Financial Econometrics

Wang Yanchu

Shanghai University of Finance and Economics

September 2024

Outline

- Chapter 2 in Tsay (2005).
- Stochastic process.
- Stationarity.
- Correlation and Autocorrelation functions.
- Autoregressive (AR) models.
- Moving average (MA) models.
- ARMA models.
- Nonstationary process.

Features of Financial Time Series

- Trends and nonstationarity.
- Persistency.
- Comovement and cointegration.
- Autoregressive Conditional Heteroskedasticity (ARCH).
- Other features.

Stochastic Process

- A stochastic process $\{y_t\}$ is simply a sequence of random variables.
- For example:

  1. $y_t \overset{i.i.d.}{\sim} D$, where D is a known distribution.
  2. $y_t = y_{t-1} + \varepsilon_t$, where ε_t follows an $i.i.d.$ distribution.

  This is called random walk.

Stationary Process

- A time series $\{r_t\}$ is called a strictly stationary process if:
  The joint distribution of $\{r_{t1}, \ldots, r_{tk}\}$ is identical to that of $\{r_{t1+t}, \ldots, r_{tk+t}\}$ for any t.
- In other words, the joint distribution of $\{r_{t1}, \ldots, r_{tk}\}$ is invariant under time shift.
  A strong condition and hard to verify empirically.

Stationary Process

- A time series $\{r_t\}$ is called a covariance (weakly) stationary process if:

$$\begin{aligned} E[r_t] &= \mu, \\ Var(r_t) &= \sigma^2, \\ Cov(r_t, r_{t-k}) &= \gamma_k. \end{aligned}$$

  That is, $\{r_t\}$ has a constant mean and the covariance between r_t and r_{t-k} only depends on k, but not on t.

- The covariance between random variables X and Y:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]. \tag{1}$$

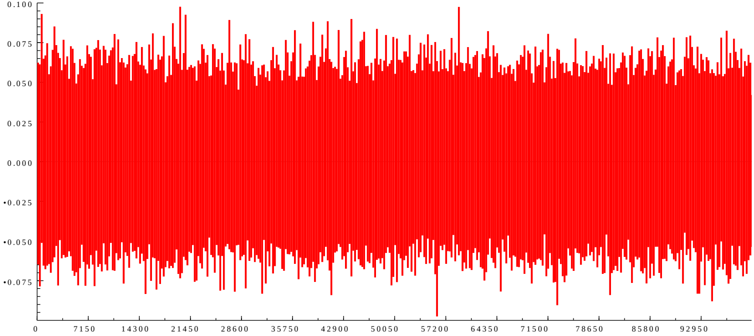
- A stationary process fluctuates around a constant mean and thus enables one to make inferences for its future realizations.
- The covariance γ_k is called lag-k autocovariance:
  $Var(r_t) = \gamma_0$ and $\gamma_k = \gamma_{-k}$.

- A Simulated stationary process

$$r_t \sim N(\mu, \sigma^2),$$

where $\mu = 0.00055$, $\sigma = 0.0224$ and $T = 100,000$.

Figure: A Simulated Stationary Process



IID, White Noise and the Information Set

- A sequence of random variables u_t is said to be iid$(0, \sigma^2)$ if u_t are independent and identically distributed with $E(u_t) = 0$ and $var(u_t) = \sigma^2$.
- A process $\{u_t\}$ is known as a white noise process if:
  1. $E[u_t] = 0$, for any t.
  2. $E[u_t^2] = Var(u_t) = \sigma^2 < \infty$, for any t.
  3. $E(u_t u_s) = 0$ for $\forall t \neq s$.
- A white noise process consists a sequence of uncorrelated but can be dependent random variables.
- It is different from the i.i.d. sequence.

Martingale Difference Sequence

- The information set at time t is denoted $\mathcal{F}_t$ or I_t. It includes realizations of any relevant variable which are known on or before time t.
  It is obvious that $\mathcal{F}_{t-1} \subset \mathcal{F}_t$.
- We call a stochastic process $\{u_t\}$ a Martingale Difference Sequence (MDS) if:
  1. $u_t \in \mathcal{F}_t$.
  2. $E[u_t | \mathcal{F}_{t-1}] = 0$.

Martingale Difference Sequence

- Proposition I: If u_t is a MDS and $E[u_t^2] = \sigma^2$, then u_t is white noise.
  Proof: We can show that
  $E[u_t] = E[E[u_t|\mathcal{F}_{t-1}]] = 0$, and
  $E[u_t u_{t+j}] = E[E[u_t u_{t+j}|\mathcal{F}_t]] = E[u_t E[u_{t+j}|\mathcal{F}_t]] = 0$.
- However, the converse is not true. Consider the following sequence:

  $$u_t = \varepsilon_t + \theta\varepsilon_{t-1}\varepsilon_{t-2},$$

  where $\varepsilon_t \sim N(0,1)$.
  It is easy to show that u_t is white noise, yet
  $E[u_t|\mathcal{F}_{t-1}] = \theta\varepsilon_{t-1}\varepsilon_{t-2} \neq 0$.
- Proposition II: *iid* $\Rightarrow$ *MDS* $\Rightarrow$ *white noise*.

