Maximum Likelihood Estimation of Time Series Models

(Hamilton: Chapters 5)

Isabel Casas

Syddansk Universitet

Lecture Outline

- Introduction
- MLE of a Gaussian AR(p)
- MLE of a Gaussian MA(q)
- MLE of a Gaussian ARMA(p,q)

Introduction

In order to make use of the maximum likelihood method of parameter estimation, one has to make distributional assumptions about $\{Y_t\}$ or the errors $\{\epsilon_t\}$ in the model

$$Y_t = \mathsf{E}(Y_t|\mathcal{F}_{t-1}) + \epsilon_t$$

where $\mathcal{F}_{t-1}$ is the past information (the information one is conditioning on).

The idea behind the maximum likelihood estimation is to find the value of the parameter vector that is the most likely one, given the observations.

Example: Linear regression, x_t fixed

$$Y_i = \mathbf{x}_i \beta + \epsilon_t, \qquad \epsilon_t \sim \text{IID } (0, \sigma^2),\ i = 1, ..., N.$$

The likelihood function is the joint density of the observations Y_i, $i = 1, ..., N$:

$$f_Y(y_1, ..., y_N; \beta, \sigma^2) = \prod_{i=1}^{N} f_i(y_i; \beta, \sigma^2). \tag{1}$$

The most likely value of (β, σ^2), where $y_1, ..., y_N$ are fixed, is the one that maximizes (1).

This value also maximizes $\log f_Y(y_1, ..., y_N; \beta, \sigma^2)$ since the logarithmic transformation is a monotonic one.

Example: Linear regression, x_t fixed

Thus equation (1) is equivalent to (2):

$$\mathcal{L}(\beta, \sigma^2) = \log f_Y(y_1, ..., y_N; \beta, \sigma^2) = \sum_{i=1}^{N} \log f_i(y_i; \beta, \sigma^2) \qquad (2)$$

Popular assumption: $\{\epsilon_t\} \sim IIDN(0, \sigma^2)$:

$$\mathcal{L}(\beta, \sigma^2) = -\frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \mathbf{x}_i \beta)^2.$$

The log-likelihood is maximized by first maximizing the function with respect to β and then, given the estimated value of β, with respect to σ^2.

Example: Linear regression, x_t fixed

Maximizing (1) or (2) with respect to β is equivalent to minimizing

$$\sum_{i=1}^{N}(y_i - \mathbf{x}_i\beta)^2$$

with respect to β.

Minimization yields

$$\widehat{\beta} = (\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i)^{-1}\sum_{i=1}^{N}\mathbf{x}_i y_i.$$

which equals the **least squares estimator** of β.

Example: Linear regression, x_t fixed

Then, the first-order condition

$$\frac{\partial}{\partial \sigma^2} \sum_{i=1}^{N} \log f_i(y_i; \beta, \sigma^2) = -\frac{N}{2\sigma^2} + \frac{\sum_{i=1}^{N} (y_i - \mathbf{x}_i \widehat{\beta})^2}{2\sigma^4} = 0.$$

Solving for σ^2 gives the maximum likelihood estimator

$$\widehat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \mathbf{x}_i \widehat{\beta})^2.$$

Consequently, can estimate β first, then σ^2.

Maximum likelihood of a dependent process

When we have a stochastic process such that Y_t and Y_{t-j} are dependent for at least one $j \neq 0$,

$$f_Y(y_1, ..., y_T; \boldsymbol{\theta}) \neq \prod_{t=1}^{T} f_j(y_j; \boldsymbol{\theta}) \tag{3}$$

where $f_Y(y_1, ..., y_T; \boldsymbol{\theta})$ is the joint density of $Y_1, ..., Y_T$, and $\boldsymbol{\theta}$ is the parameter vector.

Q: Can we get an expression of the joint density?

