时间序列分析

Time Series Analysis

Lecture 5

Review

1. Linear Time Series
2. Simple AR Models
3. Simple MA Models
4. Simple ARMA Models

Today's Topics

1. The Autocovariance-Generating Function

2. Sums of *ARMA* processes

3. Wold's Decomposition and the Box-Jenkins Modeling Philosophy

4. Maximum Likelihood Estimation

Today's Topics

1. The Autocovariance-Generating Function
   - The Autocovariance-Generating Function
   - Filters

2. Sums of *ARMA* processes

3. Wold's Decomposition and the Box-Jenkins Modeling Philosophy

4. Maximum Likelihood Estimation

1. The Autocovariance-Generating Function

For a covariance-stationary process Y_t, we define the *autocovariance-generating function (AGF)*:

$$g_Y(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j,$$

where $\{\gamma_j\}_{j=-\infty}^{\infty}$ is the sequence of autocovariances (if it is absolutely summable), and z is a complex scalar.

If two different processes share the same autocovariance-generating function, then the two processes exhibit the identical sequence of autocovariances.

Example 1: Calculating the *AGF* for an *MA* process

From Lecture 4, an $MA(1)$ process $Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$ has *AGF*

$$\begin{aligned} g_Y(z) &= [\theta\sigma^2]z^{-1} + [(1+\theta^2)\sigma^2]z^0 + [\theta\sigma^2]z^1 \\ &= \sigma^2[\theta z^{-1} + (1+\theta^2)\theta z] \\ &= \sigma^2(1+\theta z)(1+\theta z^{-1}). \end{aligned}$$

The form suggests that for the $MA(q)$ process,

$$Y_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t,$$

the autocovariance-generaing function might be calculated as

$$\begin{aligned} g_Y(z) = \sigma^2&(1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q) \\ &\times (1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \cdots + \theta_q z^{-q}). \end{aligned}$$

Proof (self-reading)

This conjecture can be converted to

$$
\begin{aligned}
&(1+\theta_1 z+\theta_2 z^2+\cdots+\theta_q z^q)\times(1+\theta_1 z^{-1}+\theta_2 z^{-2}+\cdots+\theta_q z^{-q})\\
&\quad=(\theta_q)z^q+(\theta_{q-1}+\theta_q\theta_1)z^{q-1}+(\theta_{q-2}+\theta_{q-1}\theta_1+\theta_q\theta_2)z^{q-2}\\
&\quad+\cdots+(\theta_1+\theta_2\theta_1+\theta_3\theta_2+\cdots+\theta_q\theta_{q-1})z^1\\
&\quad+(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2)z^0\\
&\quad+(\theta_1+\theta_2\theta_1+\theta_3\theta_2+\cdots+\theta_q\theta_{q-1})z^{-1}+\cdots+\theta_q z^{-q}
\end{aligned}
$$

Extension to $MA(\infty)$ case

This method for finding $g_Y(z)$ extends to the $MA(\infty)$ case. If

$$Y_t = \mu + \psi(L)\varepsilon_t$$

with

$$\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots$$

and

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

then

$$g_Y(z) = \sigma^2 \psi(z)\psi(z^{-1}).$$

Example 2: Calculating the AGF for an $AR(1)$ process

The stationary $AR(1)$ process can be written as

$$Y_t - \mu = (1 - \phi L)^{-1} \varepsilon_t.$$

The AGF for an $AR(1)$ process could therefore be calculated from

$$g_Y(z) = \frac{\sigma^2}{(1 - \phi z)(1 - \phi z^{-1})}$$

Verify this claim:

$$\begin{aligned} \frac{\sigma^2}{(1 - \phi z)(1 - \phi z^{-1})} &= \sigma^2 (1 + \phi z + \phi^2 z^2 + \phi^3 z^3 + \cdots) \\ &\quad \times (1 + \phi z^{-1} + \phi^2 z^{-2} + \phi^3 z^{-3} + \cdots), \end{aligned}$$

from which the coefficient on z^j is

$$\gamma_j = \sigma^2 (\phi^j + \phi^{j+1}\phi + \phi^{j+2}\phi^2 + \cdots) = \sigma^2 \phi^j / (1 - \phi^2).$$

Example 3: Calculating the *AGF* for an *ARMA*(p, q) process

For a stationary *ARMA*(p, q) process

$$\begin{aligned} Y_t &= c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} \\ &+ \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}, \end{aligned}$$

or,

$$\begin{aligned} (1 - \phi_1 L &- \ldots - \phi_p L^p) Y_t = \\ c &+ (1 + \theta_1 L + \ldots + \theta_q L^q) \varepsilon_t, \end{aligned}$$

its autocovariance-generating function can be shown as

$$g_Y(z) = \frac{\sigma^2 (1+\theta_1 z+\theta_2 z^2+\cdots+\theta_q z^q)(1+\theta_1 z^{-1}+\theta_2 z^{-2}+\cdots+\theta_q z^{-q})}{(1-\phi_1 z-\phi_2 z^2-\cdots-\phi_p z^p)(1-\phi_1 z^{-1}-\phi_2 z^{-2}-\cdots-\phi_p z^{-p})}.$$

Today's Topics

Filters

Suppose that the original data Y_t follows an $MA(1)$ process,

$$Y_t = (1 + \theta L)\varepsilon_t.$$

But we're interested in analyzing X_t, representing the change in Y_t:

$$X_t = Y_t - Y_{t-1} = (1 - L)Y_t = [1 + (\theta - 1)L - \theta L^2]\varepsilon_t \equiv [1 + \theta_1 L + \theta_2 L^2]\varepsilon_t,$$

with $\theta_1 \equiv (\theta - 1)$ and $\theta_2 \equiv -\theta$.

