A Notebook of Time Series Analysis

Yuwen Long

2024

1 Introduction

This notebook focuses on fundamental time series models in finance, which of course can be scaled up to other fields. The basic ideas, concepts, theorems and propositions are discussed in details to provide further insight on specific models.

Before the content, some notation is to be introduced. The price of a certain underlying at time t is denoted by P_t. It is common to use the logarithmic rate of return, $r_t = \ln\left(\frac{P_t - P_{t-1}}{P_{t-1}} + 1\right)$, in financial analysis.

2 Stationarity and White Noise Series

2.1 Stationarity

A stochastic process $\{a_t\}$ is called *stationary* if any joint distribution of a collection of its random variables $\{a_t, a_{t+1}, \ldots, a_{t+k}\}_k$ is only related to k, that is its distribution is *time invariant*. However, this *strict* stationarity is hardly satisfied in real data. So the *weak stationarity* is discussed more, which says $\mathrm{Cov}\left(a_t, a_{t+k}\right) = \gamma_k$ which is a function only of the lagged time k. And, a weak stationary series also defined to hold that $\mathbb{E}\left(a_t\right) = \mu, \mathrm{Var}\left(a_t\right) = \sigma^2 < \infty$ for all $t = 1, 2, \ldots, T$.

2.2 White Noise Series

A series of random variables $\{a_t\}$ is called *white noise series* if

$$\begin{aligned} \mathbb{E}\left(a_t\right) &= 0, \\ \mathrm{Var}\left(a_t\right) &= \sigma^2 < \infty, \\ \mathrm{Cov}\left(a_t, a_{t+k}\right) &= 0, \end{aligned} \tag{2.1}$$

which is called *Gauss-Markov Condition* in linear models. It is worthwhile to pointed that a white noise series is self-uncorrelated but not necessarily independent.

2.3 Martingale Difference Sequence

We a stochastic process $\{a_t\}$ a *Martingale Difference Sequence*(MDS) if $a_t \in \mathcal{F}_t$, where $\mathcal{F}_t$ is the information set(the information available) at time t, and $\mathbb{E}\left(a_t | \mathcal{F}_{t-1}\right) = 0$ which means we can know nothing about a_t dependent on the past information. It is obvious that an *iid* series of random variables must be MDS. And using Iterated expectation theorem, one can get a MDS must be white noises. At last, from the definition of stationarity and white noises, it is obvious s stationary series is a subset of white noise series.

3 AR Model

3.1 ACF

The correlation of a time series $\{r_t\}$ is about the components itself, so it is called *Auto Correlation Function*(ACF). First consider the covariance $\mathrm{Cov}\left(r_t, r_{t+k}\right)$. If that time series is stationary, then

$$\mathrm{Cov}\left(r_t, r_{t+k}\right) = \gamma_k. \tag{3.1}$$

And it can be seen that

$$\mathrm{Cov}\left(r_t, r_{t+k}\right) = \gamma_k = \gamma_{-k} = \mathrm{Cov}\left(r_t, r_{t-k}\right), \tag{3.2}$$

since $\mathrm{Cov}\left(r_t, r_{t-k}\right) = \mathrm{Cov}\left(r_{t-k}, r_t\right)$ and let $l = t - k$, then $\mathrm{Cov}\left(r_t, r_{t-k}\right) = \mathrm{Cov}\left(r_l, r_{l+k}\right) = \gamma_k$. Also, $\mathrm{Var}\left(a_t\right) = \gamma_0$.

Then, we can make notes about the correlation as

$$\rho_k = \frac{\mathrm{Cov}\left(r_t, r_{t+k}\right)}{\sqrt{\mathrm{Var}\left(r_t\right)\mathrm{Var}\left(r_{t+k}\right)}} = \frac{\gamma_k}{\gamma_0}, \tag{3.3}$$

Which is the auto correlation function of $\{r_t\}$. More specifically, if we have the sample of $\{r_t\}$, we can calculate the sample covariance as

$$\hat{\gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}\left(r_t - \bar{r}\right)\left(r_{t-k} - \bar{r}\right), \tag{3.4}$$

where $\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$. And the ACF can be calculated as

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}\left(r_t - \bar{r}\right)\left(r_{t-k} - \bar{r}\right)}{\sum_{t=1}^{T}\left(r_t - \bar{r}\right)^2}. \tag{3.5}$$

We can calculated the ACF of a sequence of k to identify which lagged r_{t-k}s significantly influence r_t. An example is illustrated by the following figure from [1].

