FE570 Financial Markets and Trading

Lecture 11. Linear Time Series Model and Its Applications (Ref. Joel Hasbrouck - *Empirical Market Microstructure*)

Steve Yang

Stevens Institute of Technology

11/27/2018

Outline

Stationarity and Ergodicity 大数定理&中心极限定理

- Much statistical inference relies on the law of large numbers (LLN) and central limit theorem (CLT).

  LLN: says that the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.
  CLT: states that, given certain conditions, the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed
  Financial time series data are by nature dependent, therefore we rely on alternative: *stationarity* and *ergodicity*.

- **Ergodicity:**
  - A time series is *ergodic*, if its local stochastic behavior is (possibly in the limit) independent of the starting point, that is, initial conditions.
  - An ergodic process eventually "forgets" where it started.

- **Stationarity:**
  - A time series r_t is said to be strictly stationary if the joint distribution of $(r_{t_1}, \ldots, r_{t_k})$ is identical to that of $(r_{t_1+t}, \ldots, r_{t_k+t})$ for all t, where k is an arbitrary positive integer and $(t_1, \ldots, t_k)$ is a collection of k positive integers.
  - A time series r_t is said to be weakly stationary if both mean of r_t and the covariance between r_t and $r_{t-\ell}$ are time invariant, where ℓ is an arbitrary integer.

Strict: distributions are time-invariant
Weak: first 2 moments are time-invariant

- $E(r_t) = \mu$, which is a constant.
- $Cov(r_t, r_{t-\ell}) = \gamma_\ell$, which only depends on ℓ.

What does weak stationarity mean in practice?

Ans: it enables one to make inferences concerning future observations (e.g. prediction)

金融市场：return;risk

Weak Stationarity:

- The covariance $\gamma_\ell = Cov(r_t, r_{t-\ell})$ is called lag-ℓ autocovarance of r_t. It has two important properties:
  - $\gamma_0 = Var(r_t)$, which is a constant.
  - $\gamma_{-\ell} = \gamma_\ell$.

- *The second property holds because:*

$$\begin{aligned} Cov(r_t, r_{t-(-\ell)}) &= Cov(r_{t-(-\ell)}, r_t) \\ &= Cov(r_{t+\ell}, r_t) = Cov(r_{t_1}, r_{t_1-\ell}) \end{aligned}$$

,where $t_1 = t + \ell$.

- In practice:
  - It is common to assume that an asset return series is weakly stationary.
  - It can be empirically checked, provided that a sufficient number of historical returns available.

Estimation for a given sample data $\{r_1, ..., r_T\}$:

- Past: time plot of $\{r_t\}$ varies around a fixed level within a finite range.
- Future: the first 2 moments of future $\{r_t\}$ are the same as those of the data so that meaningful inferences can be made.
- Mean (or expectation) of returns:

$$\mu = E(r_t)$$

- Variance (variability) of returns:

$$Var(r_t) = E[(r_t - \mu)^2]$$

- Estimate using sample mean and sample variance of returns.

$$\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t \text{ and } Var(r_t) = \frac{1}{T-1}\sum_{t=1} T(r_t - \bar{r})^2$$

Correlation: 协方差的属性

- **The correlation coefficient between two random variables X and Y is defined as:**

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

  where μ_x and μ_y are the mean of X and Y, respectively, and it is assumed that the variances exist. This coefficient measures the strength of linear dependence between X and Y, and it can be shown that $-1 \leq \rho_{x,y} \leq 1$ and $\rho_{x,y} = \rho_{y,x}$.

- **Properties:**
  - The two random variables are uncorrelated if $\rho_{x,y} = 0$.
  - If both X and Y are normal random variables, then $\rho_{x,y} = 0$ if and only if X and Y are independent.