Maximum likelihood for any distribution

- If the shocks $\epsilon_t \sim N(0, \sigma^2)$ but they are dependent, then we can use the multivariate normal density function to calculate the log–likelihood function.
- If the shocks have any other type of density, the best way to find the log–likelihood function is to factorise the joint density function into the conditional density function multiply by the marginal density function of the first variable

$$f_{X,Y}(x, y) = f_{Y|X}(y|x)f_X(x).$$

SYDDANSK UNIVERSITET

Maximum likelihood for any distribution

Factorize $f_{\mathbf{Y}}(y_1, ..., y_T; \boldsymbol{\theta})$, assuming that $\{Y_t\}$ is strictly stationary and depends of the previous p lags:

- First step:

$$\begin{aligned} f_{\mathbf{Y}}(y_1, ..., y_T; \boldsymbol{\theta}) = & f_{Y_T | Y_{T-1} ..., Y_1}(y_{T-1}, ..., y_1; \boldsymbol{\theta}) \\ & \cdot f_{Y_1, ..., Y_{T-1}}(y_1, ..., y_{T-1}; \boldsymbol{\theta}) \end{aligned}$$

- Second step:

$$\begin{aligned} f_{\mathbf{Y}}(y_1, ..., y_T; \boldsymbol{\theta}) = & f_{Y_T | Y_{T-1} ..., Y_1}(y_{T-1}, ..., y_1; \theta) \\ & \cdot f_{Y_{T-1} | Y_{T-2} ..., Y_1}(y_{T-2}, ..., y_1; \theta) \\ & \cdot f_{Y_1, ..., Y_{T-2}}(y_1, ..., y_{T-2}; \boldsymbol{\theta}) \end{aligned}$$

Maximum likelihood for any distribution

Continue p times:

$$f_{\mathbf{Y}}(y_1, ..., y_T; \boldsymbol{\theta}) = \left\{ \prod_{t=p+1}^{T} f_{Y_t | Y_{t-1}..., Y_1}(y_t | y_{t-1}, ..., y_1; \boldsymbol{\theta}) \right\} \cdot f_{Y_1,...,Y_p}(y_1, \ldots, y_p; \boldsymbol{\theta})$$

Taking logs:

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

Q: Does it look similar to (2)?

Exact maximum likelihood for AR(1) with Gaussian errors

Consider the AR(1) process

$$Y_t = c + \phi Y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{IIDN}(0, \sigma^2),\ t = 1, ..., T.$$

Assuming $|\phi| < 1$, we have the unconditional moments

$$\mathsf{E}(Y_t) = \frac{c}{1-\phi}, \quad \mathsf{Var}(Y_t) = \frac{\sigma^2}{1-\phi^2}$$

and the conditional moments

$$\mathsf{E}(Y_t | Y_{t-1} = y_{t-1}) = c + \phi y_{t-1}$$

$$\mathsf{Var}(Y_t | Y_{t-1} = y_{t-1}) = \mathsf{E}(Y_t - c - \phi y_{t-1})^2 = \mathsf{E}(\epsilon_t^2) = \sigma^2$$

Exact maximum likelihood for AR(1) with Gaussian errors

Thus the **conditional** density

$$f_{Y_T|Y_{T-1},\dots,Y_1}(y_t|y_{t-1},\dots,y_1;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}\right\},$$

for $t = 2, \dots, T$.

The **unconditional** density

$$f_{Y_1}(y_1;\boldsymbol{\theta}) = \sqrt{\frac{1-\phi^2}{2\pi\sigma^2}}\exp\left\{-\frac{(y_1 - \frac{c}{1-\phi})^2(1-\phi^2)}{2\sigma^2}\right\}. \quad (4)$$

Exact maximum likelihood for AR(1) with Gaussian errors

Putting the pieces together, the log-likelihood of the stationary AR(1) model becomes

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(1-\phi^2) \tag{5}$$

$$-\frac{(y_1 - \frac{c}{1-\phi})^2(1-\phi^2)}{2\sigma^2} - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - c - \phi y_{t-1})^2 \tag{6}$$

Q: How would you find the estimators of $\boldsymbol{\theta} = (c, \phi, \sigma^2)$

Exact maximum likelihood for AR(1) with Gaussian errors

Putting the pieces together, the log-likelihood of the stationary AR(1) model becomes

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(1-\phi^2) \quad (5)$$

$$-\frac{(y_1 - \frac{c}{1-\phi})^2(1-\phi^2)}{2\sigma^2} - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - c - \phi y_{t-1})^2 \quad (6)$$

Q: How would you find the estimators of $\boldsymbol{\theta} = (c, \phi, \sigma^2)$

A: It requires numerical methods

Q: Can we avoid that?

