Financial Time Series

Topic 2: Linear Time Series Analysis

Hung Chen
Department of Mathematics
National Taiwan University
3/8/2002

OUTLINE

1. Fundamental Concepts
2. Ergodicity of a Stochastic Process
3. Stationarity of a Stochastic Process
4. The ACF of a Process
5. The PACF of a Process
6. Stationary Time Series Models
   - AR(1) Processes
   - MA(1) Processes
   - ARMA(1,1) Processes

Fundamental Concepts

- **Time series** data are observations with a particular order (usually time t)
- It is different from cross-sectional data.
- Examples of time series
  - Quartly unemployment rate
  - Monthly car sales
  - Interest rate spread
  - Stock market index values (Dow Jones, Hang Seng and etc)
  - log return of an asset: r_t
- $\{x_1, x_2, x_3, \ldots, x_n\}$, an observed univariate times series such as $\{30.5, 28.9, 33.1, \ldots\}$
- x_t: a single number recorded at time t
- We assume that a time series is a **realization** from a certain stochastic process.
- Purposes:
  - Study the dynamics structure.
  - Monitor price behavior (stock, currency, commodity)

- Understand the probable development of prices in the future
- Forecast the volatility of future prices.
- Understand how prices behave
- Use knowledge of price behavior to take better decisions

  Suppose we have planned a holiday abroad and will need to buy foreign currency. The purchase could be made months in advance or left until you arrive at your destination.

- How do we predict x_{n+1} at time $t = n$?
- Motives for trading:
  - Reduction of business risks
  - Purchases and sales of raw materials
  - Investment of personal or corporate wealth
- Trading possibilities
  - Spot markets: stock and share markets
  - Future markets: agree a price for exchanging goods at some later date

– Option markets: gives its owner the right to engage in a particular spot or future transaction at a formerly agreed price.

Stochastic Processes

Data:

- A **stochastic process** is a family of time indexed random variables $\{X_{t_1}, X_{t_2}, X_{t_3}, \ldots\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ where $t_i \in T$.
- Treat x_t as the realized value of some random variable X_t.
- Tomorrow's price is uncertain and it can be modelled by a probability distribution.
- What is a random variable?
- What is the distinction between a random variable and a function?
- In this course, $T = \{0, \pm 1, \pm 2, \ldots\}$.
- The observed series $(x_1, x_2, \ldots, x_T)$, $\{x_t\}_1^T$ is considered as a realization of a stochastic process $\{X_t\}_{-\infty}^{\infty}$.

Ergodicity

- Can we use a single realization to infer the unknown parameters of a joint probability distribution?
- The sample moments for finite stretches of the realization approach their population counterparts as the length of the realization becomes infinite.

Consider the following toy example.

- Suppose we have six coins with unkown probabilities of success $p_1, \cdots, p_6$, respectively. On those coins, they are marked by 1 through 6. Conduct the following experiment.
  - Roll a fair die once and denote the outcome by I.
  - Pick the coin which number matches I and flip it continuously.
  - From this experiment, can we estimate $p_1, \cdots, p_6$ consistently?

Stationarity of a Stochastic Process:

How do we describe $(X_1, X_2, \ldots, X_T)$?

- Use a T-dimensional probability distribution.
  In practice, it is difficult to find the cdf from an observed time series.
- Model its first and second moments.
  T means: $E(X_1), E(X_2), \ldots, E(X_T)$,
  T variances: $V(X_1), V(X_2), \ldots, V(X_T)$,
  $T(T-1)/2$ covariances:
  $$Cov(X_i, X_j), \quad i < j.$$
- When $(X_1, X_2, \ldots, X_T)$ are normally distributed, the first two moments completely characterize its properties.

Question: We have too many parameters $(T + T(T+1)/2)$ and too few observations (T).
Solution: Reduce the number of unknown parameters by imposing additional structure.

Strictly Stationarity:

The properties of $\{X_t\}_{-\infty}^{\infty}$ are unaffected by a change of time origin.