Sample Correlation: 样本协方差

- When the sample $\{(x_t, y_t)\}_{t=1}^{T}$ is available, the sample correlation can be estimated as:

$$\hat{\rho_{x,y}} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sum_{t=1}^{T}(y_t - \bar{y})^2}}$$

  where $\bar{x} = (\sum_{t=1} Tx_t)/T$ and $\bar{y} = (\sum_{t=1} Ty_t)/T$ are the sample mean of X and Y, respectively

- Consider a weakly stationary return series r_t, when the linear dependence between r_t and its past values r_{t-i} is of interest, the concept of correlation is generalized to autocorrelation. The correlation coefficient between r_t and $r_{t-\ell}$ is called the lag-ℓ autocorrelation of r_t and is commonly denoted by ρ_ℓ, which under the weak stationarity assumption is a function of ℓ only. Let $\bar{r}$ be the sample mean, and $\bar{r} = (\sum_{t=1}^{T} r_t)/T$.
- **Lag-ℓ autocovariance:**

$$\gamma_\ell = Cov(r_t, r_{t-\ell}) = E[(r_t - \bar{r})(r_{t-\ell} - \bar{r})].$$

Autocorrelation Function (ACF):

- **Serial (or auto-) correlations:**

$$\rho_\ell = \frac{Cov(r_t, r_{t-\ell})}{\sqrt{Var(r_t)Var(r_{t-\ell})}} = \frac{Cov(r_t, r_{t-\ell})}{Var(r_t)} = \frac{\gamma_\ell}{\gamma_0}$$

where the property $Var(r_t) = Var(r_{t-\ell})$ for a weakly stationary series is used.

Note: $\rho_0 = 1$ and $\rho_\ell = \rho_{-\ell}$ and $-1 \leq \rho_\ell \leq 1$ for $\ell \neq 0$. Why?

- A weakly stationary series r_t is not serially correlated if and only if $\rho_\ell = 0$ for all $\ell > 0$.
- *Sample autocorrelation function (ACF)*

$$\hat{\rho}_\ell = \frac{\sum_{t=1}^{T-\ell}(r_t - \bar{r})(r_{t+\ell} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2}$$

where $\bar{r}$ is the sample mean and T is the sample size.

Test zero serial correlations (market efficiency):

- Individual test: for example, asymptotically normal $N(0,1)$.

市场是否 efficient的test

$$H_0 : \rho_\ell = 0 vs. H_a : \rho_\ell \neq 0$$

$$t = \frac{\hat{\rho}_\ell}{\sqrt{1 + 2\sum_{i=1}^{\ell-1} \hat{\rho}_\ell^2}}$$

Decision rule: Reject H_0 if $|t| > Z_{\alpha/2}$ or p-value less than α.

- Joint test (Ljung-Box statistics):

$$H_0 : \rho_1 = ... = \rho_m = 0 vs. H_a : \rho_i \neq 0$$

$$Q(m) = T(T+2)\sum_{\ell=1}^{m} \frac{\hat{\rho_\ell^2}}{T-\ell}$$

Decision rule: Reject H_0 if $Q(m) > \chi_m^2(\alpha)$ or p-value less than α.

A proper perspective: at a time point t

- Available data $\{r_1, r_2, \ldots, r_{t-1}\} \equiv F_{t-1}$
- The return can be decomposed into two parts as

$$\begin{aligned} r_t &= \text{predictable part} + \text{not predictable part} \\ &= \text{function of elements of } F_{t-1} + a_t \end{aligned}$$

In other words, given information F_{t-1}

$$r_t = \mu_t + a_t = E(r_t | F_{t-1}) + \sigma_t \epsilon_t$$

- μ_t: conditional mean of r_t
- a_t: shock or innovation at time t
- ϵ_t: an iid sequence with mean zero and variance 1
- σ_t: conditional standard deviation (commonly called volatility in finance)

- Model for μ_t: **mean equation**
- Model for σ_t^2: **volatility equation**

- **White Noise:** a time series r_t is called a white noise if $\{r_t\}$ is a sequence of independent and identically distributed random variables with finite mean and variance.
  - *Gaussian White Noise:* if r_t is normally distributed with mean zero and variance σ^2. 高斯白噪声服从正态分布
  - For a white noise series, all the ACFs are zero.
  - In practice, if all sample ACFs are close to zero, then the series is a white noise series.

- **Linear Time Series:**
  - A time series r_t is said to be linear if it can be written as:

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

where μ is the mean of r_t, $\psi_0 = 1$ and $\{a_t\}$ is a sequence of independent and identically distributed random variables with mean zero and a well-defined distribution.

