Chapter 1

Time Series

1.1 Introduction

A random variable is a $(\Omega, \mathscr{F}) \setminus (\mathbb{R}^m, \mathscr{B})$ measure function. A time series is a sequence of random variables $(y_1(\omega), y_2(\omega), \ldots, y_n(\omega)) \in \mathbb{R}^{m \times n}$, and it can be extended to a doubly infinite sequence $(..., y_{t-1}, y_t, y_{t+1}, ...) \in \mathbb{R}^{m \times \infty}$. We consider discrete time series (instead of the continuous time series). For each fixed ω, the sequence is a deterministic vector $(\omega) \in \mathbb{R}^{m \times n}$; for each fixed t, $y_t(\omega)$ is a common random vector in $\mathbb{R}^m$.

1.2 Stationarity

In reality, we have only one realized sequence, but statistics needs repeated observations. We introduce the concept *stationarity* to produce "repeated" observations.

Definition 1.1. (y_t) is **covariance stationarity** or **weakly stationarity** if the mean $\mu = E[y_t]$, covariance $\Sigma = E\left[(y_t - \mu)(y_t - \mu)'\right]$ and autocovariance $\Gamma(\ell) = E\left[(y_t - \mu)(y_{t-\ell} - \mu)'\right]$ are independent of t.

- For a vector-valued weakly stationarity time series, $\Sigma = \Gamma(0)$ is a positive-definite symmetric matrix. The autovariance $\Gamma(\ell), \ell \neq 0$ is not symmetric in general, and

$$\Gamma(-\ell) = E\left[(y_t - \mu)(y_{t+\ell} - \mu)'\right] = E\left[(y_{t-\ell} - \mu)(y_t - \mu)'\right] = \Gamma(\ell)'.$$

- When $m = 1$ (scalar time series), we use $\gamma(0), \gamma(1), ...,$ for the autocovariance, and we define *autocorrelation* as $\rho(\ell) = \gamma(\ell)/\gamma(0)$. By the Cauchy-Schwarz inequality $\rho(\ell) \in [-1, 1]$.

Definition 1.2. (y_t) is *strictly stationarity*, if for every $\ell \in \mathbb{Z}^+$, joint distribution of $(y_t, y_{t+1}, \ldots, y_{t+\ell})$ is independent of t.

When mentioning "stationarity", the default is "strictly stationarity".

- If (y_t) is i.i.d, then it is strictly stationarity.
- If (y_t) is strictly stationarity, its transformation $x_t \in \phi(y_t, y_{t-1}, ...) \in \mathbb{R}^q$ is also strictly stationarity. In other words, strictly stationarity is preserved by transformation.

Series: $x_t = \sum_{j=0}^{\infty} a_j y_{t-j}$

- The infinite series x_t is convergent if the partial sum $\sum_{j=1}^{N} a_j y_{t-j}$ has a finite limit as $N \to \infty$ almost surely.
- If y_t is strictly stationarity, $E\|y\| < \infty$ and $\sum_{j=0}^{N} |a_j| < \infty$ (absolutely summable), then x_t is convergent and strictly stationarity.

1.3 Ergodicity

A time series $\{y_t\}$ is *ergodic* if all invariant events are trivial. Any event unaffected by time shift is of probability 0 or 1. "invariant" means the sequence of a random variable gets stuck somewhere. Ergodicty is preserved by transformation. If $\{y_t\}$ is stationarity and erdogic, the same is for $x_t \in \phi(y_t, y_{t-1}, ...)$ (function with infinite terms).

Example 1.1. If $x_t = \sum_{j=0}^{\infty} a_j y_{t-j}$ if convergent and (y_t) is erdogic, then x_t is also erdogic.

(Cesaro means) If $a_j \to a$ as $j \to \infty$, then $\frac{1}{n}\sum_{j=0}^{\infty} a_j \to a$ as $n \to \infty$.

Theorem 1.1. *If $y_t \in \mathbb{R}^m$ is stationarity and erdogic, and var $(y_t) < \infty$, then $\frac{1}{n}\sum_{\ell=1}^{n} cov(y_t, y_{t+\ell}) \to 0$ as $n \to \infty$*

Definition 1.3. Formal definitions

Let $\widetilde{y}_t = (..., y_{t-1}, y_t, y_{t+1}, ...)$ an event $A \in \{\widetilde{y}_t \in G\}$ for some $G \subseteq \mathbb{R}^{m\times\infty}$.

