Topic 3: Simple Autoregressive Models

Zongye Huang

International School of Economics and Management
Capital University of Economics and Business

March 6th, 2017

Outline

Linear Models for Financial Time Series I

- Introduce methods and linear models useful in modeling and forecasting financial time series

  1. Simple autoregressive (AR) models,
  2. Simple moving average (MA) models,
  3. Mixed autoregressive moving average (ARMA) models,
  4. Unit-root models including unit-root tests,
  5. Exponential smoothing,
  6. Seasonal models,
  7. Regression models with time series errors,
  8. fractionally differenced models for long-range dependence.

Outline

AR Model I

- When x_t has a statistically significant lag-1 autocorrelation, the lagged value x_{t-1} might be useful in predicting x_t.
- A simple model that makes use of such predictive power is

$$x_t = \phi_0 + \phi_1 x_{t-1} + a_t$$

  - This model is in the same form as the well-known simple linear regression model, referred to as an **AR model** of order 1, or simply an **AR(1)** model.
  - a_t is a white noise series with mean 0 and variance σ_a^2.
  - x_t is the dependent variable and x_{t-1} is the explanatory variable.

AR Model II

- Conditional on the past return x_{t-1},

$$\begin{aligned} E(x_t|x_{t-1}) &= \phi_0 + \phi_1 x_{t-1} \\ Var(x_t|x_{t-1}) &= \sigma_a^2 \end{aligned}$$

  - Given the past return x_{t-1}, the current return is centered around $\phi_0 + \phi_1 x_{t-1}$ with standard deviation σ_a.
  - This is a Markov property such that conditional on x_{t-1}, the return x_t is not correlated with x_{t-i} for $i > 1$.

- AR(p), p is a nonnegative integer

$$x_t = \phi_0 + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + a_t$$

  - It is the same form as a multiple linear regression model with lagged values serving as the explanatory variables.

AR(1) Model I

$$x_t = \phi_0 + \phi_1 x_{t-1} + a_t$$

- Assuming that the series is weakly stationary, we have

$$\begin{aligned} E(x_t) &= \mu \\ Var(x_t) &= \gamma_0 \\ Cov(x_t, x_{t-j}) &= \gamma_j \end{aligned}$$

  - μ and γ_0 are constant
  - γ_j is a function of j, not t.

AR(1) Model II

- The mean of AR(1),
  - Taking the expection and using $E(a_t) = 0$

$$\begin{aligned} E(x_t) &= \phi_0 + \phi_1 E(x_{t-1}) \\ \mu &= \phi_0 + \phi_1 \mu \end{aligned}$$

  - $E(x_t) = \mu = \frac{\phi_0}{1-\phi_1}$.
    - The mean of x_t exists if $\phi_1 \neq 1$.
    - The mean of x_t is 0 if and only if $\phi_0 = 0$.

AR(1) Model III

- The variance of AR(1)
  - The AR(1) model can be rewritten as

$$\begin{aligned} x_t - \mu &= \phi_1(x_{t-1} - \mu) + a_t \\ &= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \dots \\ &= \sum_{i=0}^{\infty} \phi_1^i a_{t-i} \end{aligned}$$

    - The AR(1) model is a linear time series, $\psi_i = \phi_1^i$.
    - $x_t - \mu$ is a linear function of a_{t-i} for $i \geqslant 0$.

AR(1) Model IV

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

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

It requires $\phi_1^2 < 1$.

AR(1) Model V

- In summary, the **necessary and sufficient condition** for the AR(1) model to be **weakly stationary** is $|\phi_1| < 1$.
- Using $\phi_0 = (1 - \phi_1)\mu$, in the finance literature, a stationary AR(1) rewrites as

$$x_t = (1 - \phi_1)\mu + \phi_1 x_{t-1} + a_t,$$

where ϕ_1 measures the persistence of the dynamic dependence of an AR(1) time series.