纯计算

- **Autoregressive Models AR(p):** says that the past p values r_{t-i} $(i = 1, ..., p)$ jointly determine the conditional expectation of r_t given the past data.

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + ... + \phi_p r_{t-p} + a_t, \qquad (1)$$

where p is a non-negative integer and $\{a_t\}$ is defined as white noise series with mean zero and variance σ_a^2.

- The AR(p) model is in the same form as a multiple linear regression model with lagged values serving as the explanatory variables.
- **AR(1) Model**
  - The lagged return r_{t-1} is useful in predicting r_t:

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t,$$

where $\{a_t\}$ is defined as white noise series with mean zero and variance σ_a^2.

AP(1) Model

- Under the weakly stationarity assumption, we have $E(r_t) = \mu$, $Var(r_t) = \gamma_0$, and $Cov(r_t, r_{t-j}) = \gamma_j$, where μ and γ_0 are constant and γ_j is a function of j, not t.
- We can easily obtain the mean, variance, and autocorrelations of the series as follows:

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

- Given $E(a_t) = 0$, and by stationarity condition, $E(r_t) = E(r_{t-1}) = \mu$ and hence:

$$\mu = \phi_0 + \phi_1 \mu, E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1}.$$

- Implications:
  - The mean of r_t exists, if $\phi_1 \neq 1$.
  - The mean of r_t is zero, if and only if $\phi_0 = 0$.

  Thus for a stationary AR(1) process, the constant term ϕ_0 is related to the mean of r_t and $\phi_0 = 0$ implies that $E(r_t) = 0$.

AR(1) Model

- Using $\phi_0 = (1 - \phi_1)\mu$, the AR(1) model can be rewritten as:

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + a_t. \tag{2}$$

- By repeated substitutions, the prior equation implies that:

$$\begin{aligned} r_t - \mu &= a_t + \phi_1 r_{t-1} + \phi_1^2 r_{t-2} + \ldots \\ &= \sum_{i=0}^{\infty} \phi_1^i a_{t-i} \end{aligned}$$

- Using this property and the independence of the series $\{a_t\}$, we obtain

$$E[(r_t - \mu)a_{t+1}] = 0$$

- By the stationarity assumption, we have

$$Cov(r_{t-1}, a_t) = E[(r_{t-1} - \mu)a_t] = 0$$

Autocorrelation Function of AR(1) Model

- Multiplying the following equation by a_t, using the independence between a_t and r_{t-1}, and taking expectation,

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + a_t. \tag{3}$$

  We then have

$$E[a_t(r_t - \mu)] = E[a_t(r_{t-1} - \mu)] + E[a_t^2] = E(a_t^2) = \sigma_a^2,$$

  where σ_a^2 is the variance of a_t.

- Multiplying Eq. 3 by $(r_{t-\ell} - \mu)$, and taking expectation, we have

$$\gamma_0 = \frac{\sigma^2}{1-\phi_1^2}, \text{and} \gamma_\ell = \phi_1 \gamma_{\ell-1}, \text{for} \ell > 0.$$

- The ACF of r_t satisfies, because $\rho_0 = 1$

$$\rho_\ell = \phi_1^\ell, \text{for} \ell \geq 0.$$

AR(2) Model

如何check stationary期末考点

- An AR(2) model assumes the form:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t. \quad (4)$$

- Using the same technique as that of the AR(1) case, we obtain:

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

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

- Using $\phi_0 = (1 - \phi_1 - \phi_2)\mu$, we can rewrite the AR(2) model as:

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

- Multiplying the prior equation by $(r_{t-\ell} - \mu)$, we we have:

$$\underline{(r_{t-\ell} - \mu)(r_t - \mu)} = (r_{t-\ell} - \mu)a_t + \phi_1(r_{t-\ell} - \mu)(r_{t-1} - \mu) + \phi_2(r_{t-\ell} - \mu)(r_{t-2} - \mu)$$

gama l 是此式的期望

AR(2) Model

- Taking expectation and using $E[(r_{t-\ell} - \mu)a_t] = 0$ for $\ell > 0$, we obtain:

$$\gamma_\ell = \phi_1 \gamma_{\ell-1} + \phi_2 \gamma_{\ell-2}, \text{for} \ell \geq 0. \quad (5)$$