The ℓ-th time shift is $\widetilde{y}_{t+\ell} = (..., y_{t-1+\ell}, y_{t+\ell}, y_{t+\ell+1}, ...)$ and a time shift of the event is $A_\ell \in \{\widetilde{y}_{t+\ell} \in G\}$.

An event is **invariant** if $A_\ell = A$

An event is **trivial** if $P(A) = 0$ or $P(A) = 1$.

Theorem 1.2. *A stationarity $\{y_t\}$ is erdogic if for all events A and B,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} P(A_\ell \cap B) = P(A)P(B)$$

Let $B = A$, and then we solve $P(A) = [P(A)]^2 \Rightarrow P(A) = 0$ or 1.

A "sufficient" condition for ergodicity is $P(A_\ell \cap B) \to P(A)P(B)$ as $\ell \to \infty$, according to Cesaro means. This sufficient condition is called "mixing".

- Mixing says that separate events (any A and B) are asymptotically independent when A is shifted to A_ℓ as $\ell \to \infty$.
- Ergodicity is slightly weaker than mixing (weak dependence), in the sense that the independence is "on average" in the form of $\frac{1}{n}\sum_{\ell=1}^{n} P(A_\ell \cap B)$.

Theorem 1.3. *Ergodic Theorem:*

$y_t \in \mathbb{R}^m$ is stationarity and erdogic, and $E\|y\| < \infty$, then $E\|\bar{y} - \mu\| \to 0$ and $\bar{y} \xrightarrow{p} \mu$.

Interpretation: Convergence in the 1st mean implies $\xrightarrow{p}$.

1.4 Information Set

- for a univariate time series, definite $E_{t-1}[y_t] = E[y_t \mid y_{t-1}, y_{t-2}, ...]$ as the condition expectation of y_t given the past history $(y_{t-1}, y_{t-2}, ...)$
- More generally, we write $\mathscr{F}_t$ as the smallest σ-field generated by the information up to time t. $\mathscr{F}_t$ is called an "information set".

$$E[y_t \mid \mathscr{F}_{t-1}] = E_{t-1}[y_t]$$

- Information sets are nested $\mathscr{F}_{t-1} \subseteq \mathscr{F}_t \subseteq \mathscr{F}_{t+1}, ...$
- Depends on the definition, when multiple random variables are involved

$$\sigma(y_t, y_{t-1}, ...) \neq \sigma(y_t, x_t, y_{t-1}, x_{t-1}, ...)$$

1.5 Martingale Difference Sequence (MDS)

- Let $\{e_t\}$ be a time series, and $\mathscr{F}_t$ be an information set, $\{e_t\}$ is **adapted** to $\mathscr{F}_t$ if $E[e_t \mid \mathscr{F}_t] = e_t$ ($\mathscr{F}_t$ contain the complete information of e_t. A **natural filtration** is $\mathscr{F}_t = \sigma(e_t, e_{t-1}, ...)$.)
- MDS: a process $\{e_t, \mathscr{F}_t\}$ is MDS if

1. e_t is adapted to $\mathscr{F}_t$
2. $E|e_t| < \infty$
3. $E[e_t \mid \mathscr{F}_{t-1}] = 0$

Interpretation: unforeseeable.
Mean independence. But it does not rule our predictability in other moments.

Example 1.2. $e_t = u_t u_{t-1}$, $u_t \sim i.i.d.\, N(0,1)$
e_t is MDS, but not i.i.d.
The covariance of e_t^2 and e_{t-1}^2 is not 0.
The filtration here is $\mathscr{F}_t = \sigma(u_t, u_{t-1}, ...)$, which subsumes $\sigma(e_t, e_{t-1}, ...)$

$$\begin{aligned} cov(e_t, e_{t-k}) &= E[e_t e_{t-k}] = E[E[e_t e_{t-k} \mid \mathscr{F}_{t-1}]] \\ &= E[e_{t-k} E[e_t \mid \mathscr{F}_{t-1}]] = 0 \end{aligned}$$

- A MDS $(e_t, \mathscr{F}_t)$ is a homoskedastic martingale difference sequence if $E\left[e_t^2 \mid \mathscr{F}_{t-1}\right] = \sigma^2$. $e_t = u_t u_{t-1}$ is MDS, but not homoskedastic.