- Time invariant: independent of time origin
- It is hard to verify in practice.

Implication:

- The joint probability distribution at any set of times $t_1, t_2, \ldots, t_m$ must be the same as the joint probability distribution at times $t_1 + k, t_2 + k, \ldots, t_m + k$, where k is an arbitrary shift along the time axis.
- For $m = 1$, the marginal probability distributions do not depend on time.
- When $E|X_t|^2 < \infty$,

  $$E(X_1) = E(X_2) = \cdots = E(X_T) = E(X_t) = \mu,$$

  and

  $$V(X_1) = V(X_2) = \cdots = V(X_T) = V(X_t) = \sigma_X^2.$$

- For $m = 2$, all bivariate distributions do not depend on t. Hence, for all k,

  $$\begin{aligned} Cov(X_1, X_{1+k}) &= Cov(X_2, X_{2+k}) = \cdots \\ &= Cov(X_{T-k}, X_T) = Cov(X_{t-k}, X_t). \end{aligned}$$

- Both autocovariances and autocorrelations depend only on the lag (or time difference) k only.

  $\gamma_k = Cov(X_{t-k}, X_t) = E[(X_{t-k}-\mu)(X_t-\mu)]$,

  and

  $$\rho_k = \frac{Cov(X_{t-k}, X_t)}{[V(X_t) \cdot V(X_{t-k})]^{1/2}} = \frac{\gamma_k}{\gamma_0}.$$

Example 1:
Sinusoid with Random Phase and Amplitude
(continuous-time stochastic process)

- $X_t = A \sin(\omega t + \theta)$ where A is uniformly distributed on the interval $[-a, a]$, θ is uniformly distributed on the interval $[0, \pi]$ independent of A. The ω and a are fixed constants.

- The first moment does not depend on t.

  $$E(X_t) = E(A) \cdot E[\sin(\omega t + \theta)] = 0$$

  because $E(A) = 0$.

- Covariance stationary:

  $Cov(X_t, X_{t-k})$

$$
\begin{aligned}
&= E(A^2) \cdot E[\sin(\omega t + \theta)\sin(\omega(t-k)+\theta)] \\
&= \frac{a^2}{3}\left\{\frac{1}{2}\cos(\omega k) - \frac{1}{2}E[\cos(\omega(2t+k)+2\theta)]\right\} \\
&= \frac{a^2}{3}\left\{\frac{1}{2}\cos(\omega k) - \frac{1}{2}\int_0^{\pi}\cos(\omega(2t+k)+2\theta)\frac{1}{2\pi}d\theta\right\} \\
&= \frac{a^2}{6}\cos(\omega k),
\end{aligned}
$$

which does not depend on time t.

Weak (Second-order) Stationarity:

- A weaker version of stationarity is often assumed.
- The series $\{X_t\}$ is *weakly stationarity* if both the mean of X_t and the covariance between X_t and X_{t-k} are time-invariant.

$$\begin{aligned}
&E(X_1) = E(X_2) = \cdots = E(X_T) = E(X_t) = \mu, \\
&V(X_1) = V(X_2) = \cdots = V(X_T) = V(X_t) = \sigma_X^2, \\
&Cov(X_1, X_{1+k}) = Cov(X_2, X_{2+k}) = \cdots \\
&\qquad = Cov(X_{T-k}, X_T) = Cov(X_{t-k}, X_t).
\end{aligned}$$

- In the finance literature, it is common to assume that an asset return series is weak stationary.
- This assumption can be checked empirically provided that sufficient number of historical returns are available.

The ACF of a Process

- Under the assumption of weak stationarity, the autocorrelations can be viewed as a function of k and will be referred to as the **autocorrelation function** (ACF), (ρ_k).

