Topic 4: MA and ARMA Models

Zongye Huang

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

March 20, 2017

Outline

Moving Average Models
- MA Models
- Estimate MA Model
- Forecasting Using MA Models
- Summary

ARMA Model
- ARMA Model
- Identifying ARMA Models
- Forecasting Using ARMA Model
- Three Representations for an ARMA Model

Outline

Moving Average Models
- MA Models
- Estimate MA Model
- Forecasting Using MA Models
- Summary

ARMA Model
- ARMA Model
- Identifying ARMA Models
- Forecasting Using ARMA Model
- Three Representations for an ARMA Model

MA Model I

- Two ways to introduce MA models
  - Treat the model as a simple extension of white noise series.
  - Treat the model as an infinite-order AR model, with some parameter constraints.
- There is no particular reason, but simplicity, to assume a priori that the order of an AR model is finite. We may entertain, at least in theory, an AR model with infinite order

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

  - Such an AR model is not realistic because it has infinite parameters.

MA Model II

- One way to make the model practical is to assume that the coefficients ϕ_i satisfy some constraints so that they are determined by a finite number of parameters.
  - A special case of this idea is

    $$x_t = \phi_0 - \theta_1 x_{t-1} - \theta_1^2 x_{t-2} - \theta_1^3 x_{t-3} - \ldots + a_t.$$

    the coefficients depend on a single parameter θ_1 via $\phi_i = -\theta_1^i$, for $i > 0$.
  - For this model to be stationary, θ_1 must be less than 1 in absolute value.
    - The contribution of x_{t-i} decays exponentially as i increases.

MA Model III

- Rewrite the model in a rather compact form

$$x_t + \theta_1 x_{t-1} + \theta_1^2 x_{t-2} + \theta_1^3 x_{t-3} + \ldots = \phi_0 + a_t.$$

  - The model for x_{t-1} is

$$x_{t-1} + \theta_1 x_{t-2} + \theta_1^2 x_{t-3} + \theta_1^3 x_{t-4} + \ldots = \phi_0 + a_{t-1}.$$

  - Multiplying by $\theta_1$and substracting the x_t model,

$$x_t = (1 - \theta_1)\phi_0 + a_t - \theta_1 a_{t-1}.$$

  - x_t is a weighted averageof shocks a_t and a_{t-1}.

MA Model IV

- A MA(1) model is

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

  where c_0 is a constant and $\{a_t\}$ is a white noise series.
- A MA(q) model is

$$
\begin{aligned}
x_t &= c_0 + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}. \\
&= c_0 + (1 - \theta_1 B - \ldots - \theta_q B^q) a_t
\end{aligned}
$$

  where $q > 0$.

Stationarity I

- MA models are always weakly stationary because they are finite linear combinations of a white noise sequence for which the first two moments are time invariant.
  - For example, consider the MA(1),

$$x_t = c_0 + a_t - \theta_1 a_{t-1}$$

Taking expectation of the model,

$$E(x_t) = c_0.$$

Taking the variance

$$Var(x_t) = (1 + \theta_1^2)\sigma_a^2.$$

Stationarity II

- For a general MA(q), the mean is

$$E(x_t) = c_0.$$

and the variance is

$$Var(x_t) = (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma_a^2.$$

Autocorrelation Function I

- Assume for simplicity that $c_0 = 0$ for an MA(1) model

$$x_t = a_t - \theta_1 a_{t-1}.$$

Multiplying the model by x_{t-k},

$$x_{t-k} x_t = x_{t-k} a_t - \theta_1 x_{t-k} a_{t-1}.$$

Taking expectation, we obtain

$$\gamma_1 = -\theta_1 \sigma_a^2 \text{ and } \gamma_k = 0, \text{ for } k > 1.$$

$$\rho_0 = 1,\ \rho_1 = \frac{-\theta_1}{1+\theta_1^2} \text{ and } \rho_k = 0, \text{ for } k > 1.$$

Autocorrelation Function II

- For an MA(1) model, the lag-1 ACF is not 0, but all higher-order ACFs are 0.
  - In other words, the ACF of an MA(1) model cuts off at lag 1.
- For the MA(2) model,

$$x_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2},$$

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, \ \rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, \text{ and } \rho_k = 0, \text{ for } k > 2.$$

  - The ACF cuts off at lag 2.
- For a general MA(q) model, the lag-q ACF is not 0, but $\rho_k = 0$ for $k > q$.
  - MA(q) is only linearly related to its first q lagged values and is a “finite memory” model.