The AGF of X_t can be calculated as:

$$\begin{aligned} g_X(z) &= \sigma^2(1 + \theta_1 z + \theta_2 z^2)(1 + \theta_1 z^{-1} + \theta_2 z^{-2}) \\ &= \sigma^2(1 - z)(1 + \theta z)(1 - z^{-1})(1 + \theta z^{-1}) \\ &= (1 - z)(1 - z^{-1}) \cdot g_Y(z). \end{aligned}$$

Generally, let $\{Y_t\}$ satisfy

$$Y_t = \mu + \psi(L)\varepsilon_t,$$

with $\psi(L) = \sum_{j=-\infty}^{\infty} \psi_j L^j, \sum_{j=0}^{\infty} |\psi_j| < \infty.$

Let's say the data are filtered according to

$$X_t = h(L)Y_t = h(1)\mu + h(L)\psi(L)\varepsilon_t \equiv \mu^* + \psi^*(L)\varepsilon_t, .$$

with $h(L) = \sum_{j=-\infty}^{\infty} h_j L^j, \sum_{j=-\infty}^{\infty} |h_j| < \infty, \mu^* \equiv h(1)\mu$ and $\psi^*(L) \equiv h(L)\psi(L).$

The AGF of X_t can accordingly be calculated as

$$\begin{aligned} g_X(z) &= \sigma^2\psi^*(z)\psi^*(z^{-1}) = \sigma^2 h(z)\psi(z)\psi(z^{-1})h(z^{-1}) \\ &= h(z)h(z^{-1})g_Y(z). \end{aligned}$$

Today's Topics

1.The Autocovariance-Generating Function

2. Sums of *ARMA* processes

Sum of an $MA(1)$ Process Plus White Noise

Adding Two MA Processes

Adding Two Autoregressive Processes

3. Wold's Decomposition and the Box-Jenkins Modeling Philosophy

4. Maximum Likelihood Estimation

2. Sums of *ARMA* processes

Example 1: Sum of an $MA(1)$ Process Plus White Noise

$MA(1)$ Process:

$$X_t = u_t + \delta u_{t-1},$$

where u_t is white noise. The autocovariances of X_t are

$$E(X_t X_{t-j}) = \begin{cases} (1+\delta^2)\sigma_u^2, & \textit{for } j = 0; \\ \delta\sigma_u^2, & \textit{for } j = \pm 1; \\ 0, & \textit{otherwise}; \end{cases}$$

Let v_t indicate a separate white noise series. Suppose that v and u are uncorrelated at all leads and lags:

$$E(u_t v_{t-j}) = 0, \quad \textit{for all } j,$$

implying

$$E(X_t v_{t-j}) = 0, \quad \textit{for all } j.$$

Define

$$
\begin{aligned}
Y_t &= X_t + v_t \\
&= u_t + \delta u_{t-1} + v_t.
\end{aligned}
$$

What are the time series properties of Y ?

Y_t has mean zero, and its autocovariances is

$$
\begin{aligned}
E(Y_t Y_{t-j}) &= E(X_t + v_t)(X_{t-j} + v_{t-j}) \\
&= E(X_t X_{t-j}) + E(v_t v_{t-j}) \\
&= \begin{cases} (1+\delta^2)\sigma_u^2 + \sigma_v^2, & \textit{for } j = 0; \\ \delta\sigma_u^2, & \textit{for } j = \pm 1; \\ 0, & \textit{otherwise}. \end{cases}
\end{aligned}
$$

Conclusion: Y has AGF similar to that of an MA(1) process!

Question: Does there exist a zero-mean $MA(1)$ representation for Y? That is, we seek for a representation such that

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

with

$$E(\varepsilon_t\varepsilon_{t-j}) = \begin{cases} \sigma^2, & \textit{for } j = 0; \\ 0, & \textit{otherwise}. \end{cases}$$

The autocovariance is

$$E(Y_t Y_{t-j}) = \begin{cases} (1+\theta^2)\sigma^2, & \textit{for } j = 0; \\ \theta\sigma^2, & \textit{for } j = \pm 1; \\ 0, & \textit{otherwise}; \end{cases}$$

In order to be consistent,

$$\begin{aligned}(1+\theta^2)\sigma^2 &= (1+\delta^2)\sigma_u^2+\sigma_v^2,\\ \theta\sigma^2 &= \delta\sigma_u^2.\end{aligned}$$

These lead to

$$\sigma^2 = \delta\sigma_u^2/\theta,$$

$$\delta\theta^2 - [(1+\delta^2)+(\sigma_v^2/\sigma_u^2)]\theta + \delta = 0. \tag{4.7.11}$$

For given values of δ, σ_u^2 and σ_v^2, two values of θ can be found from the quadratic formula:

$$\theta = \frac{[(1+\delta^2)+(\sigma_v^2/\sigma_u^2)] \pm \sqrt{[(1+\delta^2)+(\sigma_v^2/\sigma_u^2)]^2 - 4\delta^2}}{2\delta}.$$

If σ_v^2 were equal to zero, the quadratic equation is

$$\delta\theta^2 - (1+\delta^2)\theta + \delta = \delta(\theta - \delta)(\theta - \delta^{-1}) = 0. \tag{4.7.13}$$

whose solutions are $\theta = \delta$ and $\bar{\theta} = \delta^{-1}$.

For $\theta > 0$ and $\sigma_v^2 > 0$, (4.7.11) is everywhere lower than (4.7.13), implying that (4.7.11) has two real solution for θ: an invertible solution: $0 < |\theta^*| < |\delta|$ and a noninvertible solution: $0 < |\delta^{-1}| < |\tilde{\theta}^*|$.