- ACF only defined on a stationary process.
- ACF gives a measure on the correlation of one value of the process with previous values.
- It indicates the length and strength of the **memory** of the process.
- Since time series data are **correlated**, it is important to study autocorrelation of the process.
- In fact, most time series models are "models" of autocorrelations.
- The estimation of ACF from an observed time series data set later.
- The ACF matrix

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \cdots \\ \rho_1 & 1 & \rho_1 & \rho_2 & \cdots \\ \rho_2 & \rho_1 & 1 & \rho_1 & \cdots \\ \rho_3 & \rho_2 & \rho_1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

must be **positive semidefinite**.

Stationary Time Series Models

- We have to find a broad class of models which can describe a wide variety of observed time series.

- Linear time series models (linear process): A time series X_t is said to be linear if it can be written as
$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}$$
where $\psi_0 = 1$ and $\{a_t\}$ is a sequence of independent and identically distributed random variables with mean zero and a well-defined distribution.
$\{a_t\}$ is a white noise process.
$\{X_t\}$ is generated as a linear filter of strict white noise.

- White Noise Process
  - $\{a_t\}$ is a white noise process if
    (i) $E(a_t) = 0$,
    (ii) $V(a_t) = \sigma^2 < \infty$,
    (iii) $Cov(a_t, a_{t-k}) = 0$ for all $k \neq 0$.
  - $\{a_t\}$ is denoted as $a_t \sim WN(0, \sigma^2)$ later on.

- Although white noise process is not commonly encountered in applied time series, it is an important building block (called innovations) of most modern time series models.
- $\rho_k = 0$ for $k > 0$.
- If the a_t's are also independent, then the sequence $\{a_j\}$ is termed **strict white noise**, denoted $a_t \sim SWN(0, \sigma^2)$.

- Non-linear time series
- Two types of time series:
  stationary and nonstationary
- Several classes of stationary time series models:
  - Autoregressive Models: $AR(1)$, $AR(2)$, $AR(p)$
  - Moving Average Models: $MA(1)$, $MA(2)$, $MA(q)$
  - Mixed Models: $ARMA(1,1)$, $ARMA(p,q)$

Many financial time series are certainly not stationary.
They have a tendency to exhibit time-changing

means and/or variances.
Box and Jenkins (1976): ARIMA.
Here I refers to “integrated.”

Wold's Decomposition: Theorem 5.7.1 of Brockwell and Davis (1991)

- Any zero-mean stationary process $\{X_t\}$ which is not deterministic can be expressed as a sum $X_t = U_t + V_t$.
- $\{U_t\}$: an $MA(\infty)$ process
$$U_t = \psi_0 a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots,$$
where $\psi_0 = 1$, $\Sigma_{j=0}^{\infty} \psi_j^2 < \infty$.
- $\{V_t\}$: a deterministic process which is uncorrelated with $\{a_t\}$
(i.e., $E(a_t V_s) = 0$ for all (s, t).).
- A process is called deterministic if the values of X_{n+j}, $j \geq 1$, were perfectly predictable in terms of elements of closed span $\{X_t, -\infty < t \leq n\}$. i.e.,
$$\sigma^2 = E|X_{n+1} - \mathbf{P}_{\mathcal{M}_n} X_{n+1}|^2 = 0.$$
$\mathbf{P}_{\mathcal{M}_n} X_{n+1}$ is the best linear predictor of X_{n+1} based on X_t up to n.
- $a_t = X_t - \mathbf{P}_{\mathcal{M}_{t-1}} X_t$
- $\psi_j =< X_t, a_{t-j} > /\sigma^2$

Statement in the book:

Every weakly stationary, purely nondeterministic process $(X_t - \mu)$ can be written as a linear combination (or linear filter) of a sequence of uncorrelated random variables. Here

- purely nondeterministic means that any linearly deterministic component has been subtracted from X_t.
- Such a component is one that can be perfectly predicted from past values of itself and examples commonly found are a (constant) mean, as is implied by writing the process as $(X_t - \mu)$, periodic sequences, and polynomial exponential sequences in t.
- The linear filter representation is given by

$$\begin{aligned} X_t - \mu &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \\ &= \sum_{j=0}^{\infty} \psi_j a_{t-j}. \end{aligned} \tag{1}$$

Here $a_t \sim WN(0, \sigma^2)$ and the coefficients in the linear filter are known as ψ-weight.