- Example. Consider an autoregressive conditional heteroskedastic (ARCH) process:

$$r_t = \sigma_t \varepsilon_t, \tag{2}$$

where

$$\begin{aligned} \varepsilon_t &\sim iid(0,1), \\ \sigma_t^2 &= \omega + \alpha r_{t-1}^2, \end{aligned}$$

with

$$\omega > 0,\ 0 \leqslant \alpha < 1.$$

Martingale Difference Sequence

- It is obvious that the process $\{r_t\}$ is a MDS if we let $\mathcal{F}_t = \{r_t, r_{t-1}, \cdots\}$.
- To see this,

$$
\begin{aligned}
E(r_t|\mathcal{F}_{t-1}) &= E(\sigma_t \varepsilon_t|\mathcal{F}_{t-1}) \\
&= \sigma_t E(\varepsilon_t|\mathcal{F}_{t-1}) \\
&= 0.
\end{aligned}
$$

- The conditional variance of r_t

$$
\begin{aligned}
Var(r_t|\mathcal{F}_{t-1}) &= E(r_t^2|\mathcal{F}_{t-1}) \\
&= \sigma_t^2 E(\varepsilon_t^2|\mathcal{F}_{t-1}) \\
&= \sigma_t^2.
\end{aligned}
$$

  The conditional variance is time-varying, thus it is called heteroskedastic.
- Is $\{r_t\}$ a white noise process?

- The lag operator delays the time index by one period:

$$LX_t = X_{t-1}, \tag{3}$$

  and thus

$$L^2 X_t = L(LX_t) = LX_{t-1} = X_{t-2}.$$

- We may define a polynormial function of L (called lag polynormial) such that

$$\alpha(L) = \alpha_0 + \alpha_1 L + \cdots + \alpha_p L^p,$$

  then

$$\alpha(L)X_t = \alpha_0 X_t + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p}.$$

The Lag Operator

- The value p can be infinity and in this case

$$\alpha(L) = \sum_{j=0}^{\infty} \alpha_j L^j,$$

and

$$\alpha(L) X_t = \sum_{j=0}^{\infty} \alpha_j X_{t-j}.$$

- The process may not be well defined since the series can be divergent without proper constraints on $\alpha_j s$.
- We say $\{\alpha_j\}$ is absolute summable if $\sum_{j=0}^{\infty} |\alpha_j| < \infty$.
- Proposition: If $\{X_t\}$ is convariance stationary and $\{\alpha_j\}$ is absolute summable, then $\alpha(L)X_t$ is also covariance stationary.

The Filters

- Polynormial functions of the lag operator are called filters.
- Product of filters. Let $\alpha(L) = \sum_{j=0}^{\infty} \alpha_j L^j$ and $\beta(L) = \sum_{j=0}^{\infty} \beta_j L^j$, then

$$\alpha(L)\beta(L) = \delta(L) = \sum_{j=0}^{\infty} \delta_j L^j,$$

where

$$\delta_j = \sum_{k=0}^{j} \alpha_k \beta_{j-k}.$$

The Filters

- Inverse of a filter. If $\alpha(L)\beta(L) = 1$, we say $\beta(L)$ is the inverse of $\alpha(L)$. Since it is unique, we can write it as $\alpha(L)^{-1}$.

$$\alpha(L)^{-1} = \frac{1}{\alpha(L)}.$$

- Theorem: The inverse of $\alpha(L)$ exists if and only if $\alpha_0 \neq 0$.
- Solving for inverse. Let $\beta(L) = \alpha(L)^{-1}$, since $\alpha(L)\beta(L) = 1 = 1 + 0 \cdot L + 0 \cdot L^2 + \cdots$, we have

$$\begin{aligned}
\alpha_0\beta_0 &= 1, \\
\alpha_0\beta_1 + \alpha_1\beta_0 &= 0, \\
\alpha_0\beta_2 + \alpha_1\beta_1 + \alpha_2\beta_0 &= 0, \text{ and so on.}
\end{aligned}$$

- The bootstrap method, start from solving β_0 and so on.

Correlation and Autocorrelation Function

- Correlation between random variables X and Y:

$$\rho_{x,y} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{E(X - \mu_x)^2 E(Y - \mu_y)^2}}. \quad (4)$$

- Correlation measures the linear dependence between X and Y.
- By definition, $-1 \leqslant \rho_{x,y} \leqslant 1$.
- X and Y are uncorrelated if $\rho_{x,y} = 0$.
- Autocorrelation Function:

$$\rho_k = \frac{Cov(r_t, r_{t-k})}{\sqrt{Var(r_t)Var(r_{t-k})}} = \frac{\gamma_k}{\gamma_0}. \quad (5)$$

- Note ρ_k is a function of k only under the stationary condition.

Correlation and Autocorrelation Function

- For estimation, let $\bar{r}$ be the sample mean of $\{r_t\}$:

$$\bar{r} = \frac{1}{T} \sum_{t=1}^{T} r_t. \tag{6}$$

- The estimate for $\widehat{\rho}_k$ is

$$\widehat{\rho}_k = \frac{\sum_{t=k+1}^{T} (r_t - \bar{r})(r_{t-k} - \bar{r})}{\sum_{t=1}^{T} (r_t - \bar{r})^2}, \ 0 \leqslant k < T - 1. \tag{7}$$

- The statistics $\widehat{\rho}_1, \widehat{\rho}_2, \ldots$ defined above is called the sample autocorrelation function (ACF) of $\{r_t\}$.