Invertibility

- Rewriting a zero-mean MA(1) model,

$$a_t = x_t + \theta_1 a_{t-1},$$

one can use repeated substitutions to obtain

$$a_t = x_t + \theta_1 x_{t-1} + \theta_1^2 x_{t-2} + \ldots$$

- This equation expresses the current shock a_t as a linear combination of the present and past values of x_t.
  - Intuitively, θ_1^j should go to 0 as j increases because the remote return x_{t-j} should have very little impact on the current shock, if any.
  - Consequently, for an MA(1) model to be plausible, we require $|\theta_1| < 1$.
    - Such an MA(1) model is said to be invertible.
  - If $|\theta_1| = 1$, then the MA(1) model is noninvertible.

Outline

Moving Average Models
- MA Models
- Estimate MA Model
- Forecasting Using MA Models
- Summary

ARMA Model
- ARMA Model
- Identifying ARMA Models
- Forecasting Using ARMA Model
- Three Representations for an ARMA Model

Identifying MA Order I

- The ACF is useful in identifying the order of an MA model.
  - For a time series x_t with ACF ρ_k, if $\rho_q \neq 0$, but $\rho_k = 0$ for $k > q$, then x_t follows an MA(q) model.
- Sample ACF provides information on the nonzero MA lags of the model.
  - For example, a simple MA(2) model with $\theta_1 = 0$, thus

$$x_t = c_0 + a_t - \theta_2 a_{t-2}.$$

The ACF is

$$\rho_0 = 1,\ \rho_1 = 0,\ \rho_2 = \frac{-\theta_2}{1+\theta_2^2} \text{ and } \rho_k = 0, \text{ for } k > 2$$

which provides the exact information on the structure of the model.

Estimation I

- Maximum likelihood estimation is commonly used to estimate MA models.
- There are two approaches for evaluating the likelihood function of an MA model.
    - The first approach assumes that the initial shocks (i.e., a_t for $t \leq 0$) are 0. As such, the shocks needed in likelihood function calculation are obtained recursively from the model, starting with $a_1 = x_1 - c_0$ and $a_2 = x_2 - c_0 + \theta_1 a_1$.
        - This approach is referred to as the *conditional likelihood method* and the resulting estimates the conditional maximum likelihood estimates.

Estimation II

- The second approach treats the initial shocks a_t, t ≤ 0 as additional parameters of the model and estimate them jointly with other parameters. This approach is referred to as the *exact likelihood method*.
  - The exact likelihood estimates are preferred over the conditional ones, especially when the MA model is close to being noninvertible.
  - The exact method, however, requires more intensive computation.

If the sample size is large, then the two types of maximum likelihood estimates are close to each other.

Estimation III

Remark. R uses the exact likelihood method in estimation. In addition, the MA polynomial is written as $1 + \theta_1 B + \cdots + \theta_q B^q$ instead of the conventional parameterization $1 - \theta_1 B - \cdots - \theta_q B^q$. More specifically, the ARMA(p, q) model under the R command `arima` is in the form

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(x_t - \mu) = (1 + \theta_1 B + \cdots + \theta_q B^q)a_t,$$

where μ is referred to as the `intercept`. See the attached R output. □

Outline

Moving Average Models
- MA Models
- Estimate MA Model
- Forecasting Using MA Models
- Summary

ARMA Model
- ARMA Model
- Identifying ARMA Models
- Forecasting Using ARMA Model
- Three Representations for an ARMA Model

Forecasting MA(1) Model I

- Forecasts of an MA model can easily be obtained. Because the model has finite memory, its point forecasts go to the mean of the series quickly.
  - Assume that the forecast origin is h, and let F_h denote the information available at time h.
- For the 1-step ahead forecast of an MA(1) process, the model says

  $$x_{h+1} = c_0 + a_{h+1} - \theta_1 a_h.$$

  Taking the conditional expectation

  $$\begin{aligned} \hat{x}_h(1) &= E(x_{h+1}|F_h) = c_0 - \theta_1 a_h \\ e_h(1) &= x_{h+1} - \hat{x}_h(1) = a_{h+1} \end{aligned}$$

Forecasting MA(1) Model II

The variance of the 1-step ahead forecast error

$$Var\left[e_h(1)\right] = \sigma_a^2.$$

- In practice, the quantity a_h can be obtained in several ways.
  - Assume that $a_0 = 0$, then $a_1 = x_1 - c_0$, and we can compute a_t for $2 \leq t \leq h$ recursively by $a_t = x_t - c_0 + \theta_1 a_{t-1}$.
  - Alternatively, it can be computed by the AR representation of the MA(1) model.