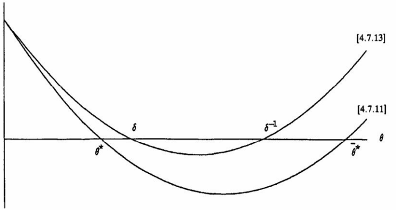

Figure 1 :

Taking the values associated with the invertible representation (θ^*, σ^{*2}),

$$(1 + \theta^* L)\varepsilon_t = (1 + \delta L)u_t + v_t,$$

or

$$\begin{aligned}
\varepsilon_t &= (1 + \theta^* L)^{-1}[(1 + \delta L)u_t + v_t] \\
&= (u_t - \theta^* u_{t-1} + \theta^{*2} u_{t-2} - \theta^{*3} u_{t-3} + \cdots) \\
&\quad +\delta(u_{t-1} - \theta^* u_{t-2} + \theta^{*2} u_{t-3} - \theta^{*3} u_{t-4} + \cdots) \\
&\quad +(v_t - \theta^* v_{t-1} + \theta^{*2} v_{t-2} - \theta^{*3} v_{t-3} + \cdots).
\end{aligned}$$

The series ε_t seems to posses a rich autocorrelation structure, however, it turns out to be white noise!!!

ε_t is White Noise !

$$g_Y(z) = (1 + \delta z)\sigma_u^2(1 + \delta z^{-1}) + \sigma_v^2,$$

so the *AGF* of $\varepsilon_t = (1 + \theta^* L)^{-1} Y_t$ is

$$g_\varepsilon(z) = \frac{(1 + \delta z)\sigma_u^2(1 + \delta z^{-1}) + \sigma_v^2}{(1 + \theta^* z)(1 + \theta^* z^{-1})},$$

But θ^* and σ^{*2} were chosen such that

$$(1 + \theta^* z)\sigma^{*2}(1 + \theta^* z^{-1}) = g_Y(z).$$

Thus,

$$g_\varepsilon(z) = \sigma^{*2}.$$

That is, ε_t is a white noise series !!!

Conclusion: $MA(1) + WN = MA(1)$!

Implications for Prediction

$$
\begin{aligned}
Y_t &= X_t + v_t = u_t + \delta u_{t-1} + v_t . (MA(1) + WN) \\
Y_t &= \varepsilon_t + \theta \varepsilon_{t-1} . (MA(1))
\end{aligned}
$$

The MA(1) suggests the optimal forecast of Y_{t+1} would be

$$\hat{E}(Y_{t+1}|Y_t, Y_{t-1}, \ldots) = \theta^* \varepsilon_t$$

with associated MSE σ^{*2}. While the MA(1)+WN suggests

$$\hat{E}(Y_{t+1}|X_t, X_{t-1}, \ldots, v_t, v_{t-1}, \ldots) = \delta u_t$$

with associated MSE $\sigma_u^2 + \sigma_v^2$. It can be shown that

$$\sigma^{*2} > \sigma_u^2 + \sigma_v^2 .$$

See Ferreira and Santa-Clara (2011). "Forecasting stock market returns:The sum of the parts is more than the whole." JFE, 100, 514-537.

Today's Topics

Example 2: Adding Two *MA* Processes

X_t is a zero-mean $MA(q_1)$ process:

$$X_t = (1 + \delta_1 L + \delta_2 L^2 + \cdots + \delta_{q_1} L^{q_1}) u_t \equiv \delta(L) u_t,$$

with

$$E(u_t u_{t-j}) = \begin{cases} \sigma_u^2, & \text{for } j = 0; \\ 0, & \text{otherwise}. \end{cases}$$

Let W_t be a zero-mean $MA(q_2)$ process:

$$W_t = (1 + \kappa_1 L + \kappa_2 L^2 + \cdots + \kappa_{q_2} L^{q_2}) v_t \equiv \kappa(L) v_t,$$

with

$$E(v_t v_{t-j}) = \begin{cases} \sigma_v^2, & \text{for } j = 0; \\ 0, & \text{otherwise}. \end{cases}$$

X and W are uncorrelated with each other at all leads and lags:

$$E(X_t W_{t-j}) = 0, \quad \textit{for all } j;$$

$$Y_t = X_t + W_t.$$

$$q \equiv \max\{q_1, q_2\}.$$

The jth autocovariance of Y is given by

$$\begin{aligned} E(Y_t Y_{t-j}) &= E(X_t + W_t)(X_{t-j} + W_{t-j}) \\ &= E(X_t X_{t-j}) + E(W_t W_{t-j}) \\ &= \begin{cases} \gamma_j^X + \gamma_j^W, & \textit{for } j = 0, \pm 1, \pm 2, \cdots, \pm q; \\ 0, & \textit{otherwise}. \end{cases} \end{aligned}$$

The autocovariances are zero beyond q lags, suggesting that Y_t might be represented as an $MA(q)$ process.

since

$$\gamma_j^Y = \gamma_j^X + \gamma_j^W,$$

it follows that

$$\begin{aligned} \sum_{j=-\infty}^{\infty} \gamma_j^Y z^j &= \sum_{j=-\infty}^{\infty} \gamma_j^X z^j + \sum_{j=-\infty}^{\infty} \gamma_j^W z^j, \\ g_Y(z) &= g_X(z) + g_W(z). \end{aligned}$$

If Y_t is to be expressed as an $MA(q)$ process,

$$Y_t = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) \equiv \theta(L)\varepsilon_t,$$

with

$$E(\varepsilon_t \varepsilon_{t-j}) = \begin{cases} \sigma^2, & \text{for } j = 0; \\ 0, & \text{otherwise}. \end{cases}$$

then its AGF would be

$$g_Y(z) = \theta(z)\theta(z^{-1})\sigma^2.$$

Question: whether there always exist values of $(\theta_1, \theta_2, \cdots, \theta_q, \sigma^2)$ such that

$$\theta(z)\theta(z^{-1})\sigma^2 = \delta(z)\delta(z^{-1})\sigma_u^2 + \kappa(z)\kappa(z^{-1})\sigma_v^2.$$

It turns out that there do !!!