Testing Individual ACF

- If $\{r_t\}$ is a independent and identically distributed (IID) process, then $\sqrt{T}\widehat{\rho}_1 \sim N(0, 1)$.
- If $\{r_t\}$ is a stationary process satisfying $r_t = \mu + \sum_{i=0}^{q} \psi_i a_{t-i}$ where $\psi_0 = 1$ and $a \sim N(0, 1)$, then $\widehat{\rho}_k$ is asymptotically normal with mean zero and variance $(1 + 2\sum_{i=1}^{k-1} \widehat{\rho}_i^2)/T$. The t-ratio is

$$t\text{-ratio} = \frac{\widehat{\rho}_k}{\sqrt{(1 + 2\sum_{i=1}^{k-1} \widehat{\rho}_i^2)/T}}. \tag{8}$$

- $H_0 : \rho_k = 0$.
- We reject H_0 if $|t - raio| > Z_{\alpha/2}$ where $Z_{\alpha/2}$ is the $100(1 - \alpha/2)th$ percentile of the standard normal distribution.
- In finite sample, $\widehat{\rho}_k$ is a biased estimator of ρ_k. The bias is on the order of $1/T$.

Testing Joint ACF

- $H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0.$
- The Box and Pierce statistic:

$$Q^*(m) = T \sum_{k=1}^{m} \widehat{\rho}_k^2. \tag{9}$$

Under the null, $Q^*(m)$ is a χ^2 distributed with m degrees of freedom.

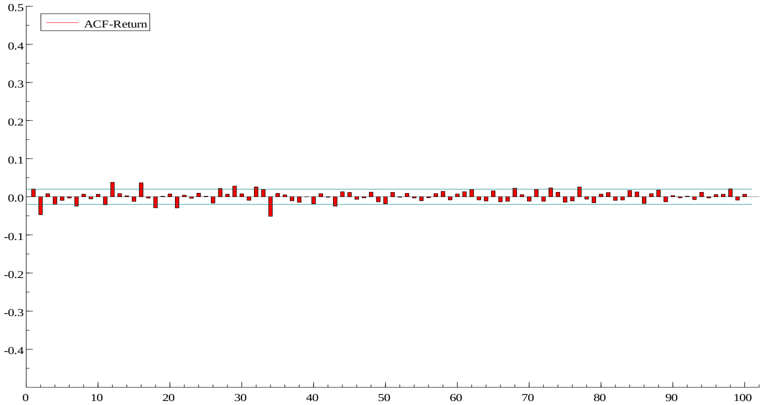
- For finite sample, Ljung and Box provide a better statistic than (9):

$$Q(m) = T(T+2) \sum_{k=1}^{m} \frac{\widehat{\rho}_k^2}{T-k}. \tag{10}$$

Under the null, $Q(m)$ is a χ^2 distributed with m degrees of freedom.

Empirical Example

- Again use the daily closing price of SP500 Index from 1970/01/05 $\sim$ 2009/07/31.
- The ACF plot up to lag 100 for log returns:



Empirical Example

- Most of the autocorrelations ρ_k, $k = 1, 2, \cdots, 100$, are not significant.
- However, several autocorrelations, i.e., ρ_2, ρ_7, ρ_{12} and others, are significant.
- The Ljung-Box statistics:
  1. $m = 10$, $Q(10) = 39.09^{**}(0.00)$.
  2. $m = 100$, $Q(100) = 248.81^{**}(0.00)$.
- The selection of m will affect the performance of the $Q(m)$ test.
- Simulation studies suggest that $m \approx \ln(T)$ will provide better power performance.
- The results violate the market efficiency hypothesis.

Autoregressive (AR) Models

- The reject of H_0 that $\rho_1 = \rho_2 = \cdots = \rho_m = 0$ in the previous example may indicate serial dependence in $\{r_t\}$.
- We can model r_t in autoregressive (AR) models.
- Consider a simple $AR(1)$ model:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \varepsilon_t. \tag{11}$$

  where ε_t is assumed to be a *white noise* series with $E[\varepsilon_t] = 0$ and $Var(\varepsilon_t) = \sigma_\varepsilon^2$. We call ε_t as the innovation or shock at time t.
- It can be written as: $(1 - \phi_1 L) r_t = \phi_0 + \varepsilon_t$.
- In general, an $AR(p)$ model is

$$\begin{aligned} r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \ldots & \quad + \phi_p r_{t-p} + \varepsilon_t, \qquad (12) \\ or\ (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) r_t & = \phi_0 + \varepsilon_t. \end{aligned}$$

Autoregressive (AR) Models

- In $AR(1)$ model, under the assumption that $E[\varepsilon_t] = 0$ and $Var(\varepsilon_t) = \sigma_\varepsilon^2$, we have

$$E(r_t | r_{t-1}) = \phi_0 + \phi_1 r_{t-1}, \tag{13}$$

and

$$Var(r_t | r_{t-1}) = \sigma_\varepsilon^2. \tag{14}$$

- Assume r_t is a stationary process, then $E(r_t) = \mu$, $Var(r_t) = \gamma_0$ and $Cov(r_t, r_{t-j}) = \gamma_j$.
- From (13), we have

$$\mu = \frac{\phi_0}{1 - \phi_1}. \tag{15}$$

- The mean of r_t exists if $\phi_1 \neq 1$ and $E(r_t) = 0$ iff $\phi_0 = 0$.