AR(1) Model VI

Autocorrelation Function of an AR(1) Model. Multiplying Equation (2.10) by a_t, using the independence between a_t and x_{t-1}, and taking expectation, we obtain

$$E[a_t(x_t - \mu)] = \phi_1 E[a_t(x_{t-1} - \mu)] + E(a_t^2) = E(a_t^2) = \sigma_a^2,$$

where σ_a^2 is the variance of a_t. Multiplying Equation (2.10) by $(x_{t-\ell} - \mu)$, taking expectation, and using the prior result, we have

$$\gamma_\ell = \begin{cases} \phi_1\gamma_1 + \sigma_a^2 & \text{if } \ell = 0 \\ \phi_1\gamma_{\ell-1} & \text{if } \ell > 0, \end{cases}$$

where we use $\gamma_\ell = \gamma_{-\ell}$. Consequently, for a weakly stationary AR(1) model in Equation (2.8), we have

$$\mathrm{Var}(x_t) = \gamma_0 = \frac{\sigma^2}{1 - \phi_1^2} \quad \text{and} \quad \gamma_\ell = \phi_1\gamma_{\ell-1}, \quad \text{for} \quad \ell > 0.$$

AR(1) Model VII

From the latter equation, the ACF of x_t satisfies

$$\rho_\ell = \phi_1 \rho_{\ell-1}, \quad \text{for} \quad \ell > 0.$$

Because $\rho_0 = 1$, we have $\rho_\ell = \phi_1^\ell$. This result says that the ACF of a weakly stationary AR(1) series decays exponentially with rate ϕ_1 and starting value $\rho_0 = 1$. For a positive ϕ_1, the plot of ACF of an AR(1) model shows a nice exponential decay. For a negative ϕ_1, the plot consists of two alternating exponential decays with rate ϕ_1^2. Figure 2.8 shows the ACF of two AR(1) models with $\phi_1 = 0.8$ and $\phi_1 = -0.8$.

AR(1) Model VIII

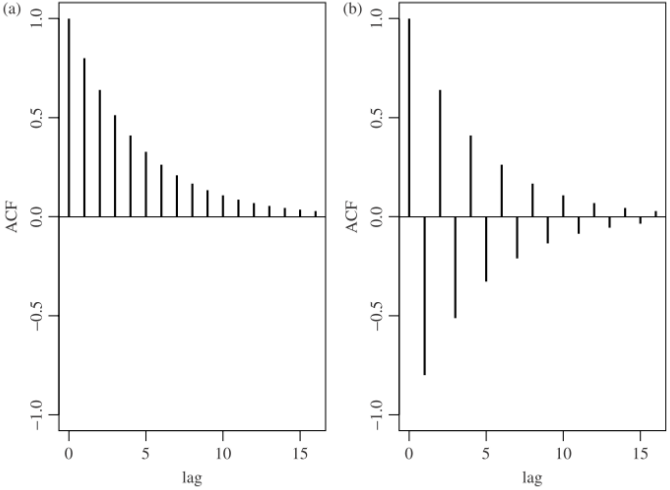

Figure 2.8. The autocorrelation function of an AR(1) model: (a) for $\phi_1 = 0.8$ and (b) for $\phi_1 = -0.8$.

Outline

AR(2) Model I

- An AR(2) model assumes

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + a_t$$

- Taking the expection and using $E(a_t) = 0$

$$\begin{aligned} E(x_t) &= \phi_0 + \phi_1 E(x_{t-1}) + \phi_2 E(x_{t-2}) \\ \mu &= \phi_0 + (\phi_1 + \phi_2)\mu \end{aligned}$$

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

AR(2) Model II

- The AR(2) model rewrites as

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

Multiplying by $(x_{t-k} - \mu)$

$$\begin{aligned}(x_{t-k} - \mu)(x_t - \mu) &= \phi_1(x_{t-k} - \mu)(x_{t-1} - \mu) \\ &\quad + \phi_2(x_{t-k} - \mu)(x_{t-2} - \mu) + (x_{t-k} - \mu)a_t\end{aligned}$$

Taking expectation

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

- This result is referred to as the moment equation of a stationary AR(2) model. Dividing the above equation by γ_0,

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

AR(2) Model III

- For the lag-1 ACF

$$\rho_1 = \phi_1\rho_0 + \phi_2\rho_{-1}$$

- Since $\rho_0 = 1$, for a stationary AR(2) series,

$$\begin{aligned} \rho_1 &= \frac{\phi_1}{1-\phi_2} \\ \rho_k &= \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, \; k > 1. \end{aligned}$$

- This result says that the ACF of a stationary AR(2) series satisfies the second-order difference equation

$$(1 - \phi_1 B - \phi_2 B^2)\rho_k = 0$$

where B is called the **backshift** operator such that $B\rho_k = \rho_{k-1}$.