Conditional likelihood for a Gaussian AR(1)

If the density (4) of Y_1 is excluded, one obtains the conditional log-likelihood

$$\log f_Y(y_1, ..., y_T; \boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - c - \phi y_{t-1})^2. \tag{7}$$

Estimating $\boldsymbol{\theta}$ from the equation (7) above is the same than estimating it using least squares, i.e. minimising

$$\sum_{t=2}^{T}(y_t - c - \phi y_{t-1})^2 \tag{8}$$

with respect to c and ϕ. and then after σ^2.

SYDDANSK UNIVERSITET

Conditional likelihood for a Gaussian AR(1)

Writting $X_t = [1, y_{t-1}]$, minimisation of (8) yields

$$
\begin{aligned}
\begin{bmatrix} \widehat{c} \\ \widehat{\phi} \end{bmatrix} &= (X_t' X_t)^{-1}(X_t' Y_t) \\
&= \left(\sum_{t=2}^{T} \begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix} \begin{bmatrix} 1 & y_{t-1} \end{bmatrix} \right)^{-1} \left(\sum_{t=2}^{T} \begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix} y_t \right) \\
&= \frac{1}{(T-1)\sum y_{t-1}^2 - (\sum y_{t-1})^2} \\
&\quad \times \begin{bmatrix} \sum y_{t-1}^2 & -\sum y_{t-1} \\ -\sum y_{t-1} & T-1 \end{bmatrix} \begin{bmatrix} \sum y_t \\ \sum y_t y_{t-1} \end{bmatrix}
\end{aligned}
$$

Conditional likelihood for a Gaussian AR(1)

$$
\begin{aligned}
\widehat{c} =& \frac{(\sum y_{t-1}^2)(\sum y_t) - (\sum y_t y_{t-1})(\sum y_{t-1})}{(T-1)\sum y_{t-1}^2 - (\sum y_{t-1})^2} \\
\approx& \overline{y} \frac{\sum y_{t-1}(y_{t-1} - y_t)}{\sum y_{t-1}^2 - \frac{1}{T-1} \sum y_{t-1})^2}
\end{aligned}
$$

and

$$
\widehat{\phi} = \frac{\sum y_t y_{t-1} - \frac{1}{T-1}(\sum y_{t-1})(\sum y_t)}{\sum y_{t-1}^2 - \frac{1}{T-1}(\sum y_{t-1})^2} \approx \frac{\sum (y_t - \overline{y})(y_{t-1} - \overline{y})}{\sum (y_{t-1} - \overline{y})^2}.
$$

where

$$
\overline{y} = \frac{1}{T-1} \sum y_{t-1}.
$$

Conditional likelihood for a Gaussian AR(1)

$$\hat{\sigma} = \frac{1}{T-1} \sum_{t=2}^{T} \{y_t - \hat{c} - \hat{\phi} y_{t-1}\}^2$$

These OLS estimators are consitent if the process is ergodic for the second moments.

Conditional likelihood for a Gaussian AR(1)

- Thus, exact maximum likelihood estimation of ϕ requires numerical optimisation
- Conditional maximum likelihood estimation yields an analytical solution.

Q: Would it be better just to use the conditional log-likelihood because it gives an analytical solution to the maximisation problem instead of the exact log-likelihood? Asymptotically they must give the same answer.

Conditional likelihood for a Gaussian AR(1)

- Thus, exact maximum likelihood estimation of ϕ requires numerical optimisation
- Conditional maximum likelihood estimation yields an analytical solution.