Autoregressive (AR) Models

- By applying $\phi_0 = \mu(1 - \phi_1)$, we can rewrite (11) as
$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \varepsilon_t. \tag{16}$$
- By repeated substitutions, it is shown that
$$r_t - \mu = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}. \tag{17}$$
- Taking on variance of (16), we have
$$Var(r_t) = \phi_1^2 Var(r_{t-1}) + \sigma_\varepsilon^2. \tag{18}$$
- Under the stationary assumption that $Var(r_t) = Var(r_{t-1}) = \gamma_0$, we have
$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}. \tag{19}$$
- It implies that a necessary and sufficient condition for stationarity of an $AR(1)$ process is that $|\phi_1| < 1$.

Autoregressive (AR) Models

- The ACF of $AR(1)$ process

  1. Multiplying ε_t on both sides of (16) and taking expectation:

$$E[(r_t - \mu)\varepsilon_t] = E[\phi_1(r_{t-1} - \mu)\varepsilon_t] + E(\varepsilon_t^2) = \sigma_\varepsilon^2. \tag{20}$$

  2. Multiplying $(r_{t\ \ k} - \mu)$ on both sides of (16), taking expectation and using result from (20), we have

$$\gamma_k = \begin{cases} \phi_1\gamma_1 + \sigma_\varepsilon^2, & \text{if } k = 0. \\ \phi_1\gamma_{k-1}, & \text{if } k > 0. \end{cases}$$

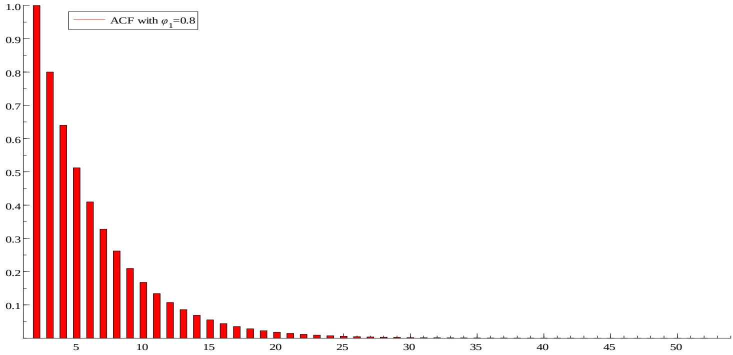
- For a stationary $AR(1)$ model, we obtain

$$\rho_k = \phi_1\rho_{k-1}, \text{ for } k > 1. \tag{21}$$

- Since $\rho_0 = 1$, it indicates that $\rho_k = \phi_1^k$. The ACF decays exponentially with rate ϕ_1.

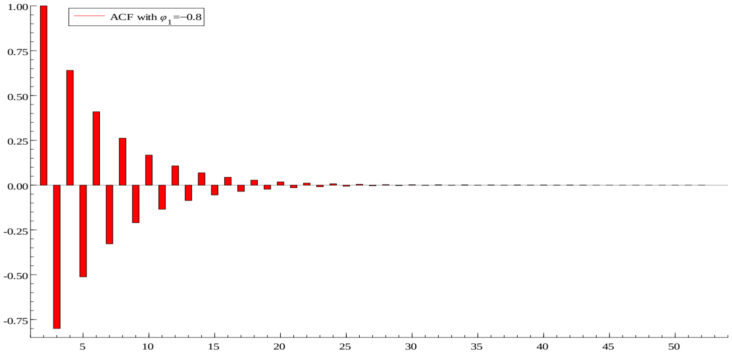
Autoregressive (AR) Models

- Examples:
  1. ACF plot with $\phi_1 = 0.8$.



- Examples:

  2. ACF plot with $\phi_1 = -0.8$.



Autoregressive (AR) Models

- $AR(2)$ model

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \varepsilon_t \tag{22}$$

- Assume stationarity, the mean of r_t is obtained

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}, \tag{23}$$

  provided that $\phi_1 + \phi_2 \neq 1$.

- Using $\phi_0 = (1 - \phi_1 - \phi_2)\mu$, rewrite (22) as

$$(r_t - \mu) = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + \varepsilon_t \tag{24}$$

- Multiplying both sides of (24) by $(r_{t-k} - \mu)$ and taking expectation, we get

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \text{ for } k > 2. \tag{25}$$

Autoregressive (AR) Models

- Dividing (25) by γ_0, we obtain

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \text{ for } k > 0. \tag{26}$$

- In particular, $\rho_0 = 1$ and $\rho_1 = \frac{\phi_1}{1-\phi_2}$.
- Eq. (26) says that a stationary $AR(2)$ process satisfies a second-order difference equation:

$$(1 - \phi_1 L - \phi_2 L^2)\rho_k = 0, \tag{27}$$

where L is the lag operator such that $L\rho_k = \rho_{k-1}$ and $L^2 \rho_k = \rho_{k-2}$.

Autoregressive (AR) Models

- Corresponding to (27), we have a second-order polynomial equation:

$$1 - \phi_1 x - \phi_2 x^2 = 0. \tag{28}$$

- Solutions of this equation are

$$x = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}. \tag{29}$$

- The inverses of the two solutions, $1/x_1$ and $1/x_2$, are called characteristic roots of $AR(2)$ model. Denote the two solutions by ω_1 and ω_2, if $\phi_1^2 + 4\phi_2 > 0$, we can rewrite (28) as

$$(1 - \omega_1 L)(1 - \omega_2 L) = 0.$$

- The stationarity condition of an $AR(2)$ time series is that $|\omega_1| < 1$ and $|\omega_2| < 1$. This ensures that ρ_k decays to zero as k increases.