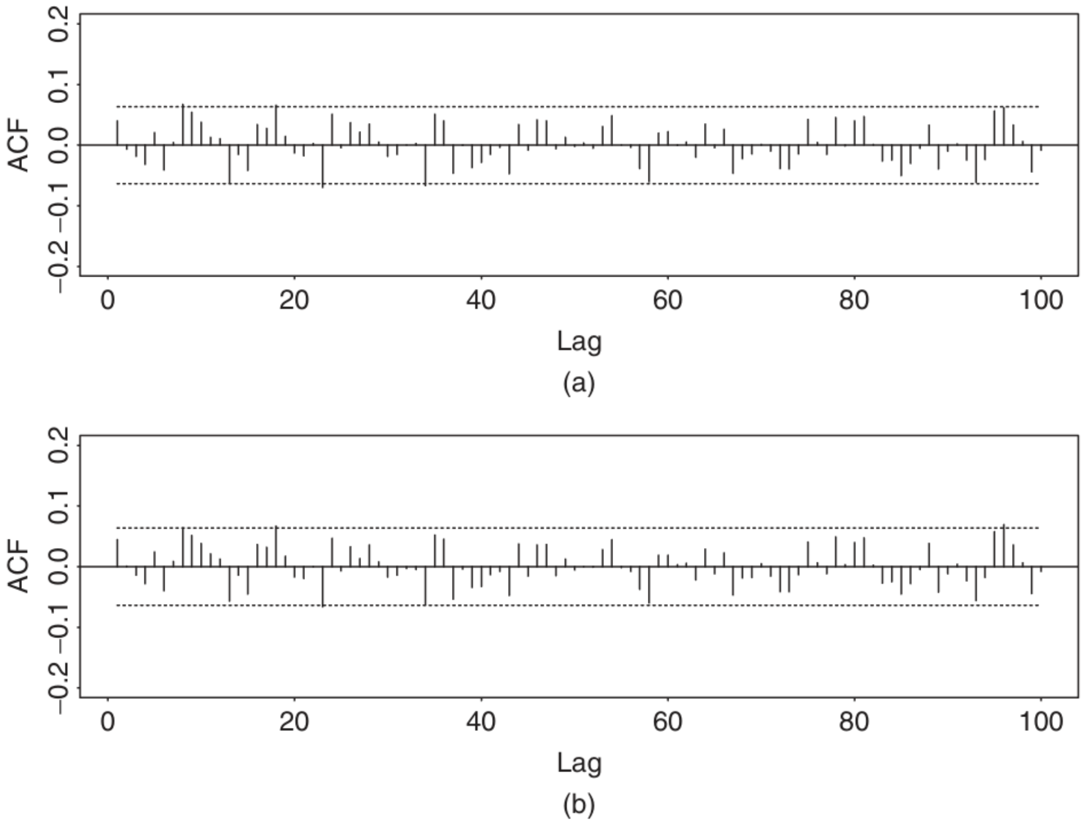

Figure 1: Sample autocorrelation functions of monthly (a) simple returns and (b) log returns of IBM stock from January 1926 to December 2008. In each plot, two horizontal dashed lines denote two standard error limits of sample ACF.

That is, those ACF which exceeds the two standard error limits is considered to be significant.

What's more, there are some test statistics for ACF.

***t*-Test**

For a given positive integer k, we test $H_0 : \rho_k = 0$ vs. $H_1 : \rho_k \neq 0$. The t-statistic is

$$t \text{ ratio} = \frac{\hat{\rho}_k}{\sqrt{\left(1 + 2\sum_{i=1}^{k-1} \hat{\rho}_k^2\right)/T}}. \tag{3.6}$$

If $\{r_t\}$ is a Gaussian stationary series satisfying $\rho_j = 0$ for $j > k$, then the t ration defined above is asymptotically distributed as a standard normal random variable. And the hypothesis test is a two-side test.

Portmanteau Test

The statistic define below test several correlations of r_t are zero jointly. The Portmanteau statistic is

$$\mathcal{Q}^*(m) = T \sum_{i=1}^{m} \hat{\rho}_i \tag{3.7}$$

which test $H_0 : \rho_1 = \cdots = \rho_m = 0$ vs. $H_1 : \rho_i \neq 0$ for some $i \in \{1, \ldots, m\}$. If $\{r_t\}$ is an iid sequence with certain moment conditions, $\mathcal{Q}^*$ is asymptotically a chi-squared random variable with m degrees of freedom.