AR(2) Model IV

- In the time series literature, some people use the notation L for **lag** operator, $Lx_t = x_{t-1}$.

- Corresponding to the prior difference equation, there is a second-order polynomial equation

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

Solutions of this equation are

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

AR(2) Model V

In the time series literature, inverses of the two solutions are referred to as the *characteristic roots* of the AR(2) model. Denote the two characteristic roots by ω_1 and ω_2. If both ω_i are real valued, then the second-order difference equation of the model can be factored as $(1-\omega_1 B)(1-\omega_2 B)$, and the AR(2) model can be regarded as an AR(1) model operates on top of another AR(1) model. The ACF of x_t is then a mixture of two exponential decays. If $\phi_1^2 + 4\phi_2 < 0$, then ω_1 and ω_2 are complex numbers (called a *complex conjugate pair*), and the plot of ACF of x_t would show a picture of damping sine and cosine waves. In business and economic applications, complex characteristic roots are important. They give rise to the behavior of business cycles. It is then common for economic time series models to have complex-valued characteristic roots. For an AR(2) model in Equation (2.12) with a pair of complex characteristic roots, the *average* length of the stochastic cycles is

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

where the cosine inverse is stated in radian. If one writes the complex solutions as $a \pm bi$, where $i = \sqrt{-1}$, then we have $\phi_1 = 2a, \phi_2 = -(a^2+b^2)$, and

$$k = \frac{2\pi}{\cos^{-1}(a/\sqrt{a^2+b^2})},$$

where $\sqrt{a^2+b^2}$ is the absolute value of $a \pm bi$. See Example 2.3 for an illustration.

AR(2) Model VI

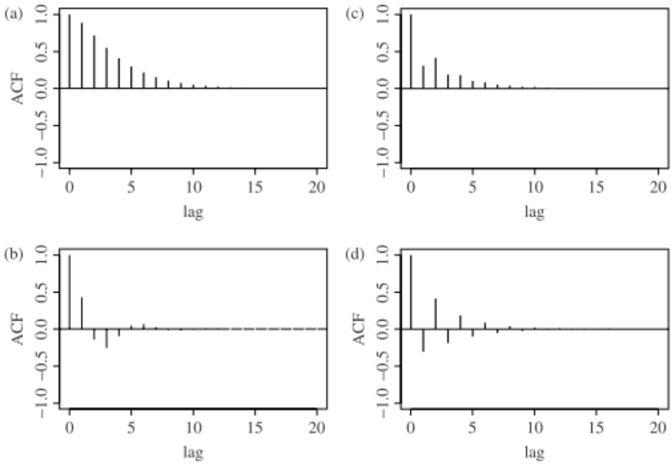

Figure 2.9. The autocorrelation function of an AR(2) model: (a) $\phi_1 = 1.2$ and $\phi_2 = -0.35$, (b) $\phi_1 = 0.6$ and $\phi_2 = -0.4$, (c) $\phi_1 = 0.2$ and $\phi_2 = 0.35$, and (d) $\phi_1 = -0.2$ and $\phi_2 = 0.35$.