Conclusion:

$$MA(q_1) + MA(q_2) = MA(max\{q_1, q_2\}).$$

Today's Topics

Example 3: Adding Two *AR* Processes

X_t and W_t are two uncorrelated $AR(1)$ processes:

$$
\begin{aligned}
(1 - \pi L)X_t &= u_t, \\
(1 - \rho L)W_t &= v_t,
\end{aligned}
$$

where u_t and v_t are each white noise with u_t uncorrelated with v_t for all t and τ.

$$Y_t = X_t + W_t,$$

Condition 1: $\pi = \rho$.

$$\begin{aligned}(1-\pi L)X_t + (1-\rho L)W_t &= u_t + v_t, \\ (1-\pi L)(X_t + W_t) &= u_t + v_t.\end{aligned}$$

$u_t + v_t$ is white noise, meaning that Y_t has an $AR(1)$ representation

$$(1-\pi L)Y_t = \varepsilon_t.$$

Condition 2: $\pi \neq \rho$.

$$\begin{aligned}(1-\rho L)(1-\pi L)X_t &= (1-\rho L)u_t, \\ (1-\pi L)(1-\rho L)W_t &= (1-\pi L)v_t. \\ (1-\rho L)(1-\pi L)(X_t+W_t) &= (1-\rho L)u_t+(1-\pi L)v_t.\end{aligned}$$

the right side has an $MA(1)$ representation, i.e,

$$(1-\phi_1 L-\phi_2 L^2)Y_t=(1+\theta L)\varepsilon_t,$$

where

$$(1-\phi_1 L-\phi_2 L^2)=(1-\rho L)(1-\pi L),$$

$$(1+\theta L)\varepsilon_t=(1-\rho L)u_t+(1-\pi L)v_t.$$

In other words,

$$AR(1)+AR(1)=ARMA(2,1).$$

In general, adding an $AR(p_1)$ process to an $AR(p_2)$ process with which it is uncorrelated at all leads and lags,

$$\pi(L)X_t = u_t, \quad \rho(L)W_t = v_t,$$

produces an $ARMA(p_1 + p_2, max\{p_1, p_2\})$ process,

$$\phi(L)Y_t = \theta(L)\varepsilon_t,$$

where

$$\begin{aligned} \phi(L) &= \pi(L)\rho(L), \\ \theta(L)\varepsilon_t &= \rho(L)u_t + \pi(L)v_t. \end{aligned}$$

For more discussion on related topics, see Granger and Morris (1976), "Time Series Modelling and Interpretation," JRSS Series A, 246-257

Today's Topics

3. Wold's (1938) decomposition

Any zero-mean covariance-stationary process Y_t can be represented in the form

$$Y_t = \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} + \kappa_t,$$

where $\psi_0 = 1$ and $\sum_{j=1}^{\infty} \psi_j^2 < \infty$. The term ε_t is white noise and represents the error made in forecasting Y_t on the basis of a linear function of lagged Y:

$$\varepsilon_t \equiv Y_t - \hat{E}(Y_t | Y_{t-1}, Y_{t-2}, \cdots). \qquad (\textit{indeterministic})$$

The value of κ_t is uncorrelated with ε_{t-j} for any j, though κ_t can be predicted arbitrary well from a linear function of the past values of Y:

$$\hat{\kappa}_t = \hat{E}(\kappa_t | Y_{t-1}, Y_{t-2}, \cdots). \qquad (\textit{deterministic})$$

Practical implications

The Wold representation requires fitting an infinite number of parameters $(\psi_1, \psi_2, \cdots)$ to the data. As a practical matter, we need to make some additional assumptions about the nature of $(\psi_1, \psi_2, \cdots)$. A typical assumption is $\psi(L)$ can be expressed as the ratio of two finite-order polynomials:

$$\sum_{j=0}^{\infty} \psi_j L^j \overset{Ass.}{=} \frac{\theta(L)}{\phi(L)} \equiv \frac{1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q}{1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p}.$$

Today's Topics

The Box-Jenkins Modeling Philosophy

The approach to forecasting advocated by Box and Jenkins (1976) can be broken into four steps:

(1) Transform the data, if necessary, so that the assumption of covariance-stationarity is a reasonable one.

(2) Make an initial guess of small values for p and q for an $ARMA(p,q)$ model that might describe the transformed series.

(3) Estimate the parameters in $\phi(L)$ and $\theta(L)$.

(4) Perform diagnostic analysis to confirm that the model is indeed consistent with the observed features of the data.

Today's Topics

4. Maximum Likelihood Estimation

Consider an *ARMA* model

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

with ε_t white noise:

$$\begin{aligned} E(\varepsilon_t) &= 0, \\ E(\varepsilon_t \varepsilon_\tau) &= \begin{cases} \sigma^2, & \text{for } t = \tau; \\ 0, & \text{otherwise}. \end{cases} \end{aligned}$$

Question: How to estimate the unknown values of

$$(c, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma^2)$$

on the basis of observations on $\mathbf{Y}$?