Autoregressive (AR) Models

- In eq. (29), if $\phi_1^2 + 4\phi_2 < 0$, then ω_1 and ω_2 are complex numbers.
  1. The plot of the ACF of r_t would show a damping sine and cosine waves.
  2. Complex characteristic roots give rise to the behavior of business cycles.
  3. For the AR(2) model, the length of the cycle is

$$k = \frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]}. \tag{30}$$

Autoregressive (AR) Models

- $AR(p)$ process

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + \varepsilon_t. \tag{31}$$

- Under the stationarity condition, the mean of r_t is

$$E(r_t) = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p},$$

  provided that $\phi_1 + \cdots + \phi_p \neq 1$.
- The $p-$order polynomial equation is

$$1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p = 0.$$

- Stationarity requires that for all characteristic roots, $|\omega_i| < 1$, $i = 1, 2, ..., p$.
- Under stationarity, the ACF satisfies

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)\rho_\ell = 0 \text{ for } \ell > 0. \tag{32}$$

Autoregressive (AR) Models

- Identifying the order of AR models
- The partial autocorrelation function (PACF)
  Consider the following regressions:

$$
\begin{aligned}
r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + \varepsilon_{1t}, \\
r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + \varepsilon_{2t}, \\
r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + \varepsilon_{3t}, \\
&\vdots
\end{aligned}
$$

- The estimate of $\widehat{\phi}_{1,1}$ is called lag-1 PACF of r_t, $\widehat{\phi}_{2,2}$ is called lag-2 PACF of r_t and so on.
- For an $AR(p)$ model, the lag-p PACF $\widehat{\phi}_{p,p}$ should not be zero, but any $\widehat{\phi}_{j,j}$ should be close to zero for all $j > p$.

- Information criteria

  1. Akaike information criteria (AIC)

  $$AIC = \frac{-2}{T}\ln(\text{likelihood}) + \frac{2}{T} \times (\text{number of parameters}). \quad (33)$$

  For a Gaussian $AR(p)$ model, AIC becomes

  $$AIC(p) = \ln(\widehat{\sigma}_\varepsilon^2) + \frac{2p}{T}.$$

  2. Bayesian information criteria (BIC)

  $$BIC = \frac{-2}{T}\ln(\text{likelihood}) + \frac{p}{T} \times \ln T.$$

  For a Gaussian $AR(p)$ model, BIC becomes

  $$BIC(p) = \ln(\widehat{\sigma}_\varepsilon^2) + \frac{p\ln T}{T}.$$

- Select the order of p that minimizes the AIC or the BIC.

Autoregressive (AR) Models

- Parameters can be estimated by the OLS method. For an $AR(p)$ model, the fitted model is

$$\widehat{r}_t = \widehat{\phi}_0 + \widehat{\phi}_1 r_{t-1} + \cdots + \widehat{\phi}_p r_{t-p}, \tag{34}$$

and the residual is

$$\varepsilon_t = r_t - \widehat{r}_t. \tag{35}$$

- We can obtain the estimator for σ_ε^2 as from

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{t=p+1}^{T} \widehat{\varepsilon}_t^2}{T - p - 1}. \tag{36}$$

Autoregressive (AR) Models

- Forecasting under $AR(P)$ model.
  The point forecast of r_{h+1} given $I_h = \{r_h, r_{h-1}, ...\}$ is the conditional expectation

$$\widehat{r}_h(1) = E(r_{h+1}|I_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i}. \tag{37}$$

  The associated forecast error is

$$e_h(1) = r_{h+1} - \widehat{r}_h(1) = \varepsilon_{h+1} \tag{38}$$

- In general, the k-step ahead forecast for $r_h(k)$ is

$$\widehat{r}_h(k) = \phi_0 + \sum_{i=1}^{p} \phi_i \widehat{r}_h(k-i). \tag{39}$$

  The $\widehat{r}_h(k)$ forecast can be computed recursively using forecasts $\widehat{r}_h(i)$ for $i = 1, ..., k-1$.

Autoregressive (AR) Models

- For a stationary $AR(p)$ model, $\widehat{r}_h(k)$ converges to $E(r_t)$ as $k \longrightarrow \infty$.
- This is called the mean reversion.
- The speed of mean reversion is measured by the half-life defined as

$$k = \ln(0.5/|\phi_1|).$$

- The variance of forecast also approaches the unconditional variance of r_t.

Moving Average (MA) Models

- Consider an AR model with an infinite order $p \longrightarrow \infty$

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + \varepsilon_t, \tag{40}$$

  where $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \sigma^2$.

- Assume that the coefficients ϕ_i are determined by a finite number of parameters. A special case is to think that $\phi_i = -\theta_1^i$ for $i \geqslant 1$ so that

$$r_t = \phi_0 - \theta_1 r_{t-1} - \theta_1^2 r_{t-2} - \cdots + \varepsilon_t. \tag{41}$$

  The contribution of r_{t-i} to r_t decays exponentially as i increases.

- Rearrange (41) to get

$$r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + \cdots = \phi_0 + \varepsilon_t. \tag{42}$$

Moving Average (MA) Models

- Similarly, the model for r_{t-1} is

$$r_{t-1} + \theta_1 r_{t-2} + \theta_1^2 r_{t-3} + \cdots = \phi_0 + \varepsilon_{t-1}. \tag{43}$$

- Let $(42) - (43) \times \theta_1$, we obtain

$$r_t = \phi_0(1 - \theta_1) + \varepsilon_t - \theta_1 \varepsilon_{t-1} = c_0 + (1 - \theta_1 L)\varepsilon_t, \tag{44}$$

  where $c_0 = \phi_0(1 - \theta_1)$ and L is the lag operator.
- (44) is called the $MA(1)$ model.
- In general, an $MA(q)$ model is defined as

$$r_t = c_0 + (1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q)\varepsilon_t, \tag{45}$$

  where $q \geqslant 1$.