AR(2) Model VII

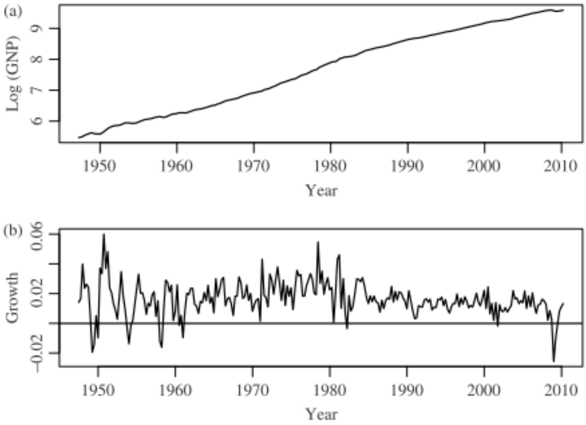

Figure 2.10. Time plots of US quarterly gross national product from 1947.I to 2010.I: (a) Log GNP series and (b) growth rate. The data are seasonally adjusted and in billions of dollars.

AR(2) Model VIII

Example 2.3. As an illustration, consider the quarterly growth rate of US gross national product (GNP), seasonally adjusted, from the second quarter of 1947 to the first quarter of 2010 for 252 observations. The log series of GNP, in billions of dollars, and its growth rate are shown in Figure 2.10. A horizontal line of zero is added to the time plot of the growth rate. The plot clearly shows that most of the growth rates are positive and the largest drop in GNP occurred in the 2008 recession.

On the basis of the model building procedure of the next section, we employ an AR(3) model for the data. The fitted model is

$$(1 - 0.438B - 0.206B^2 + 0.156B^3)(x_t - 0.016) = a_t, \quad \hat{\sigma}_a = 9.55 \times 10^{-5}. \quad (2.15)$$

The standard errors of the estimates are 0.062, 0.067, 0.063, and 0.001, respectively. See the attached R output for further information. Model (2.15) gives rise to a third-order polynomial equation

$$1 - 0.438z - 0.206z^2 + 0.156z^3 = 0,$$

AR(2) Model IX

which has three solutions, namely, $1.616 + 0.864i$, $1.616 - 0.864i$, and -1.909. The real solution corresponds to a factor $[1 - (1/-1.909)z] = (1 + 0.524z)$ that shows an exponentially decaying feature of the GNP growth rate. Focusing on the complex conjugate pair $1.616 \pm 0.864i$, we obtain the absolute value $\sqrt{1.616^2 + 0.864^2} = 1.833$ and

$$k = \frac{2\pi}{\cos^{-1}(1.616/1.833)} \approx 12.80.$$

Therefore, the fitted AR(3) model confirms the existence of business cycles in the US economy, and the average length of the cycles is 12.8 quarters, which is about 3 years. This result is reasonable as the US economy went through expansion and contraction and the length of expansion is generally believed to be around 3 years. If one uses a nonlinear model to separate US economy into "expansion" and "contraction" periods, the data show that the average duration of contraction periods is about three quarters and that of expansion periods is about 3 years; see, for instance, the analysis in Tsay (2010, Chapter 4). The average duration of 12.8 quarters is a compromise between the two separate durations. The periodic feature obtained here is common among growth rates of national economies. For example, similar features can be found for many economies in the Organization for Economic Cooperation and Development (OECD) countries.

AR(2) Model X

- Stationarity.
  - The stationarity condition of an AR(2) time series is that the absolute values of its two characteristic roots are less than 1, that is, its two characteristic roots are less than 1 in modulus.
  - Equivalently, the two solutions of the characteristic equation are greater than 1 in modulus. Under such a condition, the recursive equation ensures that the ACF of the model converges to 0 as the lag increases.
  - This convergence property is a necessary condition for a stationary time series.
  - In fact, the condition also applies to the AR(1) model, where the polynomial equation is $1 - \phi_1 z = 0$. The characteristic root is $w = 1/z = \phi_1$, which must be less than 1 in modulus for x_t to be stationary.

Outline

AR(p) Model I

- The mean of a stationary series is

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

where the denominator is not 0.

- The associated characteristic equation of the model is

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

  - If all the solutions of this equation are greater than 1 in modulus, then the series is stationary.
  - Inverses of the solutions are the characteristic roots of the model.

AR(p) Model II

- Stationarity requires that all characteristic roots are less than 1 in modulus.
- For a stationary AR(p) series, the ACF satisfies the difference equation

$$\left(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p\right)\rho_k=0, \text{ for } k>0.$$

- The plot of ACF of a stationary AR(p) model would then show a mixture of damping sine and cosine patterns and exponential decays depending on the nature of its characteristic roots.