Maximum Likelihood Principle

- Population parameters: $\boldsymbol{\theta} \equiv (c, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma^2)'$
- Observed sample: $\{y_1, \cdots, y_T\}$
- Calculate the joint probability density

$$f_{Y_T, Y_{T-1}, \cdots, Y_1}(y_T, y_{T-1}, \cdots, y_1; \boldsymbol{\theta}).$$

- The MLE of $\boldsymbol{\theta}$ is the value that maximizes the above joint probability density (viewed as the likelihood function of the parameters), i.e.,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{argmax} L(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta} \in \Theta}{argmax} f_{Y_T, Y_{T-1}, \cdots, Y_1}(y_T, y_{T-1}, \cdots, y_1; \boldsymbol{\theta})$$

Distributional Assumption

- Typically, we assume that ε_t is Gaussian white noise:

$$\varepsilon_t \sim \ i.i.d.\ N(0, \sigma^2).$$

  Although this assumption is strong, the estimate of $\boldsymbol{\theta}$ that result from it will often turn out to be sensible for non-Gaussian processes as well.

- The Box-Cox transformation (Box and Cox, 1964)

$$Y_t^{(\lambda)} = \begin{cases} \frac{Y_t^{\lambda}-1}{\lambda} & \text{for } \lambda \neq 0; \\ \log Y_t, & \text{for } \lambda = 0. \end{cases}$$

- Other distributions, such as the student t, the extreme value distributions, are also popular for financial applications.

How to find MLE ?

- Two key steps:
  1. Calculate the likelihood function.
  2. Find $\boldsymbol{\theta}$ which maximizes the likelihood.
- We shall discuss
  1. how to obtain the likelihood function for Gaussian $AR(1), MA(1)$ and the general $ARMA(p, q)$ processes.
  2. how to maximize the likelihood function.

Today's Topics

1.The Autocovariance-Generating Function

2. Sums of *ARMA* processes

3. Wold's Decomposition and the Box-Jenkins Modeling Philosophy

4. Maximum Likelihood Estimation

Introdution

Example 1: The Likelihood Function for an AR(1)

Example 2: The Likelihood Function for an MA(1)

Example 3: The Likelihood Function for an ARMA(p,q)

Numerical Optimization

Example 1: The Likelihood Function for an AR(1)

Gaussian $AR(1)$ process:

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t, \text{ with } \varepsilon_t \sim i.i.d.\, N(0, \sigma^2).$$

First consider the probability distribution of Y_1 (under stationarity, i.e., $|\phi| < 1$,),

- Mean: $E(Y_1) = \mu = c/(1-\phi)$,
- Variance: $E(Y_1 - \mu)^2 = \sigma^2/(1-\phi^2)$.

Since ε_t is Gaussian, Y_1 is also Gaussian. The density of Y_1 is

$$\begin{aligned} f_{Y_1}(y_1; \boldsymbol{\theta}) &= f_{Y1}(y_1; c, \phi, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2/(1-\phi^2)}} exp\left[\frac{-\{y_1 - [c/(1-\phi)]\}^2}{2\sigma^2/(1-\phi^2)}\right]. \end{aligned}$$

Next, consider the distribution of Y_2 conditional on observing $Y_1 = y_1$,

$$Y_2 = c + \phi Y_1 + \varepsilon_2,$$

$$(Y_2 | Y_1 = y_1) \sim N((c + \phi y_1), \sigma^2),$$

meaning

$$f_{Y_2|Y_1}(y_2|y_1; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-(y_2 - c - \phi y_1)^2}{2\sigma^2}\right].$$

The joint distribution of Y_1 and Y_2 is

$$f_{Y_2,Y_1}(y_2, y_1; \boldsymbol{\theta}) = f_{Y_2|Y1}(y_2|y_1; \boldsymbol{\theta}) f_{Y1}(y_1; \boldsymbol{\theta}).$$

Similarly, the distribution of the third observation conditional on the first two is

$$\begin{aligned} f_{Y_3|Y_2,Y_1}(y_3|y_2,y_1;\boldsymbol{\theta}) &= f_{Y_3|Y_2}(y_3|y_2;\boldsymbol{\theta}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-(y_3-c-\phi y_2)^2}{2\sigma^2}\right], \end{aligned}$$

from which

$$f_{Y_3,Y_2,Y_1}(y_3,y_2,y_1;\boldsymbol{\theta}) = f_{Y_3|Y_2,Y_1}(y_3|y_2,y_1;\boldsymbol{\theta}) f_{Y_2,Y1}(y_2,y_1\boldsymbol{\theta}).$$

In general,

$$
\begin{aligned}
& f_{Y_t|Y_{t-1},Y_{t-2},\cdots,Y_1}(y_t|y_{t-1},\cdots,y_1;\boldsymbol{\theta}) \\
&= f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\boldsymbol{\theta}) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} exp[\frac{-(y_t - c - \phi y_{t-1})^2}{2\sigma^2}].
\end{aligned}
$$

The joint density of the whole sample is then

$$
\begin{aligned}
& f_{Y_T,Y_{T-1},\cdots,Y_1}(y_T,y_{T-1},\cdots,y_1;\boldsymbol{\theta}) \\
&= f_{Y_T|Y_{T-1}}(y_t|y_{T-1};\boldsymbol{\theta}) f_{Y_{T-1},\cdots,Y1}(y_{T-1},\cdots,y_1;\boldsymbol{\theta}) \\
&= f_{Y_1}(y_1;\boldsymbol{\theta}) \prod_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\boldsymbol{\theta}).
\end{aligned}
$$