- Dividing the above equation by γ_0, we have the property:

$$\underline{\rho_\ell = \phi_1 \rho_{\ell-1} + \phi_2 \rho_{\ell-2}, \text{for} \ell \geq 0.}$$

- In particular, the lag-1 ACF satisfies

$$\underline{\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_1 = \phi_1 + \phi_2 \rho_1.}$$

- Therefore, for a stationary AR(2) series r_t, we have $\rho_0 = 1$,

$$\rho_1 = \frac{\phi_1}{1 - \phi_2},$$
$$\rho_\ell = \phi_1 \rho_{\ell-1} + \phi_2 \rho_{\ell-2}, \ell \geq 2.$$

AR(2) Model

- Equation (4) says that the ACF of a stationary AR(2) series satisfies the second-order difference equation:

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

  where B is called *back − shift* operator such that $B\rho_\ell = \rho_{\ell-1}$.
- Corresponding to the prior difference equation, there is a second-order polynomial equation

$$1-\phi_1 x+\phi_2 x^2 = 0,$$

$$\text{and solutions are:} x = \frac{\phi_1 \pm \sqrt{\phi_1^2+4\phi_2}}{-2\phi_2},$$

$x < 1$

- Denote the two solutions by ω_1 and ω_2
  - If both ω_i are real valued, the AR(2) model can be regarded as an AR(1) model operating on top of another AR(1) model.
  - If they are complex numbers, and a plot of the ACF of r_t would show a picture of damping sine and cosine waves.

- **Stationarity of AR(2) Models** 如何证明stationary
  - The stationarity condition of an AR(2) time series is that the absolute values of its two characteristic roots are less than one or, equivalently, its two characteristic roots are less than one in modulus.

- Example: suppose an ACF with $\phi_1 = 0.36$ and $\phi_2 = -0.4$
  - Next figure shows the ACFs of four stationary AR(2) models.
  - Part(b) is the ACF of the AR(2) model $(1 - 0.6B + 0.4B^2)r_t = a_t$.
  - Because $\phi_1^2 + 4\phi_2 = 0.36 + 4x(-0.4) = -1.24 < 0$, this particular AR(2) model contains two complex characteristic roots, and hence its ACF exhibits damping sine and cosine waves.
  - All the other three ACFs have real roots, and their ACFs decay exponentially.

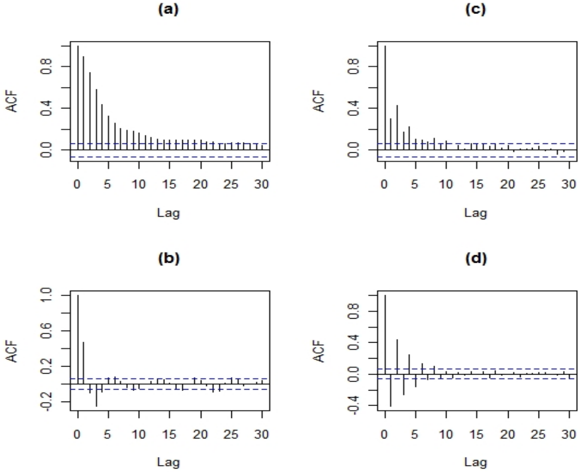

Figure: The autocorrelation function of an AR(2) model: (a) $\phi_1 = 1.2$ and $\phi_2 = -0.35$, (b) $\phi_1 = 0.6$ and $\phi_2 = -0.4$, (c) $\phi_1 = 0.2$ and $\phi_2 = 0.35$, (d) $\phi_1 = -0.2$ and $\phi_2 = 0.35$

AR(p) Model

- The mean of a staionary series is:

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

provided the denominator is not zero.

- The associated polynomial equation of the model is

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

which is referred to as the *characteristic equation* of the model.

- For a stationary AR(p) series, the ACF satisfies the difference equation:

$$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)\rho_\ell = 0, \text{for } \ell > 0.$$

A plot of the ACF of a stationary AR(p) model would then show a mixture of damping sine and cosine patterns and exponential decays.