The Portmanteau statistics is modified by Ljung and Box to increase the power of the test as

$$\mathcal{Q}\left(m\right) = T\left(T+2\right)\sum_{i=1}^{m}\frac{\hat{\rho}_i^2}{T-i}. \tag{3.8}$$

Both statistics above reject H_0 if its value is larger than $\chi^2_{1-\alpha}\left(m\right)$, the $100\left(1-\alpha\right)$th percentile of a chi-squared distribution with m degrees of freedom.

3.2 AR(1) Model

The *autoregressive*(AR) model talks about r_t is correlated with some lagged values in the same time series, which give the name "auto". First consider the AR(1) model whose form is

$$r_t = \phi_0 + \phi_1 r_{t-1} + \epsilon_t, \tag{3.9}$$

where $\{\epsilon_t\}$ is a white noise series which satisfies the Gaussian-Markov Condition (2.1)

The Mean of AR(1)

If $\{r_t\}$ satisfies $\mathbb{E}\left(r_t\right) = \mu$ for all $t = 1, 2, \ldots, T$, then take expectation of both sides of (3.9) we have

$$\begin{aligned}\mathbb{E}\left(r_t\right) &= \phi_0 + \phi_1\,\mathbb{E}\left(r_{t-1}\right)\\ \mu &= \frac{\phi_0}{1-\phi_1}.\end{aligned} \tag{3.10}$$

So the mean of $\{r_t\}$ exists if $\phi_1 \neq 1$.

The Covariance of AR(1)

Replace ϕ_0 by $\phi_0 = \mu\left(1-\phi_1\right)$, (3.9) can be written as

$$r_t - \mu = \phi_1\left(r_{t-1}-\mu\right) + \epsilon_t. \tag{3.11}$$

Multiply both sides of (3.11) by ϵ_{t+k}, use iterative equation of AR(1) model and the properties of white noise series and take expectation of both sides we have

$$\mathbb{E}\left[\left(r_t-\mu\right)\epsilon_{t+k}\right] = \begin{cases}0, & k \neq 0,\\ \sigma^2, & k = 0.\end{cases} \tag{3.12}$$

Multiply both sides of (3.11) by $\left(r_{t-k}-\mu\right)$ for positive integers k, we have

$$\mathbb{E}\left[\left(r_t-\mu\right)\left(r_{t-k}-\mu\right)\right] = \phi_1\,\mathbb{E}\left[\left(r_{t-1}-\mu\right)\left(r_{t-k}-\mu\right)\right] + \mathbb{E}\left[\left(r_{t-k}-\mu\right)\epsilon_t\right], \tag{3.13}$$

which leads to

$$\gamma_k = \begin{cases}\phi_1\gamma_1 + \sigma^2, & k = 0,\\ \phi_1\gamma_0, & k = 1,\\ \phi_1\gamma_{k-1}, & k > 1,\end{cases} \tag{3.14}$$

from which we can derive that $\mathrm{Var}\left(r_t\right) = \gamma_0 = \frac{\sigma^2}{1-\phi_1^2}$ and $\gamma_1 = \frac{\phi_1\sigma^2}{1-\phi_1^2}$.

ACF of AR(1)

Divided both sides of (3.14) by γ_0, we have

$$\begin{cases}\rho_0 = 1,\\ \rho_1 = \phi_1,\\ \rho_k = \phi_1\rho_{k-1}, & k > 1.\end{cases} \tag{3.15}$$

Stationarity of AR(1)

Theorem 3.1. *In the AR(1) model such that $r_t = \phi_0 + \phi_1 r_{t-1} + \epsilon_t$, $\{r_t\}$ is weak stationary if and only if* $|\phi_1| < 1$.

Proof. Using the iterative equation of AR(1) model, we have

$$r_t - \mu = \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i},$$

thus $\mathrm{Var}(r_t) = \mathbb{E}(r_t - \mu) = \mathbb{E}\left(\sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i}\right) = \sigma^2 \sum_{i=0}^{\infty} \phi_1^i$, which converges,the condition that a weak stationary series must holds, if and only if $|\phi_1| < 1$. $\square$

3.3 AR(2) Model

The AR(2) model tells that the r_t is correlated with the past 2 variables as follows:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \epsilon_t, \tag{3.16}$$

The Mean of AR(2)

Just apply the same argument of the AR(1) model, we have

$$\mu = \mathbb{E}(r_t) = \frac{\phi_0}{1 - \phi_1 - \phi_2}, \tag{3.17}$$

which exists if and only if $1 - \phi_1 - \phi_2 \neq 1$.