Forecasting MA(1) Model III

- For the 2-step ahead forecast from the equation

$$x_{h+2} = c_0 + a_{h+2} - \theta_1 a_{h+1}.$$

  We have

$$\begin{aligned} \hat{x}_h(2) &= E(x_{h+2}|F_h) = c_0, \\ e_h(2) &= x_{h+2} - \hat{x}_h(2) = a_{h+2} + \theta_1 a_{h+1}. \end{aligned}$$

  The variance of the 2-step ahead forecast error

$$Var[e_h(2)] = (1 + \theta_1^2)\sigma_a^2.$$

  which is greater than or equal to $Var[e_h(1)]$.
- For an MA(1) model, $\hat{x}_h(k) = c_0$, for $k > 1$.
    - Mean reverting only takes one time period.

Forecasting MA(q) Model I

- In MA(2) model,

$$x_{h+k} = c_0 + a_{h+k} - \theta_1 a_{h+k-1} - \theta_2 a_{h+k-2}.$$

  We have

$$\begin{aligned}
\hat{x}_h(1) &= c_0 - \theta_1 a_h - \theta_2 a_{h-1}, \\
\hat{x}_h(2) &= c_0 - \theta_2 a_h, \\
\hat{x}_h(k) &= c_0, \text{ for } k > 2.
\end{aligned}$$

  - The multistep ahead forecasts of an MA(2) model go to the mean of the series after two steps.
  - The variances of forecast errors go to the variance of the series after two steps.

- In general, for an MA(q) model, multistep ahead forecasts go to the mean after the first q steps.

Outline

Moving Average Models
- MA Models
- Estimate MA Model
- Forecasting Using MA Models
- Summary

ARMA Model
- ARMA Model
- Identifying ARMA Models
- Forecasting Using ARMA Model
- Three Representations for an ARMA Model

Summary of AR and MA models I

- For MA models, ACF is useful in specifying the order because ACF cuts off at lag q for an MA(q) series;
- For AR models, PACF is useful in order determination because PACF cuts off at lag p for an AR(p) process;
- MA series is always stationary.
- For an AR series to be stationary, all of its characteristic roots must be less than 1 in modulus;
- For a stationary series, as the forecast horizon increases
  - the multistep ahead forecasts converge to the mean of the series,
  - the variances of forecast errors converge to the variance of the series.

Outline

Moving Average Models
- MA Models
- Estimate MA Model
- Forecasting Using MA Models
- Summary

ARMA Model
- ARMA Model
- Identifying ARMA Models
- Forecasting Using ARMA Model
- Three Representations for an ARMA Model

ARMA Model I

- In some applications, the AR or MA models discussed in the previous sections become cumbersome because one may need a high order model with many parameters to adequately describe the dynamic structure of the data.
- An ARMA model combines the ideas of AR and MA models into a compact form so that the number of parameters used is kept small, achieving parsimony in parameterization.
  - The model is useful in modeling business, economic, and engineering time series.
  - For the return series in finance, the chance of using ARMA models is low.
  - The concept of ARMA models is highly relevant in volatility modeling.

ARMA(1,1) I

- A time series x_t follows an ARMA(1,1) model if it satisfies

$$x_t - \phi_1 x_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1},$$

where $a_t$is a white noise series.

  - The left-hand side of the Equation is the AR component of the model.
  - The right-hand side gives the MA component.
  - The constant term is ϕ_0.
  - For this model to be meaningful, we need $\phi_1 \neq \theta_1$; otherwise, there is a cancellation in the equation and the process reduces to a white noise series.