Identify AR Model in Practice

- In application, the order p of an AR time series is unknown and must be specified empirically.
- There are two general approaches:
    - **Partial Autocorrelation Function (PACF)**
    - **Information Criteria**

- Partial Autocorrelation Function (PACF)
    - The PACF of a stationary time series is a function of its ACF and is a useful tool for determining the order p of an AR model. A simple yet effective way to introduce PACF is to consider the following AR models in consecutive orders:

$$
\begin{aligned}
r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + e_{1t}, \\
r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + e_{2t}, \\
r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + e_{3t}, \\
r_t &= \phi_{0,4} + \phi_{1,4} r_{t-1} + \phi_{2,4} r_{t-2} + \phi_{3,4} r_{t-4} + \phi_{4,4} r_{t-4} + e_{4t}, \\
&\cdots
\end{aligned}
$$

Identify AR Model in Practice

- These models are in the form of a muliple linear regression and can be estimated by the least squares method.
  - The estimated $\hat{\phi}_{1,1}$ of the first equation is called the lag-1 sample PACF of r_t.
  - The estimated $\hat{\phi}_{2,2}$ of the second equation is called the lag-2 sample PACF of r_t.
  - The estimated $\hat{\phi}_{3,3}$ of the third equation is called the lag-3 sample PACF of r_t.
  - ...
- Therefore, for an AR(p) model, the lag-p sample PACF should not be zero, but $\hat{\phi}_{j,j}$ should be close to zero for all $j > p$.
- For a stationary Gaussian AR(p) model, the sample PACF:
  - $\hat{\phi}_{p,p}$ converge to ϕ_p as the sample size T goes to infinity.
  - $\hat{\phi}_{\ell,\ell}$ converge to zero for all $\ell > p$.
  - The asymptotic variance of $\hat{\phi}_{\ell,\ell}$ is $1/T$ for $\ell > p$.
- The sample PACF of an AR(p) series custs off at lag p.

Information Criteria

- Akaike Information Criterion (AIC)
  - For a Gassian $AR(\ell)$ model, AIC reduces to:

$$AIC = ln(\tilde{\sigma}_\ell^2) + \frac{2\ell}{T},$$

    where $\tilde{\sigma}_\ell^2$ is the maximum likelihood estimate of σ_a^2, which is the variance of a_t, and T is the sample size.
  - Find the AR order with *minimum* AIC.
- Bayesian Information Criterion (BIC)
  - For a Gassian $AR(\ell)$ model, BIC reduces to:

$$BIC = ln(\tilde{\sigma}_\ell^2) + \frac{\ell ln(T)}{T},$$

- The penalty for each parameter used is 2 for AIC and ln(T) for BIC. Thus, BIC tends to select a lower AR model when the sample size is moderate or large.

Estimation - Least Squares or Likelihood Method

- Conditioning on the first p observations, we have:

$$r_t = \phi_0 + \phi_1 r_{t-1} + ... + \phi_p r_{t-p} + a_t, t = p+1, ..., T,$$

which is in the form of a multiple linear regression.

- Denote the estimate of ϕ_i by $\hat{\phi}_i$, the fitted model is

$$\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} + ... + \hat{\phi}_p r_{t-p},$$

and the series $\{\hat{a}_t\}$ is called the residual series:

$$\hat{a}_t = r_t - \hat{r}_t.$$

- Using least squares method, we obtain

$$\hat{\sigma}_a^2 = \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{T - 2p - 1}.$$

- Using likelihood method, $\tilde{\sigma}_a^2 = \hat{\sigma}_a^2 \times (T - 2p - 1)/(T - p)$.

- Model Checking
  - If the model is adequate, the residual series should behave as a white noise.
  - The ACF and Ljung-Box statistics of the residuals can be used to check the closeness of $\hat{a}_t$ to a white noise.
  - For an AR(p) model the Ljung-Box statistic $Q(m)$ follows asymptotically a chi-squared distribution with $m - g$ degrees of freedom, where g denotes the number of AR coefficients used in the model.