Properties:

$$
\begin{aligned}
E(X_t) &= \mu, \\
\gamma_0 &= V(X_t) = E(\sum_{j=0}^{\infty} \psi_j a_{t-j})^2 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2, \\
\gamma_k &= E(X_t - \mu)(X_{t-k} - \mu) \\
&= E(a_t + \psi_1 a_{t-1} + \cdots + \psi_k a_{t-k} \\
&\quad + \psi_{k+1} a_{t-k-1} + \cdots) \cdot (a_{t-k} + \psi_1 a_{t-k-1} + \cdots) \\
&= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, \\
\rho_k &= r_k / r_0 = \sum_{j=0}^{\infty} \psi_j \psi_{j+k} / \sum_{j=0}^{\infty} \psi_j^2,
\end{aligned}
$$

where $\psi_0 = 1$.
The last one gives the ACF of the moving average representation of a process.

- Linear time series models are econometric and statistical models that describes the pattern of ψ-weights.
- If $\Sigma_{j=0}^{\infty} |\psi_j| < \infty$, then $\Sigma_{j=0}^{\infty} \psi_j a_{t-j}$ converges absolutely with probability one.
- $\Sigma_{j=0}^{\infty} |\psi_j| < \infty$ guarantees that the process $\{X_t\}$ to be stationary.
- The **backshift** operator B:
  It shifts time one step back.
  Hence $Ba_t = a_{t-1}$ or $B^m a_t = a_{t-m}$.

(1) can be written as

$$X_t - \mu = \psi(B)a_t,$$

where $\psi(B) = \Sigma_{j=0}^{\infty} \psi_j B^j$ with $\psi_0 = 1$.

- $\{X_t\}$ is said to be a moving average ($MA(\infty)$) of $\{a_t\}$.
- Mathematically speaking, $\Sigma_{j=0}^{\infty} |\psi_j| < \infty$ is equivalent to require all the roots of the polynomial $\psi(B)$ lie outside the unit circle.
- It means that if δ is a real-valued root of $\psi(B)$, we require $|\delta| > 1$; and if δ is a complex root, say, $\delta = c + di$, we require $|\delta| = \sqrt{c^2 + d^2} > 1$.

ARMA Processes:

$AR(1)$ process (autoregressive process of order 1):

- Choose $\psi_j = \phi^j$ in Wold's decomposition.
- Stationarity:
  $\Sigma_j |\phi^j| < \infty$ if $|\phi| < 1$.
- The linear filter representation converges if $|\phi| < 1$.
- $AR(1)$ process:

$$\begin{aligned} X_t - \mu &= 1 \cdot a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} \cdots \\ &= a_t + \phi a_{t-1} + \phi^2 a_{t-2} \cdots \\ &= a_t + \phi(a_{t-1} + \phi a_{t-2} + \cdots) \\ X_{t-1} - \mu &= a_{t-1} + \phi a_{t-2} + \cdots . \end{aligned}$$

  It means that X_t is an infinite-order weighted average of the present and past innovations. Or

$$X_t - \mu = \phi(X_{t-1} - \mu) + a_t. \qquad (2)$$

  It means that X_t depends linearly on X_{t-1} and the innovation a_t alone.
- Invertibility: $|\phi| < 1$

- If a process can be re-written as an AR representation, we call that process **invertiable**.
- Why do we call it an AR model?
  This is the same form as the well-known simple linear regression in which X_t is the dependent variable and X_{t-1} is the explanatory variable.
- Regress $X_t - \mu$ on $X_{t-1} - \mu$.
  $$\min_{a \in R} E[(X_t - \mu) - a(X_{t-1} - \mu)]^2$$
  versus
  $$\min_{\theta \in L^2} E[X_t - \theta(X_{t-1})]^2$$
- ϕ: autoregressive parameter
- For the $AR(1)$ model, it can be written as
  $$(1 - \phi B)(X_t - \mu) = a_t.$$
  Here B is the **backshift** operator.
- Hence,
  $$\begin{aligned} X_t - \mu &= (1 - \phi B)^{-1} a_t = (1 + \phi B + \phi^2 B^2 + \cdots) a_t \\ &= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots . \end{aligned} \tag{3}$$
  This is called an MA (moving average) representation of a process.

