Financial Time Series

Topic 10: Revisit Linear and Nonlinear Time Series

Hung Chen
Department of Mathematics
National Taiwan University
6/12/2002

OUTLINE

1. Basic Concepts
2. Linear Time Series:
   - $AR(p)$ Processes
   - $MA(q)$ Processes
   - $ARMA(p, q)$ Processes
   - $ARIMA$ Processes (Non-stationary)
3. Nonlinear Time Series:

Basic Concepts

Time series is a sequence of random variables measuring certain quantity of interest over time.

- A time series is a record of values of certain quantity of interest taken at different time points.
- Usually, data are observed at equally spaced time intervals, resulting in a discrete-time series.
- If treated as a stochastic process over time (continuous time), then we have a continuous-time time series.
- X_t can be a continuous variable or a discrete random variable, e.g. counts.

The objective of time series analysis is to find the dynamic dependence of X_t on its past values $\{X_{t-1}, X_{t-2}, \ldots\}$.

- Consider $\phi(B)X_t = c$ where c is a constant and $\phi(B) = \Sigma_{i=0}^{p} \phi_i B^i$.
  This equation is used in time series analysis to describe the dynamic dependence of X_t on its past values.

- A variety of dynamic dependence patterns of X_t can be generated by considering the *rational* lag polynomial $\pi(B) = \phi(B)/\theta(B)$.
- Consider $\pi(B) = 1/(1 - \theta B)$.
  - We have
  $$\frac{1}{1-\theta B} X_t = \sum_{i=0}^{\infty} \theta^i X_{t-i}.$$
  - If the $\{X_t\}$ sequence is bounded, we might want the resulting sequence to be bounded as well. This is achieved by requiring $|\theta| < 1$.
  - This special rational polynomial shows that X_t is an infinite-order moving average of its past values, $\{X_{t-1}, X_{t-2}, \ldots\}$, with weights decaying exponentially.
- First-order Difference Equations
  - $X_t = \phi X_{t-1} + c + ba_t$ where a_t is a forcing variable, which follows a well-defined probability distribution.
  - Write the model as
  $$(1 - \phi B) X_t = c + ba_t.$$

– The solution is

$$X_t = \frac{c}{1-\phi B} + \frac{b}{1-\phi B} a_t + \gamma \phi^t.$$

Or,

$$X_t = \frac{c(1-\phi^t)}{1-\phi} + b \sum \phi^i a_{t-i} + \phi^t X_0.$$

– If $|\phi| < 1$ and a_t is bounded with mean zero, then X_t approaches $c/(1-\phi)$ from any starting point.

- Second-order Difference Equations

  – $X_t = \phi_1 X_{t-1} + \phi_1 X_{t-2} + \delta + b a_t$

  – Write the model as

  $$(1 - \phi_1 B - \phi_2 B^2) X_t = \delta + b a_t.$$

  – Write $1 - \phi_1 B - \phi_2 B^2$ as $(1 - \lambda_1 B)(1 - \lambda_2 B)$.

  – The solution of $(1 - \phi_1 B - \phi_2 B^2) X_t = 0$ is

  $$X_t = c_1 \lambda_1^t + c_2 \lambda_2^t.$$

  Here $1/\lambda_1$ and $1/\lambda_2$ are the zeros of the polynomial $1 - \phi_1 B - \phi_2 B^2$.
  If the solution is to remain bounded, we

would require $|\lambda_i| < 1$ for $i = 1, 2$, or equivalently that the zeros of the polynomial $1 - \phi_1 B - \phi_2 B^2$ lie outside the unit circle (modulus ¿ 1.

– For a second-order equation, we have three possibilities: (a) distinct real roots, (b) equal real roots, and (c) complex roots.

– Consider complex roots, we can write

$$\begin{aligned} X_t &= c_1 r^t e^{it\theta} + c_2 r^t e^{-it\theta} \\ &= r^t[(c_1 + c_2)\cos(t\theta) + i(c_1 - c_2)\sin(t\theta)] \\ &= k r^t \cos(t\theta + w). \end{aligned}$$

We can see how oscillatory solutions are possible.

– Consider equal root case.

$$(1 - \lambda B)^2 X_t = 0.$$

The solution is

$$X_t = c_1 \lambda^t + C_2 t \lambda^t.$$

Stationary Time Series

Weak stationary: $\{X_t\}$

- If we would like to predict the future of the process, we must be able to identify some key features of the distribution of the process that are **time invariant**.
- A particular time-invariant feature which has proven to be useful is the stationarity.
- A time series $\{X_t\}$ is (strictly) stationary if

$$F_{X_t,\cdots,X_{t+s}}(*) = F_{X_{t+r},\cdots,X_{t+r+s}}(*)$$

  for all r and s.