ARMA(1,1) II

- Taking expectation

$$E(x_t) - \phi_1 E(x_{t-1}) = \phi_0 + E(a_t) - \theta_1 E(a_{t-1}),$$

the mean is

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

This is exactly the same as that of the AR(1) model.

- For simplicity, assume $\phi_0 = 0$.

$$x_t = \phi_1 x_{t-1} + a_t - \theta_1 a_{t-1}.$$

The variance is

$$Var(x_t) = \phi_1^2 Var(x_{t-1}) + \sigma_a^2 + \theta_1^2 \sigma_a^2 - 2\phi_1 \theta_1 E(x_{t-1} a_{t-1}).$$

ARMA(1,1) III

Since $E(x_{t-1} a_{t-1}) = \sigma_a^2$, we have

$$Var(x_t) = \frac{(1 + \theta_1^2 - 2\phi_1 \theta_1)\sigma_a^2}{1 - \phi_1^2}.$$

We need $\phi_1^2 < 1$. Again, this is precisely the same stationarity condition as that of the AR(1) model.

Autocorrelation I

- To obtain the autocovariance function, we assume that $\phi_0 = 0$. and multiply the model by x_{t-k} to obtain

$$x_t x_{t-k} - \phi_1 x_{t-1} x_{t-k} = a_t x_{t-k} - \theta_1 a_{t-1} x_{t-k}.$$

  - For $k = 1$, taking expectation

$$\gamma_1 - \phi_1 \gamma_0 = -\theta_1 \sigma_a^2.$$

    which is different from that of the AR(1) case($\gamma_1 - \phi_1 \gamma_0 = 0$).
  - For $k = 2$,

$$\gamma_2 - \phi_1 \gamma_1 = 0.$$

    which is identical to that of the AR(1) case.
  - For $k > 1$,

$$\gamma_k - \phi_1 \gamma_{k-1} = 0.$$

Autocorrelation II

- The ACF of an ARMA(1,1),

$$\rho_1 = \phi_1 - \frac{\theta_1 \sigma_a^2}{\gamma_0}, \ \rho_k = \phi_1 \rho_{k-1}, \ \text{for } k > 1$$

.

  - The ACF of an ARMA(1,1) model behaves very much similar to that of an AR(1) model except that the exponential decay starts with lag 2.
  - The ACF of an ARMA(1,1) model does not cut off at any finite lag.

Autocorrelation III

- Turning to PACF, one can show that the PACF of an ARMA(1,1) model does not cut off at any finite lag either.
  - It behaves very much similar to that of an MA(1) model except that the exponential decay starts with lag 2 instead of lag 1.
- In summary, the stationarity condition of an ARMA(1,1) model is the same as that of an AR(1) model, and the ACF of an ARMA(1,1) exhibits a pattern similar to that of an AR(1) model except that the pattern starts at lag 2.

General ARMA Models I

- A General ARMA model,

$$x_t = \phi_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + a_t - \sum_{i=1}^{q} \theta_i a_{t-i}.$$

  - The AR and MA models are special cases of the ARMA(p, q) model.
  - Using the backshift operator,

$$(1 - \phi_1 B - ... - \phi_p B^p) x_t = \phi_0 + (1 - \theta_1 B - ... - \theta_q B^q) a_t.$$

  - We require that there are no common factors between the AR and MA polynomials; otherwise, the order (p, q) of the model can be reduced.

General ARMA Models II

- Similar to a pure AR model, the AR polynomial introduces the characteristic equation of an ARMA model.
  - If all of the characteristic roots are less than 1 in absolute value, then the ARMA model is weakly stationary.
  - In this case, the unconditional mean of the model is

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

Outline

Moving Average Models
- MA Models
- Estimate MA Model
- Forecasting Using MA Models
- Summary

ARMA Model
- ARMA Model
- Identifying ARMA Models
- Forecasting Using ARMA Model
- Three Representations for an ARMA Model

EACF I

- The ACF and PACF are not informative in determining the order of an ARMA model.
- Tsay and Tiao (1984) propose a new approach that uses the extended autocorrelation function (EACF) to specify the order of an ARMA process.
    - The basic idea of EACF is relatively simple.
        - If we can obtain a consistent estimate of the AR component of an ARMA model, then we can derive the MA component.
        - From the derived MA series, we can use ACF to identify the order of the MA component.
    - The output of EACF is a two-way table, where the rows correspond to AR order p and the columns to MA order q.
        - The theoretical version of EACF for an ARMA(1,1) model is given in the table.