ACF of an $AR(1)$ process:

It follows from (2) that for all $k > 0$,

$$
\begin{aligned}
E(X_t - \mu)(X_{t-k} - \mu) &= \phi E(X_{t-1} - \mu)(X_{t-k} - \mu) \\
&\quad + E a_t (X_{t-k} - \mu) \\
V(X_t) &= \phi^2 Var(X_{t-1}) + Var(a_t) \\
V(X_t) &= \frac{\sigma^2}{1 - \phi^2} \\
\gamma_k &= \phi \gamma_{k-1} = \phi^k \gamma_0. \qquad (4)
\end{aligned}
$$

ACF ($\rho_k = \phi^k$): tail-off at lag 1
Observe that

- $\phi > 0$: ACF decays exponentially to zero,
- $\phi < 0$: ACF decays in an oscillatory pattern,
- ϕ is close to the non-stationary boundaries of 1 and -1: ACF decays slowly.

Question:

- For all AR processes, their ACFs damp out.
- It is difficult to distinguish between processes of different orders.

Partial Autocorrelation Function (PACF)

- The correlation between two random variables is often due to both variables being correlated with a third.
- A large portion of the correlation between X_t and X_{t-k} may due to the correlation this pair have with the intervening observations $X_2, \cdots, X_k$.
- The partial autocorrelation $\alpha(k)$ at lag k may be regarded as the correlation between X_1 and X_{k+1}, adjusted for the intervening observations $X_{t-1}, \cdots, X_{t-k-1}$.
- $\alpha(k)$ is the correlation of the two residuals obtained after regressing X_{k+1} and X_1 on the intermediate observations $X_2, \cdots, X_k$.
- $\alpha(1) = Corr(X_2, X_1) = \rho_1$

- residuals after regressing X_{k+1} on the intermediate observations $X_2, \cdots, X_k$:

$$X_{k+1} - \alpha_0 - \sum_{i=2}^{k} \alpha_i X_i$$

where, for $j = 2, \ldots, k$,

$$\begin{aligned} \alpha_0 E(X_j) + \sum_{i=2}^{k} \alpha_i E(X_i X_j) &= E(X_{k+1} X_j), \\ \alpha_0 + \sum_{i=2}^{k} \alpha_i E(X_i) &= E(X_{k+1}), \end{aligned}$$

and $\alpha_0 + \Sigma_{i=2}^{k} \alpha_i X_i$ is the so-called best linear prediction of X_{k+1} based on $X_2, \cdots, X_k$.

- $\alpha(k) = Corr(X_t, X_{t+k} | X_{t+1}, \cdots, X_{t+k-1})$

Example: zero mean $AR(1)$

$$X_t = 0.9 X_{t-1} + a_t.$$

We have

$$\begin{aligned} \alpha(1) &= Corr(X_2, X_1) \\ &= Corr(0.9 X_1 + a_2, X_1) = 0.9. \end{aligned}$$

The best linear prediction of X_{k+1} based on $1, X_2, \ldots, X_k$ is $0.9 X_k$ and the best linear prediction of X_1 based on $1, X_2, \ldots, X_k$ is also

$0.9X_k$ since $(X_1, \ldots, X_k)$ has the same covariance matrix as $(X_{k+1}, X_k, \ldots, X_2)$.
Hence, for $k \geq 2$,

$$\begin{aligned} \alpha(k) &= Corr(X_{k+1} - 0.9X_k, X_1 - 0.9X_2) \\ &= Corr(a_{k+1}, X_1 - 0.9X_2) = 0. \end{aligned}$$

- PACF of $AR(1)$: cut-off at lag 1
  It is a useful tool to determine the order p of an AR model.