Outline

Determining the Order of AR models I

- The order determination (or order specification) of AR models,
  - The order p of an AR time series is unknown. It must be specified empirically.
- Two general approaches are available for determining the value of p.
  - The first approach is to use the partial autocorrelation function (PACF).
  - The second approach uses some information criteria.

Partial Autocorrelation Function (PACF) I

- The PACF of a stationary time series is a function of its ACF.
- A simple, yet effective way to introduce PACF is to consider the following AR models in consecutive orders:

$$
\begin{aligned}
x_t &= \phi_{0,1} + \phi_{1,1} x_{t-1} + e_{1t} \\
x_t &= \phi_{0,2} + \phi_{1,2} x_{t-1} + \phi_{2,2} x_{t-2} + e_{2t} \\
x_t &= \phi_{0,3} + \phi_{1,3} x_{t-1} + \phi_{2,3} x_{t-2} + \phi_{3,3} x_{t-3} + e_{3t} \\
&\vdots
\end{aligned}
$$

- These models are in the form of a multiple linear regression and can be estimated by the least squares (LS) method.
- As a matter of fact, they are arranged in a sequential order that enables us to apply the idea of partial F test in multiple linear regression analysis.

Partial Autocorrelation Function (PACF) II

- The estimate $\hat{\phi}_{1,1}$ of the first equation is called the lag-1 sample PACF of x_t.
- The estimate $\hat{\phi}_{2,2}$ of the second equation is the lag-2 sample PACF of x_t.
  - It shows the added contribution of x_{t-2} to x_t over the AR(1) model.
- The estimate $\hat{\phi}_{3,3}$ of the third equation is the lag-3 sample PACF of x_t, and so on.
  - It shows the added contribution of x_{t-3} to x_t over the AR(2) model.

Partial Autocorrelation Function (PACF) III

- To **determine the order** p:
  For an AR(p) model, the lag-p sample PACF should not be 0, but $\hat{\phi}_{j,j}$ should be close to 0 for all $j > p$.
- For a stationary Gaussian AR(p) model, it can be shown that the sample PACF has the following properties
  - $\hat{\phi}_{p,p}$ converges to ϕ_p as the sample size T goes to infinity.
  - $\hat{\phi}_{k,k}$ converges to 0 for all $k > p$.
  - The asymptotic variace of $\hat{\phi}_{k,k}$ is $1/T$ for $k > p$.

  These results imply that, for an AR(p) series, the sample PACF cuts off at lag p.

Information Criteria I

- Several information criteria available to determine the order p of an AR process.
  - All of them are likelihood based.
- The well-known **Akaike Information Criterion (AIC)** (Akaike, 1973) is defined as

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

where the likelihood function is evaluated at the maximum likelihood estimates, T is the sample size.

  - The first term of the AIC measures the goodness of fit of the AR() model to the data.

Information Criteria II

- The second term is called the penalty function of the criterion because it penalizes a candidate model by the number of parameters used.

- For a Gaussian AR(k) model, AIC reduces to

$$\mathrm{AIC}(k) = \ln(\tilde{\sigma}_k^2) + \frac{2k}{T}$$

where $\tilde{\sigma}_k^2$ is the maximum likelihood estimate of σ_a^2, which is the variance of a_t.

- Another commonly used criterion function is the **Schwarz–Bayesian criterion (BIC, Bayesian information criterion)**.

Information Criteria III

- For a Gaussian AR(k) model, the criterion is

$$\text{BIC}(k) = \ln(\tilde{\sigma}_k^2) + \frac{k \ln(T)}{T}.$$

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

- Selection Rule
  - To use AIC to select an AR model in practice, one computes AIC(k) for = 0, . . . , P, where P is a prespecified positive integer and selects the order k that has the **minimum** AIC value.
  - The same rule applies to BIC.