- Weak stationarity:
  Both the mean of X_t and covariance between X_t and $X_{t-\ell}$ are time-invariant.
- Correlation coefficient between two random variables X and Y:

$$\begin{aligned}\rho_{X,Y} &= \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \\ &= \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{E(X-\mu_X)^2E(Y-\mu_Y)^2}}.\end{aligned}$$

A sample estimate is

$$\hat{\rho}_{X,Y} = \frac{\Sigma_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\Sigma_{t=1}^{T}(x_t - \bar{x})^2 \, \Sigma_{t=1}^{T}(y_t - \bar{y})^2}}.$$

- Law of Large Numbers:
  Chebyschen inequality:

  $$P(|U - E(U)| > \epsilon) < \frac{Var(U)}{\epsilon^2}.$$

  Consistency: $\{T_n\}$

  $$\lim_{n\to\infty} P(|T_n - c| > \epsilon) = 0$$

  for any given $\epsilon > 0$.

- For a weak stationary process X_t, the lag-k autocorrelation is

  $$\rho_\ell = Cov(X_{t-\ell}, X_t)/Var(X_t).$$

- Treating ρ_k as a function of k, we call ρ_k the autocorrelation function of X_t.

- How do we estimate the parameters of a particular time series model using observations of a single realization?
  - Since a stationary time series is time invariant, the distribution of X_t is the same as the distribution of X_s.

- We can treat the single realization $X_1, X_2, \ldots, X_n$ as a sample of n observations from a distribution.
- The mean μ of X_t can be estimated by

$$\hat{\mu} = n^{-1} \sum_{i=1}^{n} X_i.$$

- $\hat{\mu}$ is a *time* average.
- If a time series satisfies the requirement that the *time* averages converges to the *ensemble* averages, then the series is said to be ergodic.

• the lag-ℓ sample autocorrelation, $0 \leq \ell < T - 1$,

$$\hat{\rho}_\ell = \frac{\Sigma_{t=\ell+1}^{T}(X_{t-\ell} - \bar{X})(X_t - \bar{X})}{\Sigma_{t=1}^{T}(X_t - \bar{X})^2}$$

where $\bar{X} = \Sigma_{t=1}^{T} X_t / T$.

• Assume that $\rho_j = 0$ for all $j > \ell$ and r_t is weakly stationary with finite ℓ-th order moment.
The asymptotic variance of $\hat{\rho}_\ell$ is

$$T^{-1}\left(1 + 2\sum_{i=1}^{\ell-1} \rho_i^2\right).$$

- The above result can be used to perform the hypothesis testing $H_0 : \rho_\ell = 0$ versus $H_a : \rho_\ell \neq 0$.

Linear time series

- Linear filter representation:
  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 $N(0,1)$ random variables.

- The coefficients ψ_i are called the ψ-weights in the time series literature.

- Mean and autocorrelations:

$$E(X_t) = \mu + \sum_{i=0}^{\infty} \psi_i E(a_{t-i}) = \mu,$$

$$\begin{aligned} Var(X_t) &= Var(\mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}) \\ &= \sum_{i=0}^{\infty} \psi_i^2 Var(a_{t-i}), \end{aligned}$$

$$\rho_\ell = \frac{\Sigma_{i=\ell}^{\infty} \psi_i \psi_{i-\ell}}{1 + \Sigma_{i=1}^{\infty} \psi_i^2}.$$

Properties of AR models

$AR(1)$: $X_t = \phi_0 + \phi_1 X_{t-1} + a_t$

- Note that

$$\begin{aligned} \mu &= \phi_0 + \phi_1 \mu, \\ E(X_t) &= \mu = \frac{\phi_0}{1-\phi_1} \end{aligned}$$

  by $E(X_t) = \phi_0 + \phi_1 E(X_{t-1})$.

- Using $\phi_0 = (1-\phi_1)\mu$, $AR(1)$ can be written as

$$X_t - \mu = \phi_1 (X_{t-1} - \mu) + a_t.$$

- Note that $Cov(X_{t-1}, a_t) = 0$. We have

$$\begin{aligned} Var(X_t) &= \phi_1^2 Var(X_{t-1}) + \sigma_a^2 \\ Var(X_t) &= \frac{\sigma_a^2}{1-\phi_1^2} \end{aligned}$$

  provided that $\phi_1^2 < 1$.

$AR(2)$: $X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t$

- Note that

$$E(X_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

if $\phi_1 + \phi_2 \neq 1$.