Moving Average (MA) Models

- MA models are always stationary. For example, for $MA(1)$ models in (44), we have:

$$E(r_t) = c_0,$$

and

$$Var(r_t) = (1 + \theta_1^2)\sigma^2.$$

- For $MA(q)$ models:

$$Var(r_t) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2. \tag{46}$$

Moving Average (MA) Models

- Autocorrelation Function.
  Consider $MA(1)$ model of (44), assume $c_0 = 0$. Multiply r_{t-k} on both sides

$$r_{t-k} r_t = r_{t-k}\varepsilon_t - r_{t-k}\theta_1\varepsilon_{t-1}.$$

  Take expectation, we can see that

$$\gamma_1 = -\theta_1\sigma^2, \text{ and } r_k = 0 \text{ for } k > 1,$$

  which implies that

$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2} \text{ and } \rho_k = 0 \text{ for } k > 1.$$

- In general, for $MA(q)$ models, $\rho_k = 0$ for all $k > q$. An $MA(q)$ series is a "finite memory" model.
- The "finite memory" property can be used to identify the MA order.

Moving Average (MA) Models

- *MA* models can be estimated by Maximum likelihood.
  1. Treat the initial innovation $\varepsilon_0 = 0$. *Conditional likelihood method.*
  2. Treat the initial innovation ε_0 as a parameter to be estimated.

  Exact likelihood method.
- Example. Consider a following $MA(1)$ model as from (44)

$$r_t = \phi_0(1-\theta_1) + a_t - \theta_1\varepsilon_{t-1} = c_0 + (1-\theta_1 L)\varepsilon_t.$$

If let $\varepsilon_0 = 0$, we have $\varepsilon_1 = r_1 - c_0$, $\varepsilon_2 = r_2 - c_0 + \theta_1\varepsilon_1, \cdots$ and $\boldsymbol{\theta} = (c_0, \theta_1, \sigma^2)'$.
Assume that $\varepsilon_t \sim N(0, \sigma^2)$, then

$$f(\varepsilon_t | I_t) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{\varepsilon_t^2}{\sigma^2}}.$$

Moving Average (MA) Models

- If y is a function of x, we have $f(x) = \frac{dy}{dx} f(y)$.
- Use this relation and since $f(r_t) = \frac{d\varepsilon_t}{dr_t} f(\varepsilon_t)$, we get

$$f(r_t; \boldsymbol{\theta} | I_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{(r_t - c_0 + \theta_1 \varepsilon_{t-1})^2}{\sigma^2}]. \tag{47}$$

- The joint density of $\{r_t\}$ is

$$f(r_1, r_2, \cdots, r_t; \boldsymbol{\theta}) = f(r_1; \boldsymbol{\theta}) \prod_{t=2}^{T} f(r_t; \boldsymbol{\theta} | I_t). \tag{48}$$

Moving Average (MA) Models

- Take the log of (48), we obtain

$$\ell(r_1, r_2, \cdots, r_t; \boldsymbol{\theta}) = \ln f(r_1; \boldsymbol{\theta}) - \frac{1}{2}\sum_{t=2}^{T}\Big[\ln(2\pi) + \ln(\sigma^2) + \frac{(r_t - c_0 + \theta_1 \varepsilon_{t-1})^2}{\sigma^2}\Big]. \quad (49)$$

- $\boldsymbol{\theta}$ is chosen to maximize (49).
- If we treat ε_0 is an additional parameter to be estimated, then $\boldsymbol{\theta} = (\varepsilon_0, c_0, \theta_1, \sigma^2)$ and the same method applies.
- If the sample size is large, then the two approaches lead to results that are close to each other.
- ACF can be used to identify the order of a MA model.

ARMA Models

- *ARMA* models: a combination of *AR* and *MA* models.
- *ARMA* models are useful in volatility modelling, i.e. the GARCH model.
- An $ARMA(p,q)$ model is

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}. \tag{50}$$

- (50) can be written as

$$(1 - \sum_{i=1}^{p} \phi_i L^i) r_t = \phi_0 + (1 - \sum_{j=1}^{q} \theta_j L^j) \varepsilon_t.$$

- The properties of ARMA models under stationary conditions are similar to the corresponding AR models.
- Identification of ARMA models.

Nonstationary Process

- Many financial series are not stationary, like interest rate, exchange rate, prices etc.
- The nonstationary is due to the fact that there is no fixed level for the series.
- Consider the following models:

$$p_t = p_{t-1} + e_t \tag{51}$$

  where e_t is a stationary process.
- We call such a process as a unit root nonstationary time series.
- We may define $x_t = p_t - p_{t-1}$, thus x_t becomes a stationary process. x_t is called the first order difference of p_t.

ARIMA Process

- A time series $\{y_t\}$ is an $ARIMA(p, 1, q)$ process if

$$c_t = y_t - y_{t-1} = (1 - L)y_t \tag{52}$$

  follows an $ARMA(p, q)$ stationary process.