Theorem 1.4. *CLT for MDS: If $\{u_t\}$ is strictly stationary, ergodic and MDS, then*

$$S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \xrightarrow{d} N\left(0, \sum\right)$$

where $\sum = E[u_t u_t']$. There is the t.s. counterpart of the Lindeberg-Lévy CLT.

1.6 Mixing

we will loose the restriction of MDS. The price are stronger assumptions on the dependence than ergodicity.

- $\alpha(A, B) = |P(AB) - P(A) P(B)|$
- Let two σ-fields be $\mathscr{F}_{-\infty}^{t} = \sigma(..., y_{t-1}, y_t)$ and $\mathscr{F}_t^{\infty} = \sigma(y_t, y_{t+1}, ...)$
- Strong mixing coefficients
$$\alpha(\ell) = \sup_{A \in \mathscr{F}_{-\infty}^{t-\ell}, B \in \mathscr{F}_t^{\infty}} \alpha(A, B)$$
y_t is strong mixing if $\alpha(\ell) \to 0$ as $\ell \to \infty$.
- In general, the α-coefficients should have a sup over t
$$\alpha(\ell) = \sup_t \sup_{A \in \mathscr{F}_{-\infty}^{t-\ell}, B \in \mathscr{F}_t^{\infty}} \alpha(A, B)$$
- A mixing process is ergodic.
- Absolute regularity (β-mixing)
$$\beta(\ell) = \sup_{A \in \mathscr{F}_t^{\infty}} \left| P\left(A \mid \mathscr{F}_{-\infty}^{t-\ell}\right) - P(A) \right|$$
β mixing is stronger than α mixing.
- Strong mixing is preserved by finite transformation.

Theorem 1.5. *y_t has mxing coefficients $\alpha_y(\ell)$. $x_t = \sigma(y_t, y_{t-1}, ..., y_{t-q})$*
Then $\alpha_x(\ell) < \alpha_y(\ell - q)$ for $\ell \geq q$.
The α-coefficients satisfy the same rate and summation properties.

- Rate conditions $\alpha(\ell) = O(e^{-r})$. Summation restriction $\sum_{\ell=0}^{\infty} \alpha(e)^r < \infty$ or $\sum_{\ell=0}^{\infty} e^s \alpha(e)^r < \infty$.
- Thm 14.13 bounds covariances with functions of α-coefficients.

1.7 CLT for Correlated Variables

$$
\begin{aligned}
var\left(S_n\right) &= var\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_t\right) \\
&= \frac{1}{n}\mathbf{I}_N' E\left[YY'\right]\mathbf{I}_N \\
&= \frac{1}{n}\mathbf{I}_N' \begin{bmatrix} \sigma^2 & \gamma(1) & \gamma(2) & & \\ \gamma(1) & \sigma^2 & \gamma(1) & & \\ \gamma(2) & \gamma(1) & \sigma^2 & & \\ & & & \ddots & \\ & & & & \sigma^2 \end{bmatrix} \mathbf{I}_N \\
&= \frac{1}{n}\left(n\sigma^2 + 2(n-1)\gamma(1) + 2(n-2)\gamma(2) + ... + 2\gamma(n-1) + 2\times 0\times\gamma(n)\right) \\
&= \sigma^2 + 2\sum_{\ell=1}^{n}\left(1-\frac{\ell}{n}\right)\gamma(\ell)
\end{aligned}
$$

As $\gamma(-\ell) = \gamma(\ell)$, $var\left(S_n\right) = \sum_{\ell=-n}^{n}\left(1-\frac{|\ell|}{n}\right)\gamma(\ell)$

In vector case, similarly we have

$$
var\left(S_n\right) = \mathit{\Gamma}(0) + \sum_{\ell=1}^{n}\left(1-\frac{\ell}{n}\right)\left(\gamma(\ell)+\gamma(\ell)'\right) = \sum_{\ell=-n}^{n}\left(1-\frac{|\ell|}{n}\right)\gamma(\ell)
$$

- For CLT to work, $var\left(S_n\right)$ must be convergent in the limit

$$
\begin{aligned}
\sum_{\ell=1}^{n}\left(1-\frac{\ell}{n}\right)\gamma(\ell) &= \frac{1}{n}\sum_{\ell=1}^{n}(n-\ell)\gamma(\ell) \\
&= \frac{1}{n}\sum_{\ell=1}^{n-1}\sum_{j=1}^{\ell}\gamma(j) \\
&\to \sum_{j=1}^{\infty}\gamma(j) = \sum_{\ell=1}^{\infty}\gamma(\ell)
\end{aligned}
$$

  by the Theorem of Cesaro means if $\sum_{\ell=1}^{\infty}\gamma(\ell)$ is convergent.