- Using $\phi_0 = (1 - \phi_1 - \phi_2)\mu$, $AR(2)$ can be written as

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + a_t.$$

- Multiplying the above equation by $X_{t-\ell} - \mu$, we have

$$\gamma_\ell = \phi_1 \gamma_{\ell-1} + \phi_2 \gamma_{\ell-2}$$

for $\ell > 0$ and

$$\rho_\ell = \phi_1 \rho_{\ell-1} + \phi_2 \rho_{\ell-2}$$

for $\ell > 2$.

- ACF satisfies the 2nd order difference equation

$$(1 - \phi_1 B - \phi_2 B^2)\rho_\ell = 0.$$

$AR(p)$:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + a_t.$$

- Note that

$$E(X_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}.$$

- ACF satisfies the difference equation

$$(1 - \phi_1 B - \cdots - \phi_p B^p)\rho_\ell = 0.$$

- parameter estimation:
  conditional least squares method

  – fitted model:

  $$\hat{X}_t = \hat{\phi}_0 + \hat{\phi}_1 X_{t-1} + \cdots + \hat{\phi}_p X_{t-p}$$

  – residual: $\hat{a}_t = X_t - \hat{X}_t$

  – variance estimate:

  $$\hat{\sigma}_a^2 = \frac{\Sigma_{t=p+1}^{T} \hat{a}_t^2}{(T-p)-(p+1)}$$

- Multi-step ahead forecast:

  – one step:

  $$\begin{aligned} \hat{X}_h(1) &= E(X_{h+1} | X_h, X_{h-1}, \cdots) \\ &= \phi_0 + \sum_{i=1}^{p} \phi_i X_{h+1-i} \end{aligned}$$

  – forecast error: $X_{h+1} - \hat{X}_h(1) = a_{h+1}$

– uncertainty in ϕ_i

– two-step:

$$\begin{aligned}\hat{X}_h(2) &= E(X_{h+2}|X_h, X_{h-1}, \cdots) \\ &= \phi_0 + \phi_1 \hat{X}_h(1) + \sum_{i=2}^{p} \phi_i X_{h+2-i}\end{aligned}$$

– forecast error:

$$X_{h+2} - \hat{X}_h(2) = a_{h+2} + \phi_1 a_{h+1}$$

Properties of MA models

An AR model with infinite order:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + a_t.$$

Put constraint on the parameters ϕ_i's.

- For $MA(1)$: $\phi_i = -\theta_1^i$ for $i \geq 1$.
  Note that

$$\begin{aligned} \phi_0 + a_t &= X_t + \theta_1 X_{t-1} + \theta_1^2 X_{t-2} + \cdots, \\ \phi_0 + a_{t-1} &= X_{t-1} + \theta_1 X_{t-2} + \theta_1^2 X_{t-3} + \cdots. \end{aligned}$$

- Then

$$\begin{aligned} X_t &= \phi_0(1 - \theta_1) + a_t - \theta_1 a_{t-1} \\ &= c_0 + a_t - \theta_1 a_{t-1}. \end{aligned}$$

  It is a weighted average of shocks a_t and a_{t-1}.

- $MA(2)$: $X_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$
  $MA(q)$:

$$X_t = c_0 + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

- MA models are always weakly stationary because they are finite linear combinations of a white noise sequence.

- $E(X_t) = c_0$ and
$$Var(X_t) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_a^2.$$

Maximum Likelihood Estimation

- Conditional MLE:
  - Assume that $a_t = 0$ for $t \leq 0$.
  - $a_1 = X_1 - c_0$, a_2 can be computed recursively.
- Exact likelihood method:
  Assume that a_t, $t \leq 0$, as additional parameters.
- Exact estimates are preferred over the conditional one.
- The two types of MLE are close to each other if the sample size is large.

Forecasting:

- short memory
- $MA(1)$:
$$\begin{aligned}
\hat{X}_h(1) &= E(X_{h+1}|X_h, X_{h-1}, \cdots) = c_0 - \theta_1 a_h, \\
e_h(1) &= X_{h+1} - \hat{X}_h(1) = a_{h+1}, \\
\hat{X}_h(2) &= E(X_{h+2}|X_h, X_{h-1}, \cdots) = c_0, \\
e_h(2) &= X_{h+2} - \hat{X}_h(2) = a_{h+2} - \theta_1 a_{h+1}.
\end{aligned}$$

- For $MA(1)$, $Var[e_h(2)] = (1 + \theta_1^2)\sigma_a^2$ and $\hat{X}_h(\ell) = c_0$
- For an $MA(q)$ model, multi-step ahead forecasts go to the mean after the first q steps.