Determine the Order I

TABLE 2.1. Sample Partial Autocorrelation Function and Some Information Criteria for the Monthly Simple Returns of CRSP Value-Weighted Index From January 1926 to December 2008

p	1	2	3	4	5	6
PACF	0.115	−0.030	−0.102	0.033	0.062	−0.050
AIC	−5.838	−5.837	−5.846	−5.845	−5.847	−5.847
BIC	−5.833	−5.827	−5.831	−5.825	−5.822	−5.818
p	7	8	9	10	11	12
PACF	0.031	0.052	0.063	0.005	−0.005	0.011
AIC	−5.846	−5.847	−5.849	−5.847	−5.845	−5.843
BIC	−5.812	−5.807	−5.805	−5.798	−5.791	−5.784

- With T = 996, the asymptotic standard error of the sample PACF is approximately 0.032. Therefore, using the 5% significant level, we identify an AR(3) or AR(9) model for the data (i.e., p = 3 or 9). If the 1% significant level is used, we specify an AR(3) model.

Determine the Order II

- The AIC values are close to each other with minimum -5.849 occurring at $p = 9$, suggesting that an AR(9) model is preferred by the criterion.
- The BIC, on the other hand, attains its minimum value -5.833 at $p = 1$ with -5.831 as a close second at $p = 3$. Thus, the BIC selects an AR(1) model for the value-weighted return series.

Parameter Estimation I

- For a specified AR(p) model, the conditional LS method, which starts with the $(p+1)$th observation, is used to estimate the parameters.
  - Specifically, conditioning on the first p observations,

$$x_t = \phi_0 + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + a_t, \quad t = p+1, \ldots T,$$

    which is in the form of a multiple linear regression and can be estimated by the LS method.
  - The fitted model is

$$\hat{x}_t = \hat{\phi}_0 + \hat{\phi}_1 x_{t-1} + \ldots + \hat{\phi}_p x_{t-p}$$

    with residual series $\hat{a}_t = x_t - \hat{x}_t$,

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

Model Checking I

- A fitted model must be examined carefully to check for possible model inadequacy.
  - If the model is adequate, then the residual series should behave as a white noise.
- The ACF and the 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)

$$Q(m) = T(T+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^{\,2}}{T-k},$$

    follows asymptotically a chi-squared distribution with $m-g$ degrees of freedom, where g denotes the number of AR coefficients used in the model.

Model Checking II

- The adjustment in the degrees of freedom is made based on the number of constraints added to the residuals $\hat{a}_t$ from fitting the AR(p) to an AR(0) model.

- If a fitted model is found to be inadequate, it must be refined.
  - If some of the estimated AR coefficients are not significantly different from 0, then the model should be simplified by removing those insignificant parameters.
  - If residual ACF shows additional serial correlations, then the model should be extended to take care of the those correlations.

Outline

Goodness of Fit I

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

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

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

$$R^2 = 1 - \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{\sum_{t=p+1}^{T} (x_t - \overline{x})^2},$$

where $\overline{x} = \sum_{t=p+1}^{T} x_t/(T-p)$. 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. However, this is only true for a stationary time series. For the unit-root nonstationary series discussed later in this chapter, R^2 of an AR(1) fit converges to 1 when the sample size increases to infinity, regardless of the true underlying model of x_t.

Goodness of Fit II

For a given data set, 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 proposed, which is defined as

$$\begin{aligned}\text{Adj}(R^2) &= 1 - \frac{\text{Variance of residuals}}{\text{Variance of } x_t}\\ &= 1 - \frac{\hat{\sigma}_a^2}{\hat{\sigma}_x^2},\end{aligned}$$

where $\hat{\sigma}_x^2$ is the sample variance of x_t. This new measure takes into account the number of parameters used in the fitted model. However, it is no longer between 0 and 1.

$$Adj(R^2) = 1 - \frac{\frac{1}{T-2p-1}\sum_{t=p+1}^{T}\hat{a}_t^{\,2}}{\frac{1}{T-1}\sum_{t=1}^{T}(x_t-\bar{x})^2}.$$