The log likelihood function is

$$
\mathcal{L}(\boldsymbol{\theta}) = log\ f_{Y_1}(y_1;\boldsymbol{\theta}) + \sum_{t=2}^{T} log\ f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\boldsymbol{\theta}).
$$

Recall the log likelihood function is

$$\mathcal{L}(\boldsymbol{\theta}) = log\ f_{Y_1}(y_1; \boldsymbol{\theta}) + \sum_{t=2}^{T} log\ f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \boldsymbol{\theta}).$$

The log likelihood function for a sample of size T from a Gaussian $AR(1)$ process is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= -\frac{1}{2}log(2\pi) - \frac{1}{2}log[\sigma^2/(1-\phi^2)] - \frac{\{y_1 - [c/(1-\phi)]\}^2}{2\sigma^2/(1-\phi^2)} \\
&\quad -[(T-1)/2]log(2\pi) - [(T-1)/2]log(\sigma^2) \\
&\quad -\sum_{t=2}^{T}\left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}\right].
\end{aligned}$$

The MLE for a Gaussian $AR(1)$ process does not have an explicit solution. Numerical methods are required to find the MLE.

Conditional Maximum Likelihood

Regard the value of y_1 as deterministic and maximize the likelihood conditioned on the first observation,

$$f_{Y_t,Y_{t-1},\cdots,Y_2|Y_1}(y_t,y_{t-1},\cdots,y_2|y_1;\boldsymbol{\theta}) = \prod_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\boldsymbol{\theta})$$

The objective of conditional maximum likelihood is to maximize

$$\begin{aligned} & log\, f_{Y_t,Y_{t-1},\cdots,Y_2|Y_1}(y_t,y_{t-1},\cdots,y_2|y_1;\boldsymbol{\theta}) \\ = \; & -[(T-1)/2]log(2\pi) - [(T-1)/2]log(\sigma^2) \\ & -\sum_{t=2}^{T}\left[\frac{(y_t-c-\phi y_{t-1})^2}{2\sigma^2}\right], \end{aligned}$$

which is equivalent to the familiar least square estimation!!!

This leads to

$$\begin{bmatrix} \hat{c} \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} T-1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y_{t-1} \\ \Sigma y_{t-1} y_t \end{bmatrix},$$

$$\hat{\sigma}^2 = \sum \left[\frac{(y_t - \hat{c} - \hat{\phi} y_{t-1})^2}{T-1} \right].$$

where Σ denotes summation over $t = 2, 3, \cdots, T$.

The exact *MLE* and conditional *MLE* turn out to have the same large sample distribution, provided that $|\phi| < 1$. And when $|\phi| > 1$, the conditional *MLE* continues to provide consistent estimates, whereas exact *MLE* does not.

Today's Topics

1.The Autocovariance-Generating Function

2. Sums of *ARMA* processes

3. Wold's Decomposition and the Box-Jenkins Modeling Philosophy

4. Maximum Likelihood Estimation

- Introduction
- Example 1: The Likelihood Function for an AR(1)
- Example 2: The Likelihood Function for an MA(1)
- Example 3: The Likelihood Function for an ARMA(p,q)
- Numerical Optimization

Example 2: The Likelihood Function for an MA(1)

Gaussian $MA(1)$ process:

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

with $\varepsilon_t \sim \ i.i.d.\, N(0, \sigma^2)$.

Let $\boldsymbol{\theta} = (\mu, \theta, \sigma^2)'$,

$$Y_t|\varepsilon_{t-1} \sim N((\mu + \theta\varepsilon_{t-1}), \sigma^2),$$

$$f_{Y_t|\varepsilon_{t-1}}(y_t|\varepsilon_{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[\frac{-(y_t - \mu - \theta\varepsilon_{t-1})^2}{2\sigma^2}].$$

Suppose $\varepsilon_0 = 0$, then

$$Y_1|\varepsilon_0 \sim N(\mu, \sigma^2).$$

Given observation of y_1, the value of ε_1 is known as

$$\varepsilon_1 = y_1 - \mu.$$

$$f_{Y_2|Y_1,\varepsilon_0=0}(y_2|y_1, \varepsilon_0 = 0; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[\frac{-(y_2 - \mu - \theta\varepsilon_1)^2}{2\sigma^2}].$$

Since ε_1 is known, ε_2 can be calculated from

$$\varepsilon_2 = y_2 - \mu - \theta\varepsilon_1.$$

As a result, given $\varepsilon_0 = 0$, the full sequence $\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T\}$ can be calculated from

$$\varepsilon_t = y_t - \mu - \theta\varepsilon_{t-1},$$

for $t = 1, 2, \cdots, T$. The conditional density of the tth observation can be calculated as

$$\begin{aligned}
& f_{Y_t|Y_{t-1},Y_{t-2},\cdots,Y_1,\varepsilon_0=0}(y_t|y_{t-1}, y_{t-2}, \cdots, y_1, \varepsilon_0 = 0; \boldsymbol{\theta}) \\
&= f_{Y_t|\varepsilon_{t-1}}(y_t|\varepsilon_{t-1}; \boldsymbol{\theta}) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} exp[\frac{-\varepsilon_t^2}{2\sigma^2}].
\end{aligned}$$

The sample likelihood is

$$f_{Y_T,Y_{T-1},\cdots,Y_1|\varepsilon_0=0}(y_T,y_{T-1},\cdots,y_1|\varepsilon_0=0;\boldsymbol{\theta})$$
$$=f_{Y_1|\varepsilon_0=0}(y_1|\varepsilon_0=0;\boldsymbol{\theta})\prod_{t=2}^{T}f_{Y_{t-1},Y_{t-2},\cdots,Y_1,\varepsilon_0=0}(y_{t-1},y_{t-2},\cdots,y_1,\varepsilon_0=0;\boldsymbol{\theta})$$