$ARMA$ Models

- For the return series in finance, the chance of using $ARMA$ models is not high.
- The idea of $ARMA$ model is highly relevant in modeling the volatility of asset returns.
- Volatility is not directly measurable, it has some basic properties that are commonly seen in asset returns.
  - volatility clusters:
    Volatility may be high for certain time periods and low for other periods
  - volatility jump:
    Volatility evolves over time in a continuous manner, that is, volatility jumps are rare.
  - volatility stationarity:
    Volatility does not diverge to infinity. (Volatility varies within some fixed range.)
- Let X_t be the log return of a stock at time index t.
  X_t is serially uncorrelated or with minor

lower lag serial correlations but it is dependent.

- A common approach is to model the squared returns X_t^2 by an $ARMA$ model.
- Consider $\mu_t = E(X_t|F_{t-1})$ and

$$\sigma_t^2 = Var(X_t|F_{t-1})$$

where F_{t-1} denotes the information set available at time $t-1$.
Typically, F_{t-1} consists of all linear function of the past returns.

- It is common that the serial dependence in X_t is weak.
Therefore, it is assumed that $X_t = \mu_t + a_t$ and

$$\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{i=1}^{q} \theta_i a_{t-i}.$$

- Note that

$$\sigma_t^2 = Var(X_t|F_{t-1}) = var(a_t|F_{t-1}).$$

We use conditional heteroscedastic models to model the evolution of σ_t^2.

$ARMA$ Models

- $ARMA(1,1)$ model:

$$X_t - \phi_1 X_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1}.$$

If $\phi_1 = \theta_1$, it reduces to a white noise series.

- $ARMA(p,q)$ model:

$$\begin{aligned}&(1 - \phi_1 B - \cdots - \phi_p B^p) X_t \\ &= \phi_0 + (1 - \theta_1 B - \cdots - \theta_q B^q) a_t.\end{aligned}$$

- AR representation:

$$\frac{\phi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B^2 + \cdots = \pi(B).$$

We have

$$\begin{aligned}X_t \;=\; & \frac{\phi_0}{1 - \theta_1 - \cdots - \theta_q} \\ & + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + a_t.\end{aligned}$$

  - Show the dependence of the current return on the past return.
  - When the π coefficient decay to zero as i increases, it is said to be invertible.
  - Consider $X_t = (1 - \theta_1 B) a_t$.
    It is invertible if $|\theta_1| < 1$.

In general, the zero of the equation $\theta(B) = 0$ are greater than unity in modulus.

- MA representation:

$$\frac{\theta(B)}{\phi(B)} = 1 + \psi_1 B + \psi_2 B^2 + \cdots = \psi(B)$$

We have

$$X_t = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p} + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots .$$

  – Show the impact of the past shock on the current return.
  – This representation is useful in computing the variance of forecast error.
    The ℓ-step ahead forecast:

$$\hat{X}_h(\ell) = \mu + \psi_\ell a_h + \psi_{\ell+1} a_{h-1} + \cdots ,$$

    where $\mu = \phi_0/(1 - \phi_1 - \cdots - \phi_p)$.
  – forecast error:

$$e_h(\ell) = a_{h+\ell} + \psi_1 a_{h+\ell-1} + \cdots + \psi_{\ell-1} a_{h+1}.$$

    The variance of forecast error:

$$Var[e_h(\ell)] = (1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2)\sigma_a^2.$$

- Mean reversion of a stationary time series:
  * $\psi_i \to 0$ as $i \to \infty$ by the stationarity
  * $\hat{X}_h(\ell) \to \mu$ as $\ell \to \infty$
  * In the long-term, the return series is expected to approach its mean.

Volatility Modeling

$ARCH$ models:

- The (mean-corrected) asset return $a_t = X_t - \mu_t$ is serially uncorrelated but dependent.
- The dependence of a_t can be described by a simple quadratic function.
- $ARCH(m)$:

  $$\begin{aligned} a_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2, \end{aligned}$$

  where $\{\epsilon_t\}$ are IID with mean zero and variance 1, $\alpha_0 > 0$, and $\alpha_i \geq 0$.
- Large past squared returns imply a large conditional variance σ_t^2 for the return a_t.
- Large returns tend to be followed by another large return.
- Alternative expression:
  $\{\eta_t\}$ is an un-correlated series with zero mean where $\eta_t = a_t^2 - \sigma_t^2$. Then

  $$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2 + \eta_t.$$

It is in the form of an $AR(m)$ model for a_t^2, except that $\{\eta_t\}$ is not an iid series. Recall that $PACF$ is an effective method to determine AR order.