Q: Would it be better just to use the conditional log-likelihood because it gives an analytical solution to the maximisation problem instead of the exact log-likelihood? Asymptotically they must give the same answer.

A: In small samples (short time series), the exact log-likelihood gives more accurate answers and never allows $|\widehat{\phi}| \geq 1$, due to the terms $\frac{1}{2}\log(1-\phi^2)$ and $-(2\sigma^2)^{-1}(y_1 - \frac{c}{1-\phi})^2(1-\phi^2)$ in (6).

Maximum likelihood estimation of non-Gaussian time series

When $\{Y_1, ..., Y_T\}$ is not multivariate normal, the log-likelihood function (7) is not the correct one.

- Nevertheless, if one estimates the parameters from the conditional log-likelihood, this does not matter in the sense that the estimates are the same.
- This is due to the fact that maximising the conditional log-likelihood, which is equivalent to minimising

$$\sum_{t=p+1}^{T} (y_t - c - \sum_{j=1}^{p} \phi_j y_{t-j})^2$$

does not depend on distributional assumptions (a least-squares estimation problem).

Maximum likelihood estimation of non-Gaussian time series

- However, estimator $\hat{\sigma}^2$ depends on the functional form of the log-likelihood function
- Statistical inference, therefore, is affected by the fact that the log-likelihood function is not the correct one.
- Basically the standard errors cannot be trusted.

Likelihood function of a Gaussian MA(1)

Consider the MA(1) process

$$Y_t = \mu + \epsilon_t + \theta \epsilon_{t-1}, \qquad \epsilon_t \sim \text{IIDN}(0, \sigma^2),\ t = 1, ..., T.$$

Note that

$$\mathsf{E}\{Y_t | \epsilon_{t-1}\} = \mu + \theta \epsilon_{t-1}.$$

Let $\boldsymbol{\theta} = (\mu, \theta, \sigma^2)$ be the parameter vector.

Then, the conditional density function

$$f_{Y_t|\epsilon_{t-1}}(y_t | \epsilon_{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_t - \mu - \theta\epsilon_{t-1})^2}{2\sigma^2} \right\}$$

Q: How does this expression look like?

SYDDANSK UNIVERSITET

Likelihood function of a Gaussian MA(1)

Assume, that $\epsilon_0 = 0$. Then

$$Y_1 = \mu + \epsilon_1 \quad \Rightarrow Y_1 \sim N(\mu, \sigma^2) \quad \epsilon_1 = Y_1 - \mu$$

$$f_{Y_2|\epsilon_1,\epsilon_0=0}(y_2|y_1,\epsilon_0;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_2-\mu-\theta\epsilon_1)^2}{2\sigma^2}\right\}$$

Note: $Y_2 = \mu + \epsilon_2 + \theta\epsilon_1 \quad \Rightarrow \epsilon_2 = Y_2 - \mu - \theta\epsilon_1$

Then,

$$f_{Y_2|\epsilon_1,\epsilon_0=0}(y_2|F_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\epsilon_2^2}{2\sigma^2}\right\}$$

Try:

$$f_{Y_3|\,Y_2,Y_1,\epsilon_0=0}(y_2|y_1,\epsilon_0;\boldsymbol{\theta})?$$

Likelihood function of a Gaussian MA(1)

and so on...

$$\epsilon_t = Y_t - \mu - \theta\epsilon_{t-1}$$

$$\begin{aligned} f_{Y_t|Y_{t-1},\dots,Y_1,\epsilon_0}(y_t|y_{t-1},\dots,y_1,\epsilon_0=0;\boldsymbol{\theta}) &= f_{Y_t|\epsilon_{t-1}}(y_t|\epsilon_{t-1};\boldsymbol{\theta}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\epsilon_t^2}{2\sigma^2}\right\} \end{aligned}$$

Likelihood function of a Gaussian MA(1)