- Goodness of Fit
  - A commonly used statistic to measure goodness of fit of a stationary model is the R-square (R^2) defined as

$$R^2 = 1 - \frac{\text{Residual sum of squares}}{\text{Total sum squares}}.$$

  - For a stationary AR(p) time series model with T observations $\{r_t | t = 1, ..., T\}$ the measure becomes

- Goodness of Fit
  - Coefficient of Determination R^2

$$R^2 = 1 - \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{\sum_{t=p+1}^{T}(r_t - \bar{r})^2}.$$

    where $\bar{r} = (\sum_{t=p+1}^{T} r_t)/(T-p)$. It is easy to show that $0 \leq R^2 \leq 1$.
  - Typically, a larger R^2 indicates that the model provides a closer fit to the data.
  - It is well-known that R^2 is a nondecreasing function of the number of parameters used. To overcome this weakness, an *adjusted* R^2 is used:

$$Adj - R^2 = 1 - \frac{\text{Variance of residuals}}{\text{Variance of} r_t} = 1 - \frac{\hat{\sigma}_a^2}{\hat{\sigma}_r^2}.$$

  - This measure takes into account the number of parameters used in the fitted model, and it is no longer between 0 and 1.

- Forecasting
  - Suppose that we are at the time index h and interested in forecasting $r_{h+\ell}$, where $\ell \geq 1$. The time index h is called the *forecast origin* and the positive integer ℓ is the *forecase horizen*. Let $\hat{r}_h(\ell)$ be the forecast of $r_{h+\ell}$ using the minimum squared error loss function and F_h be the collection of information available at the forecast origin h. Then the forecast $\hat{r}(\ell)$ is chosen such that:

$$E[r_{h+\ell} - \hat{r}(\ell)]^2 | F_h] \leq \min_g E[(r_{h+\ell} - g)^2 | F_h].$$

    where g is a function of the information available at time h (inclusive), that is, a function of F_h.
  - We referred to $\hat{r}_h(\ell)$ as ℓ-step ahead forecast of r_t at the forecast origin h.
- 1-Step Ahead Forecast
  - From the AR(p) model, we have

$$r_{h+1} = \phi_0 + \phi_1 r_h + \ldots + \phi_p r_{h+1-p} + a_{h+1},$$

1-Step Ahead Forecast

- The point forecast of r_{h+1} given $F_h = \{r_h, r_{h-1}, \ldots\}$ is the conditional expectation

$$\hat{r}_h(1) = E(r_{h+1}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i},$$

and the associated forecast error is

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1},$$

- Consequently, the variance of the 1-step ahead forecast error is $Var[e_h(1)] = Var(a_{h+1}) = \sigma_a^2$.

2-Step Ahead Forecast

2步更差

- The point forecast of r_{h+2} given $F_h = \{r_h, r_{h-1}, \ldots\}$ is the conditional expectation

$$\hat{r}_h(2) = E(r_{h+2}|F_h) = \phi_0 + \phi_1 \hat{r}_h(1) + \sum_{i=2}^{p} \phi_i r_{h+2-i},$$

2-Step Ahead Forecast

- The associated forecast error is

$$e_h(2) = r_{h+2} - \hat{r}_h(2) = \phi_1[r_{h+1} - \hat{r}_h(1)] + a_{h+2} = \underline{a_{h+2} + \phi_1 a_{h+1}},$$

- The variance of the 2-step ahead forecast error is $Var[e_h(2)] = Var(a_{h+2}) + \phi_1 Var(a_{h+1}) = (1 + \phi_1^2)\sigma_a^2$.

Multistep Ahead Forecast

- The point forecast of $r_{h+\ell}$ given $F_h = \{r_h, r_{h-1}, ...\}$ is

$$\hat{r}_h(\ell) = E(r_{h+\ell}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(\ell - i),$$

  where it is understood that $\hat{r}_h(i) = r_{h+i}$ if $i \leq 0$. This forecast can be computed recursively using forecast $\hat{r}_h(i)$ for $i = 1, ..., \ell - 1$.
- The <u>ℓ-step ahead forecast error is $e_h(\ell) = r_{h+\ell} - \hat{r}_h(\ell)$</u>. 考试计算
- It can be shown that $\hat{r}_h(\ell)$ converges to $E(r_t)$ as $\ell \to \infty$ (*mean reversion*).

Example

Use R AR and ARIMA module to fit an AR model

- Estimate Parameters
- Model Checking
- Forecast Using Fitted Model