- The kth partial autocorrelation is the coefficient ψ_{kk} in the $AR(k)$ process

$$x_t = \psi_{k1} x_{t-1} + \cdots + \psi_{kk} x_{t-k} + a_t. \quad (5)$$

- For general stationary process, refer to page 19 for ψ_{kk} or $\alpha(k)$.

$MA(1)$ process
(first-order moving average process):

There are several ways to introduce MA models.
We first consider the approach by treating the models as a simple extension of white noise series.
Choose $\psi_1 = -\theta$ and $\psi_j = 0$, $j \geq 2$. Then we have

$$
\begin{aligned}
X_t - \mu &= a_t - \theta a_{t-1} \\
X_t - \mu &= (1 - \theta B) a_t. \qquad (6)
\end{aligned}
$$

- Except for the constant term, X_t is a weighted average of shocks a_t and a_{t-1}.
- Stationarity: MA are always weakly stationary, because they are finite linear combinations of a white noise sequence. (a finite number of nonzero ψ_j)
- X_t is a linear combination of the present and immediately preceding innovations.
- Observations one period apart are correlated, observations more than one period apart are uncorrelated.

The **memory** of the process is just one period.

- Another approach to treat MA as infinite-order AR models with some parameter constraints.
  AR representation:
  Consider $(1-\theta B)^{-1}X_t = a_t$.
  Expanding $(1-\theta B)^{-1}$ yields

  $$(1+\theta B+\theta^2 B^2+\cdots)(X_t-\mu) = a_t.$$

  Hence

  $$X_t-\mu = \pi_1(X_{t-1}-\mu)+\pi_2(X_{t-2}-\mu)+\cdots+a_t,$$

  where $\pi_j = -\theta^j$ and the π-weights converge ($\Sigma\,|\pi_j| < \infty$) if $|\theta| < 1$.

- Invertibility: if $|\theta| < 1$

- When $|\theta| < 1$, it means that the effect of past observations decreases with age.

- Not every process is invertible. Consider $X_t = (1-2B)a_t$.

  $$\begin{aligned}(1-2B)^{-1}X_t &= a_t \\ (1-2B+4B^2-8B^3+\cdots)X_t &= a_t.\end{aligned}$$

The sum $|1| + |-2| + |4| + |-8| + \cdots$ is explosive, therefore, this process is noninvertible.

ACF of an $MA(1)$ process:
It follows from (6) that for all $k > 0$,

$$\begin{aligned}
E(X_t - \mu)(X_{t-k} - \mu) &= E(a_t - \theta a_{t-1}) \\
&\quad \cdot (a_{t-k} - \theta a_{t-k-1}), \\
\gamma_k &= 0 \quad k > 1, \\
\gamma_1 &= -\sigma^2\theta, \\
\gamma_0 &= \sigma^2(1+\theta^2) \\
\rho_1 &= \frac{-\theta}{1+\theta^2} = \frac{-1/\theta}{1+(1/\theta)^2}, \\
\rho_k &= 0 \quad \text{for } k > 1.
\end{aligned}$$

Remarks:

- $-0.5 < \rho_1 < 0.5$ by $\rho_1 = -\theta/(1+\theta^2)$
- Both the following $MA(1)$ models have the same ACF:

$$\begin{array}{ll} \text{Model A} & X_t - \mu = (1 - 0.4B)a_t \\ \text{Model B} & X_t - \mu = (1 - 2.5B)a_t \end{array}$$

  which is $\rho_1 = -0.34483$.

- Model A is invertible while the Model B is not.
- For simplicity, we only consider invertible models.
- ACF: cut-off at lag 1
  The ACF is useful in identifying the order of an MA model.
- PACF: tail-off at lag 1
- This model is said to have *short memory*.