EACF II

- The key feature of the table is that it contains a triangle of "O" with the upper left vertex located at the order (1,1).
- This is the characteristic we use to identify the order of an ARMA process.
- In general, for an ARMA(p, q) model, the triangle of "O" will have its upper left vertex at the (p, q) position.

TABLE 2.4. Theoretical EACF Table For an ARMA(1,1) Model, Where "X" Denotes Nonzero, "O" Denotes Zero, and "*" Denotes Either Zero or Nonzero[a]

	MA							
AR	0	1	2	3	4	5	6	7
0	X	X	X	X	X	X	X	X
1	X	O	O	O	O	O	O	O
2	*	X	O	O	O	O	O	O
3	*	*	X	O	O	O	O	O
4	*	*	*	X	O	O	O	O
5	*	*	*	*	X	O	O	O

Using Information Criteria

- The information criteria discussed earlier can also be used to select the order of an ARMA model.
    - For some prespecified positive integers P and Q, one computes AIC (or BIC) for ARMA(p, q) models, where $0 \leq p \leq P$ and $0 \leq q \leq Q$, and selects the model that gives the minimum AIC (or BIC).
    - Once an ARMA(p, q) model is specified, its parameters can be estimated by either the conditional or exact likelihood method.
        - The Ljung–Box statistics of the residuals can be used to check the adequacy of a fitted model. If the model is correctly specified, then Q(m) follows asymptotically a chi-squared distribution with $m - g$ degrees of freedom, where g denotes the number of AR or MA coefficients fitted in the model.

Outline

Moving Average Models
- MA Models
- Estimate MA Model
- Forecasting Using MA Models
- Summary

ARMA Model
- ARMA Model
- Identifying ARMA Models
- Forecasting Using ARMA Model
- Three Representations for an ARMA Model

Forecasting ARMA Model

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

and the associated forecast error is $e_h(1) = x_{h+1} - \hat{x}_h(1) = a_{h+1}$. The variance of 1-step ahead forecast error is $\mathrm{Var}[e_h(1)] = \sigma_a^2$. For the ℓ-step ahead forecast, we have

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

where it is understood that $\hat{x}_h(\ell - i) = x_{h+\ell-i}$ if $\ell - i \leq 0$ and $a_h(\ell - i) = 0$ if $\ell - i > 0$ and $a_h(\ell - i) = a_{h+\ell-i}$ if $\ell - i \leq 0$. Thus, the multistep ahead forecasts of an ARMA model can be computed recursively. The associated forecast error is

$$e_h(\ell) = x_{h+\ell} - \hat{x}_h(\ell),$$

Outline

Moving Average Models
- MA Models
- Estimate MA Model
- Forecasting Using MA Models
- Summary

ARMA Model
- ARMA Model
- Identifying ARMA Models
- Forecasting Using ARMA Model
- Three Representations for an ARMA Model

Three Representations I

The first representation of the ARMA(p, q) model is

$$(1-\phi_1 B-\ldots-\phi_p B^p)x_t=\phi_0+(1-\theta_1 B-\ldots-\theta_q B^q)a_t.$$

- This representation is compact and useful in parameter estimation.
- It is also useful in computing recursively multistep ahead forecasts of x_t.

Three Representations II

For the other two representations, we use long division of two polynomials. Given two polynomials $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$, we can obtain, by long division, that

$$\frac{\theta(B)}{\phi(B)} = 1 + \psi_1 B + \psi_2 B^2 + \cdots \equiv \psi(B) \tag{2.30}$$

and

$$\frac{\phi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B^2 - \cdots \equiv \pi(B). \tag{2.31}$$

For instance, if $\phi(B) = 1 - \phi_1 B$ and $\theta(B) = 1 - \theta_1 B$, then

$$\psi(B) = \frac{1 - \theta_1 B}{1 - \phi_1 B} = 1 + (\phi_1 - \theta_1)B + \phi_1(\phi_1 - \theta_1)B^2 + \phi_1^2(\phi_1 - \theta_1)B^3 + \cdots$$

$$\pi(B) = \frac{1 - \phi_1 B}{1 - \theta_1 B} = 1 - (\phi_1 - \theta_1)B - \theta_1(\phi_1 - \theta_1)B^2 - \theta_1^2(\phi_1 - \theta_1)B^3 - \cdots.$$