The **exact** likelihood function conditioned to $\epsilon_0 = 0$ is:

$$f_{Y_T,\ldots,Y_1|\epsilon_0}(y_T,\ldots,y_1|\epsilon_0) = \\ = \left\{\prod_{\mathbf{t=2}}^{T} f_{Y_t|Y_{t-1},\ldots,Y_2,\epsilon_0=0}(y_T|\ldots)\right\} \\ \times f_{Y_1|\epsilon_0=0}(y_1|\epsilon_0)$$

The **conditional** log likelihood function only has the left term:

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=1}^{T} -\frac{\epsilon_t^2}{2\sigma^2} \qquad (9)$$

Q: Does it look easy to solve?

Conditional maximum likelihood estimation of a Gaussian MA(1)

- Remember that $\epsilon_t = y_t - \mu - \theta\epsilon_{t-1}$
- This means that maximising (9) is a nonlinear optimisation problem (no analytical solution)

Q: Does this numerical optimisation have always a solution?

CMLE of a Gaussian MA(1)

- The solution requires that the MA(1) process is invertible. Consider

$$\begin{aligned}\epsilon_t &= y_t - \mu - \theta(y_{t-1} - \mu - \theta\epsilon_{t-2}) \\ &= y_t - \mu - \theta(y_{t-1} - \mu) + \theta^2(y_{t-2} - \mu - \theta\epsilon_{t-3}) \\ &= \dots \\ &= \sum_{j=0}^{t-1}(-\theta)^j(y_{t-j} - \mu) + (-\theta)^t\epsilon_0.\end{aligned}$$

  The sequence $\{(-\theta)^j(y_{t-j} - \mu)\}$ converges only if $|\theta| < 1$.
- What if the estimate $\hat{\theta} > 1$?

CMLE of a Gaussian MA(1)

- The solution requires that the MA(1) process is invertible. Consider

$$\begin{aligned}\epsilon_t &= y_t - \mu - \theta(y_{t-1} - \mu - \theta\epsilon_{t-2}) \\ &= y_t - \mu - \theta(y_{t-1} - \mu) + \theta^2(y_{t-2} - \mu - \theta\epsilon_{t-3}) \\ &= \dots \\ &= \sum_{j=0}^{t-1}(-\theta)^j(y_{t-j} - \mu) + (-\theta)^t\epsilon_0.\end{aligned}$$

  The sequence $\{(-\theta)^j(y_{t-j} - \mu)\}$ converges only if $|\theta| < 1$.
- What if the estimate $\hat{\theta} > 1$?

Discard it and start the numerical optimisation with $1/\hat{\theta}$ as starting point for the numerical search.

CMLE of a Gaussian MA(1)

What starting point $\tilde{\boldsymbol{\theta}}_{\mathbf{0}} = (\hat{\mu}, \hat{\theta})$ should we used for the optimiser?

- $\hat{\mu} = \overline{y}$.
- $\hat{\theta}$, solution of

$$\hat{\rho} = \hat{\theta}/(1 + \hat{\theta}^2)$$

where

$$\hat{\rho} = \frac{\sum (y_t - \overline{y})(y_{t-1} - \overline{y})}{\sum (y_t - \overline{y})^2}.$$

Exact likelihood function of a MA(1)

Consider the prediction error decomposition of the covariance matrix of the MA(1) process

$$\mathbf{\Omega} = \mathsf{E}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})' = \sigma^2 \begin{bmatrix} 1+\theta^2 & \theta & 0 & \dots & 0 & 0 \\ \theta & 1+\theta^2 & \theta & \dots & 0 & 0 \\ 0 & \theta & 1+\theta^2 & \dots & 0 & 0 \\ & \dots & & & & \\ 0 & 0 & 0 & & \theta & 1+\theta^2 \end{bmatrix}$$

where

$$\mathbf{y} = (y_1, \dots, y_T)',\ \boldsymbol{\mu} = \mu\mathbf{1},\ \mathbf{1} = (1, \dots, 1)' \text{ is a } T \times 1 \text{ vector.}$$