Estimation:

- Fact:

$$P(\cap_{i=1}^n A_i) = P(A_1)\prod_{i=2}^n P(A_i|\cap_{j=1}^{i-1} A_j).$$

- Under the normality assumption, the likelihood function of an $ARCH(m)$ model is

$$\begin{aligned}
&f(a_1,\cdots,a_T|\alpha)\\
&= f(a_T|F_{T-1})\cdots f(a_{m+1}|F_m)f(a_1,\cdots,a_m|\alpha)\\
&\prod_{t=m+1}^{T}\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left[-\frac{a_t^2}{2\sigma_t^2}\right]\times f(a_1,\cdots,a_m|\alpha),
\end{aligned}$$

  where σ_t^2 can be evaluated recursively.

  - Conditional likelihood function: drop the last term
  - We may assume that ϵ_t follows a heavy-tailed distribution such as a standardized Student-t distribution

$GARCH$ Model

- The conditional mean μ_t can be adequately described by an $ARMA$ model.
- $a_t = X_t - \mu_t$.
- $GARCH(m, n)$:

$$\begin{aligned} a_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{n} \beta_m \sigma_{t-j}^2, \end{aligned}$$

  where $\{\epsilon_t\}$ are IID with mean zero and variance 1.

- Alternative expression:
  $\{\eta_t\}$ is an un-correlated series with zero mean where $\eta_t = a_t^2 - \sigma_t^2$. Then

$$a_t^2 = \alpha_0 + \sum_{i=1}^{\max(m,n)} (\alpha_i + \beta_i) a_{t-i}^2 + \eta_t - \sum_{j=1}^{n} \beta_j \eta_{t-j}.$$

  It is in the form of an $ARMA$ model for a_t^2. Application of the $ARMA$ idea to the squared series a_t^2.

Model Identification

- Variance-stabilizing transformations
  - If the variance of the series changes with the level, then a logarithmic transformation will often stabilize the variance.
  - Power transformation
- Degree of differencing
  - When the series is not mean stationary, we have to determine the proper degree of differencing.
    Watch for changing levels and slowly decaying autocorrelations.
  - Examine plots of the time series.
  - Examine plots of the SACF.
  - Examine sample variances of the successive differences.
- Specification of p and q
- Inclusion of a trend parameter: Check for the inclusion of a deterministic trend in the model.

Unit-root Nonstationarity

- In some studies, price series are of interest which tend to be nonstationarity.
- The nonstationarity is mainly caused by the fact that there is no fixed level for price series.
  unit-root time series
- The random walk model:

$$p_t = p_{t-1} + a_t$$

  where p_0 is the starting value of the process and $\{a_t\}$ is a white noise series.
  - If a_t has a symmetric distribution around zero, then conditional on p_{t-1}, p_t has 50−50% to go up and down at random.
  - Treat the random walk model as a special $AR(1)$ model.
  - The coefficient of p_{t-1} is unity, which does not satisfy the weak stationarity condition of $AR(1)$ model.
  - A random walk series is called a unit-root nonstationarity time series.

Forecast under Random Walk Model

- The stock price is not predictable because

$$\hat{p}_h(1) = E(p_{h+1}|p_h, p_{h-1}, \cdots) = p_h.$$

  It is the price of the stock at the forecast origin.

- The 2-step ahead forecast is

$$\begin{aligned} \hat{p}_h(2) &= E(p_{h+2}|p_h, p_{h-1}, \cdots) \\ &= E(p_{h+1} + a_{h+2}|p_h, p_{h-1}, \cdots) = p_h. \end{aligned}$$

- For any forcast horizon $\ell > 0$, we have $\hat{p}_h(\ell) = p_h$.

- MA representation of the random walk model is

$$p_t = a_t + a_{t-1} + \cdots.$$

- The ℓ-step ahead forecast error is

$$e_h(\ell) = a_{h+\ell} + \cdots + a_{h+1},$$

  do that $var[e_h(\ell)] = \ell\sigma_a^2$, which diverges to infinity as $\ell \to \infty$.

- The model is not predictable.

- The unconditional variance of p_t approach infinity as ℓ increases.

- The impact of any past shock a_{t-i} on p_t does not decay over time.
  The series has strong memory as it remembers all of the past shocks.
- Random walk with a drift:

$$p_t = \mu + p_{t-1} + a_t,$$

  where $\mu = E(p_t - p_{t-1})$.
- μ: It represents the time-trend of the log price p_t.
- If we graph p_t against time index t, we have a time trend with slope μ.