Three Representations III

AR Representation. Using the result of long division in Equation (2.31), the ARMA(p, q) model can be written as

$$x_t = \frac{\phi_0}{1 - \theta_1 - \cdots - \theta_q} + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \pi_3 x_{t-3} + \cdots + a_t. \tag{2.32}$$

This representation shows the dependence of the current return x_t on the past returns x_{t-i}, where $i > 0$. The coefficients $\{\pi_i\}$ are referred to as the π-*weights* of an ARMA model. To show that the contribution of the lagged value x_{t-i} to x_t is diminishing as i increases, the π_i coefficient should decay to 0 as i increases. An ARMA(p, q) model that has this property is said to be invertible. For a pure AR model, $\theta(B) = 1$ so that $\pi(B) = \phi(B)$, which is a finite-degree polynomial. Thus, $\pi_i = 0$ for $i > p$, and the model is invertible. For other ARMA models, a sufficient condition for invertibility is that all the zeros of the polynomial $\theta(B)$ are greater than unity in modulus. For example, consider the MA(1) model $x_t = (1 - \theta_1 B)a_t$. The zero of the first-order polynomial $1 - \theta_1 B$ is $B = 1/\theta_1$. Therefore, an MA(1) model is invertible if $|1/\theta_1| > 1$. This is equivalent to $|\theta_1| < 1$.

From the AR representation in Equation (2.32), an invertible ARMA(p, q) series x_t is a linear combination of the current shock a_t and a weighted average of the past values. The weights decay exponentially for more remote past values.

Three Representations IV

MA Representation. Again, using the result of long division in Equation (2.30), an ARMA(p, q) model can also be written as

$$x_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \mu + \psi(B) a_t, \tag{2.33}$$

where $\mu = E(x_t) = \phi_0/(1 - \phi_1 - \cdots - \phi_p)$. This representation shows explicitly the impact of the past shock a_{t-i} $(i > 0)$ on the current return x_t. The coefficients $\{\psi_i\}$ are referred to as the *impulse response function* of the ARMA model. For a weakly stationary series, the ψ_i coefficients decay exponentially as i increases. This is understandable as the effect of shock a_{t-i} on the return x_t should diminish over time. Thus, for a stationary ARMA model, the shock a_{t-i} does not have a permanent impact on

Three Representations V

the series. If $\phi_0 \neq 0$, then the MA representation has a constant term, which is the mean of x_t (i.e., $\phi_0/(1-\phi_1-\cdots-\phi_p)$).

The MA representation in Equation (2.33) is also useful in computing the variance of a forecast error. At the forecast origin h, we have the shocks $a_h, a_{h-1}, \ldots$ Therefore, the ℓ-step ahead point forecast is

$$\hat{x}_h(\ell) = \mu + \psi_\ell a_h + \psi_{\ell+1} a_{h-1} + \cdots, \tag{2.34}$$

and the associated forecast error is

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

Consequently, the variance of ℓ-step ahead forecast error is

$$\mathrm{Var}[e_h(\ell)] = (1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2)\sigma_a^2, \tag{2.35}$$

which, as expected, is a nondecreasing function of the forecast horizon ℓ.

Three Representations VI

Finally, the MA representation in Equation (2.33) provides a simple proof of mean reversion of a stationary time series. The stationarity implies that ψ_i approaches 0, as $i \to \infty$. Therefore, by Equation (2.34), we have $\hat{x}_h(\ell) \to \mu$, as $\ell \to \infty$. Because $\hat{x}_h(\ell)$ is the conditional expectation of $x_{h+\ell}$ at the forecast origin h, the result says that in the long term, the return series is expected to approach its mean, that is, the series is mean reverting. Furthermore, using the MA representation in Equation (2.33), we have $\text{Var}(x_t) = (1 + \sum_{i=1}^{\infty} \psi_i^2)\sigma_a^2$. Consequently, by Equation (2.35), we have $\text{Var}[e_h(\ell)] \to \text{Var}(x_t)$, as $\ell \to \infty$. The speed by which $\hat{r}_h(\ell)$ approaches μ determines the speed of mean reverting.