Necessary condition: $\gamma(\ell)\to 0$ as $\ell\to\infty$.

Sufficient: $\sum_{\ell=1}^{\infty}|\gamma(\ell)| < \infty$

It can be show if $E\left\|u_t\right\|^r < \infty$ and $\sum_{\ell=0}^{\infty}\alpha(\ell)^{1-2/\gamma} < \infty$ for some $\gamma > 2$, then $\sum_{\ell=1}^{\infty}|\mathit{\Gamma}(\gamma)| < \infty$ is absolutely convergent.

Theorem 1.6. *(CLT) If y_t is strictly stationarity with α-mixing coefficients $\sum_{\ell=0}^{\infty}\alpha(\ell)^{1-2/\gamma} < \infty$ and $E\left\|u_t\right\|^r < \infty$ for some $\gamma > 2$, $E\left[u_t\right] = 0$, then $S_n \xrightarrow{d} N(0,\mathit{\Omega})$ where $\mathit{\Omega} = \sum_{\ell=-\infty}^{\infty}\mathit{\Gamma}(\gamma)$ is the long-run variance.*

1.8 Linear Projection

- In regression problems, $\mathscr{P}(y \mid X) = X\beta^* = X'(E[XX'])^{-1} E[Xy]$
- Extend to a projection to the infinite past history $\tilde{y}_{t-1} = (y_{t-1}, y_{t-2}, ...)$

Denote $\mathscr{P}_{t-1}(y_t) = \mathscr{P}[y_t \mid \tilde{y}_{t-1}]$, and the projection error $e_t = y_t - \mathscr{P}_{t-1}(y_t)$

Theorem 1.7. *Projection Theorem:*
If $y_t \in \mathbb{R}$ is covariance stationarity, then the projection error statistics
(1) $E[e_t] = 0$
(2) $\sigma^2 = E[e_t^2] \leq E[y_t^2]$
(3) $E[e_t e_{t-j}] = 0$ for all $j \geq 1$.
*In other words, $\{e_t\}$ is a **white noise**.*

- If $\{y_t\}$ is strictly stationarity, then $\{e_t\}$ is strictly stationarity.

Definition 1.4. A time series is a white noise if it is covariance stationarity with 0 autocovariance.
It is helpful to imagine the projection as a linear combination

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + e_t$$

The nature of projection ensures e_t is uncorrelated with all regressions.
e_{t-j} is a linear combination $y_{t-j} - \alpha_1 y_{t-j-1} - \alpha_2 y_{t-j-2} - ....$
Then e_t is uncorrelated with e_{t-j}.

1.9 Wold Decomposition

- If y_t is covariance stationarity, and the linear projection error has $\sigma^2 > 0$, then $y_t = u_t + \sum_{j=0}^{\infty} b_j e_{t-j}$, $b_0 = 1$, and $u_t = \lim_{m\to\infty} \mathscr{P}_{t-m}(y_t)$

Project y_t onto the orthogonal elements $e_t, e_{t-1,} e_{t-2}, ...$ For simplicity, we can consider the case $\mu_t = \mu$.

Definition 1.5. Lag operator: $\mathrm{L}y_t = y_{t-1}$, $\mathrm{L}^2 y_t = \mathrm{L}(\mathrm{L}y_t) = \mathrm{L}y_{t-1} = y_{t-2}$, and so on.

$$\begin{aligned} y_t &= \mu + \sum_{j=0}^{\infty} b_j e_{t-j} \\ &= \mu + \left(b_0 + b_1 \mathrm{L} + b_2 \mathrm{L}^2 + ...\right) e_t \\ &= \mu + b(\mathrm{L})\, e_t \end{aligned}$$

where $b(\mathrm{L})$ is an infinite-order polynomial.

- Autoregressive Wold Representation: If y_t is covariance stationarity with $y_t = u_t + b(\mathrm{L})\, e_t$, then with some additional technical restrictions, $y_t = \mu + \sum_{j=1}^{\infty} a_j y_{t-j} + e_j$.