The prediction error decomposition of the covariance matrix

$$\mathbf{\Omega} = \mathbf{ADA}'$$

where

Exact likelihood function of a MA(1)

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ \frac{\theta}{1+\theta^2} & 1 & 0 & \dots & 0 & 0 \\ 0 & \frac{\theta(1+\theta^2)}{1+\theta^2+\theta^4} & 1 & \dots & 0 & 0 \\ & \dots & & & & \\ 0 & 0 & 0 & & \frac{\theta(1+\theta^2+\theta^4+\dots+\theta^{2(T-2)})}{1+\theta^2+\theta^4+\dots+\theta^{2(T-1)}} & 1 \end{bmatrix}$$

and

$$\mathbf{D} = \sigma^2 \begin{bmatrix} 1+\theta^2 & 0 & 0 & \dots & 0 \\ 0 & \frac{1+\theta^2+\theta^4}{1+\theta^2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1+\theta^2+\theta^4+\theta^6}{1+\theta^2+\theta^4} & \dots & 0 \\ & \dots & & & \\ 0 & 0 & 0 & & \frac{1+\theta^2+\theta^4+\dots+\theta^{2T}}{1+\theta^2+\theta^4+\dots+\theta^{2(T-1)}} \end{bmatrix}.$$

Exact likelihood function of a MA(1)

The likelihood

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y},\boldsymbol{\theta}) &= (2\pi)^{-T/2}|\boldsymbol{\Omega}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\} \\
&= (2\pi)^{-T/2}|\mathbf{D}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'(\mathbf{ADA}')^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}
\end{aligned}
\tag{10}
$$

since $|\mathbf{A}| = 1$ ($\mathbf{A}$ is lower triangular with ones on the main diagonal).

Exact likelihood function of a MA(1)

Furthermore, can write

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

where $\widetilde{\mathbf{y}} = \mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ (and, consequently, $\mathbf{A}\widetilde{\mathbf{y}} = \mathbf{y}-\mu$).
The first row of $\tilde{\mathbf{y}}$

$$\widetilde{y}_1 = y_1 - \mu$$

generally,

$$\widetilde{y}_t = y_t - \mu - \frac{\theta(1+\theta^2+\theta^4+\ldots+\theta^{2(t-2)})}{1+\theta^2+\theta^4+\ldots+\theta^{2(t-1)}}\widetilde{y}_{t-1}. \qquad (11)$$

Exact likelihood function of a MA(1)

Matrix $\mathbf{D} = \text{diag}(d_{11}, \ldots, d_{TT})$ with $d_{tt} > 0$, $t = 1, \ldots, T$, so

$$|\mathbf{D}| = \prod_{t=1}^{T} d_{tt} \text{ and } |\mathbf{D}|^{-1} = \prod_{t=1}^{T} d_{tt}^{-1}.$$

The likelihood (10) has the form

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

and the exact log-likelihood becomes

$$\begin{aligned} \mathcal{L}_T(\boldsymbol{\theta}) &= \log f_{\mathbf{Y}}(\mathbf{y}, \boldsymbol{\theta}) \\ &= -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \log d_{tt} - \frac{1}{2} \sum_{t=1}^{T} \frac{\tilde{y}_t^2}{d_{tt}} \end{aligned}$$

Exact likelihood function of a MA(1)

Given θ the sequence $\{\widetilde{y}_t\}$ is calculated recursively from (11).

Whether or not θ is associated with an invertible MA(1) process does not matter.

Conditional likelihood function of a MA(q)

The MA(q) process

$$Y_t = \mu + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}, \qquad \epsilon_t \sim \text{ IIDN}(0, \sigma^2), \; t = 1, ..., T.$$

Starting-values (q of them) needed for the estimation. One alternative:

$$\epsilon_0 = \epsilon_{-1} = ... = \epsilon_{-q+1} = 0.$$

The conditional log-likelihood

$$\mathcal{L}_T(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - \mu - \sum_{j=1}^{q}\theta_j\epsilon_{t-j}\right)^2.$$

Conditional likelihood function of a MA(q)

- Invertibility required if this expression of the log-likelihood is to be applied, i.e. the roots of

$$1 + \theta_1 z + \ldots + \theta_q z^q = 0$$

are outside the unit circle.