The Covariance of AR(2)

Replace ϕ_0 by $\phi_0 = \mu(1 - \phi_1 - \phi_2)$, (3.16) can be written as

$$r_t - \mu = \phi_1 (r_{t-1} - 1) + \phi_2 (r_{t-2} - 1) + \epsilon_t.. \tag{3.18}$$

Multiply both sides of (3.18) by ϵ_{t+k}, use iterative equation of AR(2) model and the properties of white noise series and take expectation of both sides we have

$$\mathbb{E}\left[(r_t - \mu)\epsilon_{t+k}\right] = \begin{cases} 0, & k \neq 0, \\ \sigma^2, & k = 0 \end{cases} \tag{3.19}$$

just the same as the one of the AR(1) model.

Multiply both sides of (3.18) by $(r_{t-k} - \mu)$ for positive integers k, we have

$$\mathbb{E}\left[(r_t - \mu)(r_{t-k} - \mu)\right] = \phi_1 \mathbb{E}\left[(r_{t-1} - \mu)(r_{t-k} - \mu)\right] + \phi_2 \mathbb{E}\left[(r_{t-2} - \mu)(r_{t-k} - \mu)\right] + \mathbb{E}\left[(r_{t-k} - \mu)\epsilon_t\right], \tag{3.20}$$

which leads to

$$\gamma_k = \begin{cases} \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2, & k = 0, \\ \phi_1\gamma_0 + \phi_2\gamma_1, & k = 1, \\ \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}, & k > 1. \end{cases} \tag{3.21}$$

ACF of AR(2)

Divided both sides of (3.21) by γ_0, we have

$$\begin{cases} \rho_0 = 1, \\ \rho_1 = \dfrac{\phi_1}{1 - \phi_2}, \\ \rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, & k > 1. \end{cases} \tag{3.22}$$

Stationarity of AR(2)

We first introduce the concept of the *Lag Operator*.

Definition 3.1. The lag operator L delays the time index by 1 period as follows:

$$Lr_t = r_{t-1}.$$

We can then define a polynormial function of L (called *lag polynormial*) such that

$$f_L = \alpha_0 + \alpha_1 L + \cdots + \alpha_p L^p = \sum_{i=0}^{p} \alpha_i L^i$$

such that

$$f_L(r_t) = \alpha_0 r_t + \alpha_1 r_{t-1} + \cdots + \alpha_p r_{t-p} = \sum_{i=0}^{p} \alpha_i r_{t-i}.$$

Actually, one can verify that f_L is a linear transformation. And just consider the covariance as an inner product in a certain space, we have

Theorem 3.2. *If $\{r_t\}$ is covariance stationary, then $f_L(r_t)$ is also covariance stationary.*

Definition 3.2. A polynormial function of the lag operater is called *filter*. Let two filters be $f_L = \sum_{i=0}^{p} \alpha_i L^i$ and $g_L = \sum_{i=0}^{p} \beta_i L^i$, then the product of filters is

$$f_L g_L = h_L = \sum_{j=0}^{p} \delta_j,$$

where $\delta_i = \sum_{i=0}^{j} (\alpha_i \beta_{j-i})$.

Since the filter is a kind of linear transformation, it is natural to think about its *inverse*.

Definition 3.3. The inverse of a filter f_L is a polynormial function g_L of the lag operator such that

$$f_L g_L = 1.$$

Following the conclusion of linear transformation, the inverse of a certain filter is unique. Thus, we can write $f_L^{-1} = g_L$.

Theorem 3.3. *The inverse of a filter $f_L = \sum_{i=0}^{\infty} \alpha_i L^i$ is $g_L = \sum_{i=0}^{\infty} \beta_i L^i$, where*

$$\begin{aligned}\beta_n &= -\left(\frac{\alpha_1^n}{\alpha_0^{n+1}} + \frac{\alpha_2^{n-1}}{\alpha_0^n} + \cdots + \frac{\alpha_n}{\alpha_0^2}\right) \\ &= -\sum_{i=1}^{n} \frac{\alpha_i^{n+1-i}}{\alpha_0^{n+2-i}}, \quad n > 0 \text{ and} \\ \beta_0 &= \frac{1}{\alpha_0}.\end{aligned} \tag{3.23}$$

Proof. It suffices to solve the linear equations

$$\begin{cases} \alpha_0 \beta_0 = 1, \\ \alpha_0 \beta_1 + \alpha_1 \beta_0 = 0, \\ \cdots, \\ \alpha_0 \beta_n + \alpha_1 \beta_{n-1} + \cdots + \alpha_n \beta_0 = 0, \\ \cdots, \end{cases} \tag{3.24}$$

which can be written in matrix form as

$$\begin{pmatrix} \alpha_0 & & & \cdots & \\ \alpha_1 & \alpha_0 & & \cdots & \\ \alpha_2 & \alpha_1 & \alpha_2 & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \\ \alpha_n & \alpha_{n-1} & \alpha_{n-2} & \cdots & \alpha_0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{3.25}$$

And we can use Gaussian elimination to solve these equations to get the desired result. □

With the lag operator, we can get further insight on the stationarity of the AR model.

