lag is $\ell = 0$:

$$\text{Var}(X_t) = \text{Cov}(X_t, X_t) = \mathbb{E}[(X_t - \mu)(X_t - \mu)].$$

- **Covariance at $\ell = 0$:** When $\ell = 0$, the covariance reduces to:

$$\text{Cov}(X_t, X_{t+0}) = \mathbb{E}[(X_t - \mu)(X_t - \mu)] = \text{Var}(X_t).$$

1.2 Note 2: Covariance and Variance Relationship at lag ℓ

The covariance, $\text{Cov}(X_t, X_{t+\ell}) = \mathbb{E}[(X_t - \mu)(X_{t+\ell} - \mu)]$ will be equal to the variance of the time series, $\text{Var}(X_t) = \mathbb{E}[(X_t - \mu)^2]$, when the lag $\ell = 0$.

Explanation:

- **Variance:** The variance of a random variable X_t is a special case of the covariance where the two random variables are identical, i.e., the lag is $\ell = 0$:

$$\text{Var}(X_t) = \text{Cov}(X_t, X_t) = \mathbb{E}[(X_t - \mu)(X_t - \mu)].$$

- **Covariance at $\ell = 0$:** When $\ell = 0$, the covariance reduces to:

$$\text{Cov}(X_t, X_{t+0}) = \mathbb{E}[(X_t - \mu)(X_t - \mu)] = \text{Var}(X_t).$$

- For $\ell \neq 0$, the covariance involves the dependence between X_t and its lagged value $X_{t+\ell}$. Unless the time series is perfectly correlated for that lag ($X_{t+\ell} = X_t$), the covariance at lag ℓ will **not** equal the variance.

1.2 Note 2: Covariance and Variance Relationship at lag ℓ

The covariance, $\text{Cov}(X_t, X_{t+\ell}) = \mathbb{E}[(X_t - \mu)(X_{t+\ell} - \mu)]$ will be equal to the variance of the time series, $\text{Var}(X_t) = \mathbb{E}[(X_t - \mu)^2]$, when the lag $\ell = 0$.

Explanation:

- **Variance:** The variance of a random variable X_t is a special case of the covariance where the two random variables are identical, i.e., the lag is $\ell = 0$:

$$\text{Var}(X_t) = \text{Cov}(X_t, X_t) = \mathbb{E}[(X_t - \mu)(X_t - \mu)].$$

- **Covariance at $\ell = 0$:** When $\ell = 0$, the covariance reduces to:

$$\text{Cov}(X_t, X_{t+0}) = \mathbb{E}[(X_t - \mu)(X_t - \mu)] = \text{Var}(X_t).$$

- For $\ell \neq 0$, the covariance involves the dependence between X_t and its lagged value $X_{t+\ell}$. Unless the time series is perfectly correlated for that lag ($X_{t+\ell} = X_t$), the covariance at lag ℓ will **not** equal the variance.

Conclusion: $\text{Cov}(X_t, X_{t+\ell}) = \text{Var}(X_t) \quad$ if and only if $\ell = 0$.