- This conditional likelihood function is a generalization of the corresponding function for the MA(1) process and is based on the same triangular factorization as the MA(1) likelihood function.

SYDDANSK UNIVERSITET

Conditional likelihood function of a ARMA(p,q)

The ARMA(p, q) process

$$\begin{aligned} Y_t &= c + \sum_{j=1}^{p} \phi_j Y_{t-j} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \\ \epsilon_t &\sim \text{IIDN}(0, \sigma^2),\ t = 1, ..., T. \end{aligned} \tag{12}$$

Conditional likelihood function of a ARMA(p,q)

The conditional log-likelihood

$$\begin{aligned}\mathcal{L}_T(\boldsymbol{\theta}) &= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 \\ &\quad -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left\{y_t - c - \sum_{j=1}^{p}\phi_j Y_{t-j} - \sum_{j=1}^{q}\theta_j \epsilon_{t-j}\right\}^2.\end{aligned}$$

Starting values $y_0, y_{-1}, \ldots, y_{-p+1}$ and $\epsilon_0, \epsilon_{-1}, \ldots, \epsilon_{-q+1}$ are required for the estimation.

Conditional likelihood function of a ARMA(p,q)

- For ϵ, the natural starting values are $\epsilon_0 = \epsilon_{-1} = ... = \epsilon_{-q+1} = 0$.
- For y, it may be better to use $y_p, y_{p-1}, ..., y_1$, in which case the starting-values for ϵ are $\epsilon_p = \epsilon_{p-1} = ... = \epsilon_{p-q+1} = 0$.

The exact (unconditional) likelihood may be constructed by the triangular decomposition of the covariance matrix but is complicated.

Estimating ARMA(p, q) models using the Hannan-Rissanen method

- The MA component complicates the estimation of the ARMA(p, q) model
- Hannan and Rissanen (1982) have devised a two-step estimation method with the aim of circumventing the problem of iterative estimation.
- The method can also be applied to the estimation of parameters in pure MA models.
- It works as follows:

Estimating ARMA(p, q) models using the Hannan-Rissanen method

1. Estimate by ordinary least squares the long autoregression

$$Y_t = c^* + \sum_{j=1}^{p^*} \phi_j^* Y_{t-j} + \epsilon_t^*$$

that approximates (12) well and calculate the residuals $\widetilde{\epsilon}_t^*$, $t = p^* + 1, ..., T$.

2. Estimate ϕ_j, $j = 1, ..., p$, and θ_j, $j = 1, ..., q$, and c by ordinary least squares from

$$Y_t = c + \sum_{j=1}^{p} \phi_j Y_{t-j} + \epsilon_t + \sum_{j=1}^{q} \theta_j \widetilde{\epsilon}_{t-j}^*, \qquad t = p^* + 1, ..., T.$$

SYDDANSK UNIVERSITET

Estimating ARMA(p, q) models using the Hannan-Rissanen method

The authors have proven that, assuming $\epsilon_t \sim \text{IIDN}(0, \sigma^2)$ and other conditions, the estimators are consistent and asymptotically normal.

Likelihood ratio test

- We want the model that fits our data best
- We use a large model and obtain the estimator $\hat{\boldsymbol{\theta}}$ with loglikelihood function $\mathcal{L}(\hat{\boldsymbol{\theta}})$
- We find a nested model, less parameters and obtain the estimator $\tilde{\boldsymbol{\theta}}$ with loglikelihood function $\mathcal{L}(\tilde{\boldsymbol{\theta}})$

Which model fits the data best?

$$2(\mathcal{L}(\hat{\boldsymbol{\theta}}) - \mathcal{L}(\tilde{\boldsymbol{\theta}})) \sim_a \chi^2_m$$

where m is the number of parameters less in the restricted model.