Theorem 3.4. *The $\{r_t\}$ of the AR(2) model is weak stationary if and only if* $\left|1/\frac{\phi_1 \pm \sqrt{4\phi_2+\phi_1^2}}{-2\phi_2}\right| = \frac{\phi_1 \pm \sqrt{4\phi_2+\phi_1^2}}{2} < 1$.

Proof. From the textbook, the argument is using the iterative equations (3.22) and the lag operator to write the equation as a second-order difference equation as

$$\left(1 - \phi_1 L - \phi_2 L^2\right) \rho_t = 0,$$

which is corresponding to solving

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

The two solutions are

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

The inverse of the two solution are referred to as the *characteristic roots* of the AR(2) model. We denote the two characteristic roots as ω_1 and ω_2. If x is allowed to take complex value, then the characteristics roots give rise to the behavior of business cycles, whose average length is

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

However, we offer another argument to prove this theorem.

To simplify the proof without loss of generality, suppose $\{r_t\}$ is a centered series, which means that the raw series is minus by its mean so that the mean of the new series $\{r_t\}$ is 0. In this way, $\phi_0 = 0$. Then, the iterative equation can be written in the matrix form as

$$\begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} r_t \\ r_{t-1} \end{pmatrix} = \begin{pmatrix} r_{t+1} \\ r_t. \end{pmatrix} \tag{3.26}$$

By repeated substitution, we have

$$\begin{pmatrix} r_{t+1} \\ r_t \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}^t \begin{pmatrix} r_1 \\ r_0 \end{pmatrix}. \tag{3.27}$$

We can calculate the power of that matrix by diagonalization. The eigenvalues of that matrix are $\lambda_1 == \frac{\phi_1 + \sqrt{4\phi_2+\phi_1^2}}{2}$ and $\lambda_1 == \frac{\phi_1 - \sqrt{4\phi_2+\phi_1^2}}{2}$. Then, (3.27) can be written as

$$\begin{pmatrix} r_{t+1} \\ r_t \end{pmatrix} = P \begin{pmatrix} \lambda_1^t & 0 \\ 0 & \lambda_2^t \end{pmatrix} P^{-1} \begin{pmatrix} r_1 \\ r_0 \end{pmatrix}, \tag{3.28}$$

where P is the matrix of eigenvectors. $\{r_t\}$ is weak stationary must have $|\lambda_1|, |\lambda_2| < 1$. Otherwise, as t goes to infinity, $|r_t|$ would be infinite and the variance of $\{r_t\}$ acts similarly once we take variance of both sides of (3.28). And the other direction is straight by the definition of weak stationarity. □

3.4 AR(p) Model

The AR(p) model has the following form:

$$\begin{aligned} r_t &= \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + \epsilon_t \\ &= \phi_0 + \sum_{k=1}^{p} \phi_k r_{t-k} + \epsilon_t. \end{aligned} \tag{3.29}$$

Just apply the same arguments of AR(1) and AR(2) models, we can derive some properties of the AR(p) model.

The Mean of AR(p)

The mean of the AR(p) model is $\mu = \frac{\phi_0}{\sum_{k=1}^{p} \phi_k}$.

ACF of AR(p)

The covariance of the AR(p) model is as follows:

$$\begin{aligned} \gamma_0 &= \frac{\sigma^2}{1 - \sum_{k=1}^{p} \phi_k^2}, \\ \gamma_l &= \sum_{k=1}^{p} \phi_k \gamma_{l-k}, l > 0. \end{aligned} \tag{3.30}$$

Then, the ACF of the AR(p) is

$$\begin{aligned} \rho_0 &= 1, \\ \rho_l &= \sum_{k=1}^{p} \phi_k \rho_{l-k}, l > 0. \end{aligned} \tag{3.31}$$

Stationarity of AR(p)

Theorem 3.5. *The corresponding characteristic equation of the AR(p) model is*

$$1 - \sum_{k=1}^{p} \phi_k x^k = 0. \tag{3.32}$$

An AR(p) model is stationary if and only if all modules of its characteristic roots is less than 1.