Outline

Forecasting I

Forecasting is an important application of time series analysis. For the AR(p) model in Equation (2.9), suppose that we are at the time index h and are interested in forecasting $x_{h+\ell}$, where $\ell \geq 1$. The time index h is called the *forecast origin* and the positive integer ℓ is the *forecast horizon*. Let $\hat{x}_h(\ell)$ be the forecast of $x_{h+\ell}$ using the minimum squared error loss function. In other words, the forecast $\hat{x}_k(\ell)$ is chosen such that

$$E\{[x_{h+\ell} - \hat{x}_h(\ell)]^2 | F_h\} \leq \min_g E[(x_{h+\ell} - g)^2 | F_h],$$

where g is a function of the information available at time h (inclusive), that is, a function of F_h. We referred to $\hat{x}_h(\ell)$ as the ℓ-step ahead forecast of x_t at the forecast origin h. In the prior equation, F_h denotes the collection of information available at the forecast origin h.

Forecasting II

1-Step Ahead Forecast. From the AR(p) model, we have

$$x_{h+1} = \phi_0 + \phi_1 x_h + \cdots + \phi_p x_{h+1-p} + a_{h+1}.$$

Under the minimum squared error loss function, the point forecast of x_{h+1} given F_h is the conditional expectation

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

and the associated forecast error is

$$e_h(1) = x_{h+1} - \hat{x}_h(1) = a_{h+1}.$$

Consequently, the variance of the 1-step ahead forecast error is $\mathrm{Var}[e_h(1)] = \mathrm{Var}(a_{h+1}) = \sigma_a^2$. If a_t is normally distributed, then a 95% 1-step ahead interval forecast of x_{h+1} is $\hat{x}_h(1) \pm 1.96 \times \sigma_a$. For the linear model in Equation (2.4), a_{t+1} is also the 1-step ahead forecast error at the forecast origin t. In the econometric literature, a_{t+1} is referred to as the *shock* to the series at time $t+1$.

Forecasting III

2-Step Ahead Forecast. Next, consider the forecast of x_{h+2} at the forecast origin h. From the AR(p) model, we have

$$x_{h+2} = \phi_0 + \phi_1 x_{h+1} + \cdots + \phi_p x_{h+2-p} + a_{h+2}.$$

Taking conditional expectation, we have

$$\hat{x}_h(2) = E(x_{h+2}|F_h) = \phi_0 + \phi_1 \hat{x}_h(1) + \phi_2 x_h + \cdots + \phi_p x_{h+2-p}$$

and the associated forecast error

$$e_h(2) = x_{h+2} - \hat{x}_h(2) = \phi_1[x_{h+1} - \hat{x}_h(1)] + a_{h+2} = a_{h+2} + \phi_1 a_{h+1}.$$

The variance of the forecast error is $\mathrm{Var}[e_h(2)] = (1 + \phi_1^2)\sigma_a^2$. Interval forecasts of x_{h+2} can be computed in the same way as those for x_{h+1}. It is interesting to see that $\mathrm{Var}[e_h(2)] \geq \mathrm{Var}[e_h(1)]$, meaning that as the forecast horizon increases the uncertainty in forecast also increases. This is in agreement with common sense that we are more uncertain about x_{h+2} than x_{h+1} at the time index h for a linear time series.

Forecasting IV

Multistep Ahead Forecast. In general, we have

$$x_{h+\ell} = \phi_0 + \phi_1 x_{h+\ell-1} + \cdots + \phi_p x_{h+\ell-p} + a_{h+\ell}.$$

The ℓ-step ahead forecast based on the minimum squared error loss function is the conditional expectation of $x_{h+\ell}$ given F_h, which can be obtained as

$$\hat{x}_h(\ell) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{x}_h(\ell - i),$$