- We may make a further differencing series if c_t is not stationary, define

$$s_t = c_t - c_{t-1} = (1 - L)c_t. \tag{53}$$

- If s_t is stationary, we say y_t follows an $ARMA(p, 2, q)$ process, and so on.

Vector Autoregressive Models

- Consider
$$Y_t = \begin{bmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{mt} \end{bmatrix}.$$
- We say Y_t is stationary if Y_t has a constant mean and the covariance $Cov(Y_t, Y_{t-k})$ only depends on k, not on t.
- Let $\Gamma(k) = Cov(Y_t, Y_{t-k}) = E[(Y_t - \mu)(Y_{t-k} - \mu)']$,
$$\Gamma(k) = \begin{pmatrix} \gamma_{11}(k) & \gamma_{12}(k) & \dots & \gamma_{1m}(k) \\ \gamma_{21}(k) & \gamma_{22}(k) & \dots & \gamma_{2m}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1}(k) & \gamma_{m2}(k) & \dots & \gamma_{mm}(k) \end{pmatrix}$$
- $\Gamma(k)$ is called the autocovariance matrix at lag k.

Vector Autoregressive Models

- It is important to note that $\Gamma(k) \neq \Gamma(-k)$, but $\Gamma(k) = \Gamma(-k)'$. This is because

$$\gamma_{ij}(k) = E[(Y_{it} - \mu_i)(Y_{jt-k} - \mu_j)] = E[(Y_{jt-k} - \mu_j)(Y_{it} - \mu_i)] = \gamma_{ji}(-$$

- Long-run covariance matrix. For a stationary vector time series Y_t, define

$$\Omega = \sum_{k=-\infty}^{k=\infty} \Gamma(k) = \Gamma(0) + \sum_{k=1}^{k=\infty} [\Gamma(k) + \Gamma(k)'].$$

- A consistent estimator is given by

$$\widehat{\Omega} = \widehat{\Gamma}(0) + \sum_{k=1}^{M} (1 - \frac{k}{M+1})(\widehat{\Gamma}(k) + \widehat{\Gamma}(k)'),$$

where $\widehat{\Gamma}(k) = \frac{1}{T} \sum_{t=k+1}^{T} (Y_t - \overline{Y})(Y_{t-k} - \overline{Y})$.

- The number of M is called the truncation order and we require that $M \to \infty$ as $T \to \infty$ with $M/T \to 0$.

Examples of Vector Time Series Process

- Vector white noise

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{mt} \end{bmatrix},$$

where $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_t') = \Omega$, and $E(\varepsilon_t \varepsilon_s') = 0$ for $t \neq s$.

- $VAR(1)$ process

$$Y_t = \Phi Y_{t-1} + \varepsilon_t,$$

where Φ is a $m \times m$ matrix and ε_t is white noise. Stationarity requires that the eigenvalues of Φ be less than 1 in absolute value.

Examples of Vector Time Series Process

- $VMA(1)$ Process

$$Y_t = \varepsilon_t + \Theta \varepsilon_{t-1},$$

where Θ is a $m \times m$ matrix. $VMA(1)$ is always stationary.

- $VARMA(p, q)$

$$Y_t = \Phi_1 Y_{t-1} + ... + \Phi_p Y_{t-p} + \varepsilon_t + \Theta_1 \varepsilon_{t-1} + ... + \Theta_q \varepsilon_{t-q}.$$

Stationary requires that all roots of the determinantal equation

$$\det(I - \Phi_1 z - ... - \Phi_p z^p) = 0$$

be greater than 1 in absolute value.

Granger Causality

- Let
$$Y_t = \begin{bmatrix} X_t \\ Z_t \end{bmatrix}.$$
- Definition: Z_t Granger causes X_t if Z_t helps to forecast X_t, given past X_t.
- This implies a restriction on the *VAR* representation of Y_t. Consider
$$\begin{aligned} X_t &= a(L)X_{t-1} + b(L)Z_{t-1} + \eta_t, \\ Z_t &= c(L)X_{t-1} + d(L)Z_{t-1} + \xi_t. \end{aligned}$$
- The absence of Granger Causality from Z_t to X_t implies that $b(L) = 0$. If we rewrite the above Y_t as
$$B(L)Y_t = \varepsilon_t,$$
then it must be the case that $B(L)$ is a lower triangular matrix.
- It is important to note that Granger Causality does not have the conventional meaning of cause.

Random Walk

- A stochastic process $\{y_t\}$ is known as a random walk if

$$y_t = y_{t-1} + \varepsilon_t, \tag{54}$$

where $\varepsilon \sim iid(0, \sigma^2)$.

- A random variable is called a martingale if $E[x_T | I_t] = x_t$ for $T > t$.
- We note that
  1. $E[y_{t+h} | I_t] = y_t$, i.e. it is a martingale.
  2. $Var(y_t) = t\sigma^2$
  3. $Cov(y_t, y_{t-k}) = (t-k)\sigma^2$. Thus the autocorrelation $\rho_k = \frac{t-k}{t}$.

Unit Root Tests

- Dicky and Fuller (1979) note that

$$y_t = \phi y_{t-1} + \varepsilon_t,$$

can be transformed to

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t,$$

where $\Delta y_t = y_t - y_{t\ 1}$ and $\gamma = \phi - 1$.