The conditional log likelihood is

$$\begin{aligned}\mathcal{L}(\theta) &= log\, f_{Y_T,Y_{T-1},\cdots,Y_1|\varepsilon_0=0}(y_T,y_{T-1},\cdots,y_1|\varepsilon_0=0;\boldsymbol{\theta})\\ &= -\frac{T}{2}log(2\pi)-\frac{T}{2}log(\sigma^2)-\sum_{t=1}^{T}\frac{\varepsilon_t^2}{2\sigma^2}.\end{aligned}$$

Remarks

- The log likelihood is a complicated function of μ and θ, so even the conditional maximum likelihood estimates for an $MA(1)$ process must be found by numerical optimization.
- The effect of initial value ε_0 on iteration:

$$
\begin{aligned}
\varepsilon_t &= (y_t - \mu) - \theta(y_{t-1} - \mu) + \theta^2(y_{t-2} - \mu) - \cdots \\
&\quad + (-1)^{t-1}\theta^{t-1}(y_1 - \mu) + (-1)^t \theta^t \varepsilon_0.
\end{aligned}
$$

  If $|\theta| < 1$, the effect of ε_0 will quickly die out. If $|\theta| > 1$, the conditional approach is not reasonable.

Exact Likelihood Function (Self-reading)

$$\mathbf{\Omega} = E[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'].$$

The variance-covariance matrix for an $MA(1)$ process is

$$\mathbf{\Omega} = \sigma^2 \begin{bmatrix} (1+\theta^2) & \theta & 0 & \cdots & 0 \\ \theta & (1+\theta^2) & \theta & \cdots & 0 \\ 0 & \theta & (1+\theta^2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \cdots & (1+\theta^2) \end{bmatrix}$$

The likelihood function is then

$$f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta}) = (2\pi)^{-T/2} |\mathbf{\Omega}|^{-1/2} exp[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})' \mathbf{\Omega}^{-1} (\mathbf{y} - \boldsymbol{\mu})].$$

The triangular factorization of $\boldsymbol{\Omega}$: $\boldsymbol{\Omega} = \mathbf{ADA}'$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \cdots & 0 & 0 \\ \frac{\theta}{1+\theta^2} & 1 & 0 \cdots & 0 & 0 \\ 0 & \frac{\theta(1+\theta^2)}{1+\theta^2+\theta^4} & 1 \cdots & 0 & 0 \\ \vdots & \vdots & \vdots \ldots & \vdots & \vdots \\ 0 & 0 & 0 \cdots & \frac{\theta[1+\theta^2+\theta^4+\cdots+\theta^{2(n-2)}]}{1+\theta^2+\theta^4+\cdots+\theta^{2(n-2)}} & 1 \end{bmatrix}$$

$$\mathbf{D} = \sigma^2 \begin{bmatrix} 1+\theta^2 & 0 & 0 & \cdots & 0 \\ 0 & \frac{1+\theta^2+\theta^4}{1+\theta^2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1+\theta^2+\theta^4+\theta^6}{1+\theta^2+\theta^4} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1+\theta^2+\theta^4+\cdots+\theta^{2n}}{1+\theta^2+\theta^4+\cdots+\theta^{2(n-1)}} \end{bmatrix}$$

Then the likelihood function is

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) &= (2\pi)^{-T/2}|\mathbf{ADA'}|^{-1/2} \\ &\times exp[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'[\mathbf{A}']^{-1}\mathbf{D}^{-1}\mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})]. \end{aligned}$$

$$|\mathbf{A}| = 1 \Longrightarrow |\mathbf{ADA}'| = |\mathbf{A}||\mathbf{D}||\mathbf{A}'| = |\mathbf{D}|.$$

Defining

$$\bar{\mathbf{y}} \equiv \mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu}),$$

the likelihood can be written

$$f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) = (2\pi)^{-T/2}|\mathbf{D}|^{-1/2} exp[-\frac{1}{2}\bar{\mathbf{y}}'\mathbf{D}^{-1}\bar{\mathbf{y}}].$$

$$\mathbf{A}\bar{\mathbf{y}} = \mathbf{y} - \boldsymbol{\mu}.$$

The first row states that $\bar{y_1} = y_1 - \mu$, while the tth row implies that

$$\bar{y}_t = y_t - \mu - \frac{\theta[1 + \theta^2 + \theta^4 + \cdots + \theta^{2(t-2)}]}{1 + \theta^2 + \theta^4 + \cdots + \theta^{2(t-1)}} \bar{y}_{t-1}.$$

$$d_{tt} = \sigma^2 \frac{1 + \theta^2 + \theta^4 + \cdots + \theta^{2t}}{1 + \theta^2 + \theta^4 + \cdots + \theta^{2(t-1)}}.$$

Since $\mathbf{D}$ is diagonal, its determinant is

$$\begin{aligned} |\mathbf{D}| &= \prod_{t=1}^{T} d_{tt}. \\ \bar{\mathbf{y}}'\mathbf{D}^{-1}\bar{\mathbf{y}} &= \sum_{t=1}^{T} \frac{\bar{y}_t^2}{d_{tt}}. \end{aligned}$$

As a result,

$$f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) = (2\pi)^{-T/2}\left[\prod_{t=1}^{T} d_{tt}\right]^{-1/2} exp\left[-\frac{1}{2}\sum_{t=1}^{T}\frac{\bar{y}_t^2}{d_{tt}}\right].$$

The exact log likelihood for a Gaussian $MA(1)$ process is

$$\mathcal{L}(\theta) = log\ f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) = -\frac{T}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} log(d_{tt}) - \frac{1}{2}\sum_{t=1}^{T}\frac{\bar{y}_t^2}{d_{tt}}.$$

Today's Topics

Example 3: The Likelihood Function for an ARMA(p,q)

Gaussian ARMA(p,q) process:

$$\begin{aligned} Y_t &= c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \\ &\quad + \theta_1 \varepsilon_{t-1} \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \end{aligned}$$

where $\varepsilon_t \sim i.i.d.\, N(0, \sigma^2)$.