where it is understood that $\hat{x}_h(i) = x_{h+i}$ if $i \le 0$. This forecast can be computed recursively using forecasts $\hat{x}_h(i)$ for $i = 1, \ldots, \ell - 1$. The ℓ-step ahead forecast error is $e_h(\ell) = x_{h+\ell} - \hat{x}_h(\ell)$. It can be shown that for a stationary AR(p) model, $\hat{x}_h(\ell)$ converges to $E(x_t)$ as $\ell \to \infty$, meaning that for such a series long-term point forecast approaches its unconditional mean. This property is referred to as the *mean reversion* in the finance literature. For an AR(1) model, the speed of mean reversion is measured by the *half-life* defined as $\ell = \ln(0.5)/\ln(|\phi_1|)$. The variance of the forecast error then approaches the unconditional variance of x_t. Note that for an AR(1) model in Equation (2.8), let $x_t = x_t - E(x_t)$ be the mean-adjusted series. It is easy to see that the ℓ-step ahead forecast of $x_{h+\ell}$ at the forecast origin h is $\hat{x}_h(\ell) = \phi_1^{\ell} x_h$. The half-life is the forecast horizon such that $\hat{x}_h(\ell) = \frac{1}{2} x_h$. That is, $\phi_1^{\ell} = \frac{1}{2}$. Thus, $\ell = \ln(0.5)/\ln(|\phi_1|)$.

Example I

Table 2.2 contains the 1-step to 12-step ahead forecasts and the standard errors of the associated forecast errors at the forecast origin 984 for the monthly simple return of the value-weighted index using an AR(3) model that was reestimated using the first 984 observations. The fitted model is

$$x_t = 0.0098 + 0.1024x_{t-1} - 0.0201x_{t-2} - 0.1090x_{t-3} + a_t,$$

where $\hat{\sigma}_a = 0.054$. The actual returns of 2008 are also given in the table. Because of the weak serial dependence in the series, the forecasts and standard deviations of forecast errors converge to the sample mean and standard deviation of the data quickly. For the first 984 observations, the sample mean and standard error are 0.0095 and 0.0540, respectively.

Example II

TABLE 2.2. Multistep Ahead Forecasts of an AR(3) Model For The Monthly Simple Returns of CRSP Value-Weighted Index[a]

Step	1	2	3	4	5	6
Forecast	0.0076	0.0161	0.0118	0.0099	0.0089	0.0093
Standard Error	0.0534	0.0537	0.0537	0.0540	0.0540	0.0540
Actual	−0.0623	−0.0220	−0.0105	0.0511	0.0238	−0.0786
Step	**7**	**8**	**9**	**10**	**11**	**12**
Forecast	0.0095	0.0097	0.0096	0.0096	0.0096	0.0096
Standard Error	0.0540	0.0540	0.0540	0.0540	0.0540	0.0540
Actual	−0.0132	0.0110	−0.0981	−0.1847	−0.0852	0.0215

[a] The forecast origin is 984.

Example III

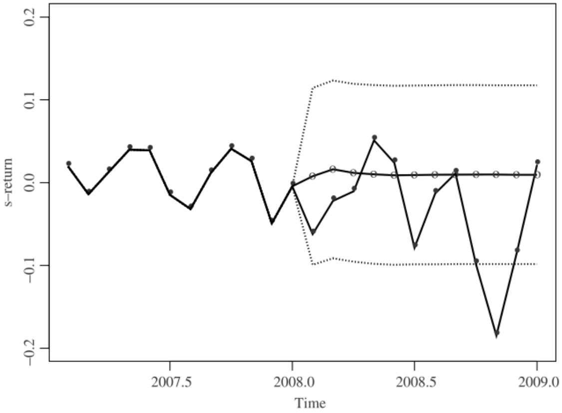

Figure 2.13. Plot of 1-step to 12-step ahead out-of-sample forecasts for the monthly simple return of the CRSP value-weighted index. The forecast origin is $t = 984$, which is December 2007. The forecasts are denoted by "o" and the actual observations by "•." The two dashed lines denote two standard error limits of the forecasts.

Example IV

Figure 2.13 shows the corresponding out-of-sample prediction plot for the monthly simple return series of the value-weighted index. The forecast origin $t = 984$ corresponds to December 2007. The prediction plot includes the two standard error limits of the forecasts and the actual observed returns for 2008. The forecasts and actual returns are marked by "o" and "•," respectively. From the plot, except for the return of October 2008, all actual returns are within the 95% prediction intervals.