Proof. See the appendix of this chapter. □

3.5 Identifying the Order of AR Models

3.5.1 PACF

The *Partial Autocorrelatiion Function* mimics the similar logic in testing the coefficients in linear regression. Consider a sequence of AR models:

$$\begin{aligned} r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + e_1 t, \\ r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + e_2 t, \\ r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{2,3} r_{t-3} e_3 t, \\ &\vdots \end{aligned}$$

If an AR model has order p, then the $\phi_{p,p}$ should not be zero and all $\phi_{j,j}$ should be zero for all $j > p$. Particularly, for a stationary Gaussian AR(p) model, its sample PACF has the following properties:

- $\hat{\phi_{p,p}}$ converges to ϕ_p as the sample size T goes to infinity.
- $\hat{\phi_{j,j}}$ converges to zero for all $j > p$.
- The asymptotic variance of $\hat{\phi_{j,j}}$ is $1/T$ for $j > p$.

3.6 Information Criteria

We can also use the information criteria to determine the order p similar to what we do in linear regression. The first typical criteria is *Akaike information criterion*(AIC), defined as

$$\text{AIC} = \frac{-2}{T} \ln \text{likelihood} + \frac{2}{T} \times (\text{number of parameters}), \tag{3.33}$$

where the likelihood is the same we use in MLE. For Gaussian AR(p) model, AIC reduces to

$$\text{AIC}(p) = \ln\left(\tilde{\sigma_p^2}\right) + \frac{2p}{T}, \tag{3.34}$$

where $\tilde{\sigma_p^2}$ is the MLE of σ_p^2.

Another similar criteria is the *Schwarz-Bayesian information criterion*(BIC). For a Gaussian AR(p) model, the criterion is

$$\text{BIC}(p) = \ln\left(\tilde{\sigma_p^2}\right) + \frac{p \ln T}{T}. \tag{3.35}$$

For both criteria introduced above, a smaller value shows a better model.

3.7 Model Checking

If the model is correct, then its random errors should be a white noise series. Thus, we can use the Box-Pierce or Ljung-Box statistic to test the hypothesis.

What's more, one can use the usual measure in linear regression like R^2 or F-test to check the goodness of fit of the fitted model without worry since after determine the order p, we just take the AR(p) model as a linear regression and use OLS to estimate it.

3.8 Forecasting

Let T denote the end time of the observed samples. For a AR(p) model, the 1-step-head forecast forward T is

$$\hat{r}_T(1) = \mathbb{E}(r_{T+1}|F_T) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{T+1-i}, \tag{3.36}$$

and the forecast error is

$$e_T(1) = r_{T+1} - \hat{r}_T(1) = \epsilon_{T+1}. \tag{3.37}$$

Its variance is σ^2.

Similarly, the 2-step-ahead forecast is

$$\hat{r}_T(2) = \mathbb{E}(r_{T+2}|F_T) = \phi_0 + \phi_1 \hat{r}_T(1) + \sum_{i=2}^{p} \phi_i r_{T+1-i}, \tag{3.38}$$

and the forecast error is

$$e_T(2) = r_{T+2} - \hat{r}_T(2) = \epsilon_{T+2} + \phi_1 (r_{T+1} - \hat{r}_T(1)) = \epsilon_{T+2} + \phi_1 \epsilon_{T+1}. \tag{3.39}$$

Its variance is $\left(1 + \phi_1^2\right) \sigma^2$.

In general, the k-step-ahead forecast is

$$\hat{r}_T(k) = \mathbb{E}(r_{T+k}|F_T) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_T(k-i), \tag{3.40}$$

where $\hat{r}_T(k-i) = r_{T+k-i}$ if $k - i \leq 0$. And the forecast error is

$$e_T(k) = r_{T+k} - \hat{r}_T(k) = \epsilon_{T+k} + \sum_{i=1}^{p} \phi_{T+k-i} e_T(k-i), \tag{3.41}$$

where $e_T(k-i)$ is ϵ_{T+k-i} if $k - i > 0$, and 0 otherwise. Thus, as k goes to infinity, the forecast converges to the unconditional mean of $\{r_t\}$. This property is referred to as the *mean reversion*.

References

[1] RUEY S. TSAY. *Analysis of Financial Time Series*. A JOHN WILEY & SONS, INC., PUBLICATION.