- The unit root test can be conducted for the null, $H_0 : \gamma = 0$ against the alternative $H_1 : \gamma < 0$.
- We may regress Δy_t on y_{t-1}, and the t-value is computed in the usual way.
- If the distribution of $\widehat{\gamma}$ were standard normal (under the null), this would be a very simple case.
- Unfortunately, it is not the case since under the null, $y_{t\ 1}$ is a unit root and the variance is growing rapidly as the number of observations increases.
- Dickey and Fuller derive the distribution rather than the standard

Unit Root Tests

- Dickey and Fuller considered three separate specifications for their test

$$
\begin{aligned}
\Delta y_t &= \gamma y_{t-1} + \varepsilon_t, \\
\Delta y_t &= \phi_0 + \gamma y_{t-1} + \varepsilon_t, \\
\Delta y_t &= \phi_0 + \delta t + \gamma y_{t-1} + \varepsilon_t,
\end{aligned}
$$

  which correspond to a unit root, a unit root with a linear time trend, and a unit root with a quadratic time trend.
- The unit root test can be conducted for the null, $H_0 : \gamma = 0$ against the alternative $H_1 : \gamma < 0$, and the null is rejected if $\widehat{\gamma}$ is sufficiently negative, which is equivalent to $\widehat{\phi}$ being significantly less than 1 in the original specification.
- The critical value under the DF distribution with $T = 200$:

	No trend	linear	quadratic
10%	-1.66	-2.56	-3.99
5%	-1.99	-2.87	-3.42
1%	-2.63	-3.49	-3.13

Unit Root Tests

- Augmented Dicky-Fuller (ADF) test generalize

$$
\begin{aligned}
\Delta y_t &= \gamma y_{t-1} + \sum_{p=1}^{P} \phi_p y_{t-p} + \varepsilon_t, \\
\Delta y_t &= \phi_0 + \gamma y_{t-1} + \sum_{p=1}^{P} \phi_p y_{t-p} + \varepsilon_t, \\
\Delta y_t &= \phi_0 + \delta t + \gamma y_{t-1} + \sum_{p=1}^{P} \phi_p y_{t-p} + \varepsilon_t.
\end{aligned}
$$

- Neither the null and alternative hypotheses nor the critical values are changed by the inclusion of lagged dependent variables.
- The intuition behind this result stems from the observation that the $y_{t\ \ p}$ are "less integrated" than y_t and so are asymptotically less informative.

Long Memory Process

- Consider the following model

$$r_t = \rho_1 r_{t-1} + \varepsilon_t \tag{55}$$

- If $|\rho_1| < 1$, the model is an $AR(1)$ model, the ACF decays exponentially, no matter how $|\rho|$ is close to 1.
- If $|\rho_1| = 1$, $\{r_t\}$ is a unit root process, the ACF almost never decays out (especially when $t \gg k$).
- Some time series have ACFs decay slowly to zero in a polynormial rate. These processes are called long memory processes.

Long Memory Process

- One usual way to model the long memory process is to use the fractional differenced process:

$$(1 - L)^d x_t = \varepsilon_t \tag{56}$$

  where ε_t is a stationary process.

- Some properties:
  1. If $d < 0.5$, then x_t is a weakly stationary process and has the infinite *MA* representation.
  2. If $d > -0.5$, then x_t is invertible and has the infinite *AR* representation.
  3. For $-0.5 < d < 0.5$, $\rho_k \sim k^{2d-1}$.

- If $(1 - L)^d x_t$ follows an $ARMA(p, q)$ process, we say that x_t follows an $ARFIMA(p, q)$ process.

Long Memory Process

- We can use the binomial theorem for the noninteger powers:

$$(1-L)^d = \sum_{i=0}^{\infty} \frac{\Gamma(i-d)}{\Gamma(-d)\Gamma(i+1)} L^i, \tag{57}$$

where $\Gamma(\cdot)$ is the Gamma function

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha} e^{-x} dx. \tag{58}$$

Regression Models with Time Series Errors

- Consider the following regression model

$$r_{it} = \alpha + \beta r_{mt} + \varepsilon_{it}. \tag{59}$$

- If $\{\varepsilon_t\}$ is a white noise series, we may use the Least-Square method to obtain consistent estimates of the parameter vector, $\boldsymbol{\theta}$.

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{T} \varepsilon_{it}(\boldsymbol{\theta})^2. \tag{60}$$

- However, if $\{\varepsilon_t\}$ is serially correlated, the LS estimates are no longer consistent.
- The regression model with time series errors is common in finance.

Regression Models with Time Series Errors

- Treatment with time series errors:
  1. Fit the linear regression model and check the serial correlations in residuals.
  2. If the residual is unit-root nonstationary, take the first difference of both independent and dependent variables. Go to step 1.
  3. If the residual series appears to be stationary, identify an ARMA model for the residuals and modify the linear model accordingly.
  4. Perform a joint estimation via the ML method.

Exercises 2 (Due Date: two weeks later)

- Download the file "sp500_1973_2008.xls" from the course homepage. The data contains daily log returns of SP500 index including dividends (vwretd) and excluding dividends (vwretx) from 1973 to 2008.
- Compute the annualized dividend yield.
- Now focus on the log returns including dividends (vwretd).
- Compute the sample mean $\widehat{\mu}$ and standard deviation $\widehat{\sigma}$. Test whether $\widehat{\mu}$ is different from zero at a 5% significance level.
- Divide the sample period into four equal subperiods. Compute $\widehat{\mu}$ and $\widehat{\sigma}$ for each subperiod. Can you conclude that r_t is a stationary process?
- Do the Ljung-Box test on the autocorrelation coefficients for the first 10, 20 and 50 lags. Make Comments on the results.