Question: how to estimate $\boldsymbol{\theta} = (c, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma^2)'$.

Taking initial values for $\mathbf{y_0} \equiv (y_0, y_{-1}, \cdots, y_{-p+1})'$ and $\boldsymbol{\varepsilon_0} \equiv (\varepsilon_0, \varepsilon_{-1}, \cdots, \varepsilon_{-q+1})'$. Then the sequence $\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T\}$ can be calculated from $\{y_1, y_2, \cdots, y_T\}$ by iterating on

$$\begin{aligned} \varepsilon_t &= y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - -\phi_p y_{t-p} \\ &\quad -\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - -\theta_q \varepsilon_{t-q}, \end{aligned}$$

for $t = 1, 2, \cdots, T$. The conditional log likelihood is then

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= log\, f_{Y_T, Y_{T-1}, \cdots, Y_1 | \mathbf{Y_0}, \boldsymbol{\varepsilon_0}}(y_T, y_{T-1}, \cdots, y_1 | \mathbf{y_0}, \boldsymbol{\varepsilon_0}; \boldsymbol{\theta}) \\ &= -\frac{T}{2} log(2\pi) - \frac{T}{2} log(\sigma^2) - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2}. \end{aligned}$$

How to set initial y's and ε's ?

(1) equal to their expected value. That is, set

$y_s = c/(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$ for $s = 0, -1, \cdots, -p+1$

and set $\varepsilon_s = 0$ for $s = 0, -1, \cdots, -p+1$.

(2) set ε's to zero but y's equal to their actual values.

Then the conditional likelihood is

$$log\, f(y_T, y_{T-1}, \cdots, y_{p+1} | y_p, \cdots, y_1, \varepsilon_p = 0, \cdots, \varepsilon_{p-q+1} = 0)$$
$$= -\frac{T-p}{2} log(2\pi) - \frac{T-p}{2} log(\sigma^2) - \sum_{t=p+1}^{T} \frac{\varepsilon_t^2}{2\sigma^2}.$$

Remarks

As in the case for the MA processes, these approximation should be used only if all values of z satisfying

$$1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q = 0$$

lie outside the unit circle.

Today's Topics

Numerical Optimization

Newton-Raphson Method

Suppose that $\boldsymbol{\theta}$ is an $(a \times 1)$ vector of parameters to be estimated.

Let

$$\underset{(a\times 1)}{\mathbf{g}(\boldsymbol{\theta}^{(0)})} = \left.\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(0)}},$$

$$\underset{(a\times a)}{\mathbf{H}(\boldsymbol{\theta}^{(0)})} = -\left.\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(0)}}.$$

Consider approximating $\mathcal{L}(\boldsymbol{\theta})$ with a second-order Taylor series around $\boldsymbol{\theta}^{(0)}$:

$$\mathcal{L}(\boldsymbol{\theta}) \cong \mathcal{L}(\boldsymbol{\theta}_0) + \mathbf{g}[\boldsymbol{\theta}^{(0)}]'[\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}] - \frac{1}{2}[\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}]'\mathbf{H}(\boldsymbol{\theta}^{(0)})[\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}].$$

Setting the derivative of the above equation with respect to $\boldsymbol{\theta}$ equal to zero results in

$$\mathbf{g}(\boldsymbol{\theta}^{(0)}) - \mathbf{H}(\boldsymbol{\theta}^{(0)})[\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}] = 0.$$

The above equation suggests that an improved estimate of $\boldsymbol{\theta}$ (denoted $\boldsymbol{\theta}^{(1)}$) would satisfy

$$\mathbf{g}(\boldsymbol{\theta}^{(0)}) - \mathbf{H}(\boldsymbol{\theta}^{(0)})[\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(0)}] = 0.$$

or

$$\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(0)} = [\mathbf{H}(\boldsymbol{\theta}^{(0)})]^{-1}\mathbf{g}(\boldsymbol{\theta}^{(0)}).$$

Similarly, the *m*th step in the iteration updates the estimate of $\boldsymbol{\theta}$ by using the formula

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + [\mathbf{H}(\boldsymbol{\theta}^{(m)})]^{-1}\mathbf{g}(\boldsymbol{\theta}^{(m)}).$$

(1) If the log likelihood function happens to be a perfect quadratic function, we will get the exact *MLE* in a single step:

$$\boldsymbol{\theta}^{(1)} = \hat{\boldsymbol{\theta}}_{MLE}.$$

(2) If the quadratic approximation is reasonable good, *NR* should converge to the local maximum quickly.

(3) If the likelihood function is not concave, *NR* behaves quite poorly.

The iteration is often modified as follows:

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + s[\mathbf{H}(\boldsymbol{\theta}^{(m)})]^{-1}\mathbf{g}(\boldsymbol{\theta}^{(m)}).$$

where s is a scalar controlling the step length.

One calculates $\boldsymbol{\theta}^{(m+1)}$ and the associated value for the log likelihood $\mathcal{L}(\boldsymbol{\theta}^{(m+1)})$ for various values of s and chooses as the estimate $\boldsymbol{\theta}^{(m+1)}$ the value that produces the biggest value for the log likelihood.