Autoregressive-moving average models: $ARMA(1,1)$ process

In some applications, the AR or MA models become cumbersome as the models may need many parameters to adequately describe the serial dependence of the data.

$$\begin{aligned}(X_t-\mu)-\phi(X_{t-1}-\mu) &= a_t-\theta a_{t-1}\\ (1-\phi B)(X_t-\mu) &= (1-\theta B)a_t. \quad (7)\end{aligned}$$

- The left-hand side is the AR component of the model and the right-hand side gives the MA component.
- For this model to be meaningful, we require that $\phi \neq \theta$; otherwise, there is a cancellation in the equation and the process reduces to a white noise series.
- The idea of $ARMA$ model is highly relevant in modeling the volatility of asset returns.
- This representation is compact and useful in parameter estimation.
  It is also useful to compute recursively multi-step forecasts of X_t.

Now we give the other two representations.

- The ψ-weights in the $MA(\infty)$ representation are

$$\psi(B) = \frac{1-\theta B}{1-\phi B}, \tag{8}$$

$$\begin{aligned} X_t - \mu &= \psi(B)a_t = \left(\sum_{i=1}^{\infty} \phi^i B^i\right)(1-\theta B)a_t \\ &= a_t + (\phi - \theta)\sum_{i=0}^{\infty} \phi^{i-1} a_{t-i}. \end{aligned} \tag{9}$$

  - This representation shows the impact of the past shock a_{t-i} $(i > 0)$ on the current return X_t.
  - It is also useful in computing the variance of forecast errors.
  - It can be used to prove the mean reversion of a stationary time series.

- The π-weights in the $AR(\infty)$ representation are

$$\begin{aligned} \pi(B) &= \frac{1-\phi B}{1-\theta B}, \\ \pi(B)(X_t - \mu) &= \left(\sum_{i=0}^{\infty} \theta^i B^i\right)(1-\psi(B)) \\ &\quad \cdot (X_t - \mu) = a_t \\ X_t - \mu &= (\phi - \theta)\sum_{i=0}^{\infty} \theta^{i-1}(X_{t-i} - \mu) \\ &\quad + a_t. \end{aligned}$$

- This representation shows the dependence of the current return X_t on the past returns X_{t-i} $(i > 0)$.
- It shows that the contribution of the lagged value X_{t-i} to X_t is diminishing as i increases, the π_i coefficient should decay to zero as i increases.

- The model is stationary if $|\phi| < 1$.
- The model is invertible if $|\theta| < 1$.
- Variance

$$\begin{aligned} V(X_t) &= \phi^2 V(X_{t-1}) + (1+\theta^2)\sigma^2 \\ &\quad +2\phi Cov(a_t, X_{t-1}-\mu) \\ &\quad -2\theta Cov(a_t, a_{t-1}) \\ &\quad -2\phi\theta Cov(a_{t-1}, X_{t-1}-\mu) \\ &= \frac{(1+\theta^2-2\phi\theta)\sigma^2}{1-\phi^2}. \end{aligned}$$

- ACF

$$\begin{aligned} &E(X_t-\mu)(X_{t-k}-\mu) \\ &= \phi E(X_{t-1}-\mu)(X_{t-k}-\mu) \\ &\quad +E[a_t(X_{t-k}-\mu)] - \theta E[a_{t-1}(X_{t-k}-\mu)]. \end{aligned}$$

$$
\begin{aligned}
\gamma_k &= \phi\gamma_{k-1}, \quad \text{for } k \geq 2, \\
\gamma_1 &= \phi\gamma_0 - \theta E\{[\phi(X_{t-2} - \mu) + a_{t-1} - \theta a_{t-2}]a_{t-1}\} \\
&= \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2}\sigma^2.
\end{aligned}
$$

Thus

$$
\rho_k = \begin{cases} 1 & k = 0 \\ \frac{(\phi-\theta)(1-\phi\theta)}{1+\theta^2-2\phi\theta} & k = 1 \\ \phi\rho_{k-1} & k \geq 2 \end{cases}.
$$

- The ACF of an $ARMA(1,1)$ process is similar to that of an $AR(1)$ process.
  The autocorrelation decay exponentially at a rate ϕ.
- PACF: tail off at lag 1
  It can be calculated easily by the formula in page 19.