SYDDANSK UNIVERSITET

Example

Assume we have a loglikelihood function for model 1:

$$\mathcal{L}(\boldsymbol{\theta}) = 1.5\theta_1^2 - 2\theta_2^2$$

First order maximization:

$$\Delta\mathcal{L}(\boldsymbol{\theta}) = (3\theta_1, -4\theta_2)' = (0,0)'$$

The maximum occurs for $\hat{\boldsymbol{\theta}} = (0,0)' \Rightarrow \mathcal{L}(\hat{\boldsymbol{\theta}}) = 0$

SYDDANSK UNIVERSITET

Example

Our hypothesis is that there is another model with only parameter θ_1 such that $H_0 : \theta_2 = \theta_1 + 1$ whose loglikelihood function is:

$$\mathcal{L}(\theta_1) = 1.5\theta_1 - 4(1 + \theta_1)^2$$

$$\Delta\mathcal{L}(\theta_1) = -7\theta_1 - 4 = 0$$

The maximum occures for $\tilde{\boldsymbol{\theta}} = (-4/7, 3/7)' \Rightarrow \mathcal{L}(\tilde{\boldsymbol{\theta}}) = -6/7$

SYDDANSK UNIVERSITET

Example

Likelihood ration test of two models:

$$2(0 + 6/7) = 12/7 = 1.71$$

```
> pchisq(1.71, df = 1, lower.tail = F)
[1] 0.1909854
```

We cannot reject the null hypothesis at 5% significance level and therefore the smaller model provides a better fit.

Exercise

Q: Which is the best model?

```
> ar4.mle = arima(mymodel, order = c(4, 0, 0))
> ar2.mle = arima(mymodel, order = c(2, 0, 0))
> ar3.mle = arima(mymodel, order = c(3, 0, 0))
> pchisq(2 * (ar4.mle$loglik - ar3.mle$loglik), df = 1, lower.tail = F)

[1] 0.3622567

> pchisq(2 * (ar3.mle$loglik - ar2.mle$loglik), df = 1, lower.tail = F)

[1] 5.746692e-57

> pchisq(2 * (ar4.mle$loglik - ar2.mle$loglik), df = 1, lower.tail = F)

[1] 3.788507e-57
```

Variance covariance matrix

The variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ may be estimated by:

$$\left[-\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \;|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right]^{-1}$$

In the example

$$\begin{bmatrix} -3 & 0 \\ 0 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} -0.33 & 0 \\ 0 & 0.25 \end{bmatrix}$$

Because the MLE is asymptotically normal, then we can construct the confidence interval of θ_2

$$0 \pm 1.96\sqrt{0.25}$$

Akaike information criteria (AIC)

AIC can be used to compared two models fit. These do not need to be nested. In this sense it is more general than the LR test. AIC of model i is:

$$AIC_i = -2\mathcal{L}_i + 2k_i$$

- $\mathcal{L}_i$ is the loglikelihood of model i and k_i is the number of parameters to estimate by model i.
- The AIC penalises for overparametrization
- The best fit that of the model with the smallest AIC

Q: Find the AIC of the models in the previous exercise

Answer

```
> ar2.mle$aic
[1] 3139.083
> ar3.mle$aic
[1] 2888.078
> ar4.mle$aic
[1] 2889.248
```

Bayesian information criteria (BIC)

BIC works similarly to the AIC but penalises overparametrisation even more.
BIC of model i is:

$$BIC_i = -2\mathcal{L}_i + k_i \log(k_i)$$

- $\mathcal{L}_i$ is the loglikelihood of model i and k_i is the number of parameters to estimate by model i.
- The BIC penalises for overparametrization
- The best fit that of the model with the smallest BIC

Q: Find the BIC of the models in the previous exercise

SYDDANSK UNIVERSITET

Answer

```
> -2 * ar2.mle$loglik + 4 * log(4)
[1] 3136.628
> -2 * ar3.mle$loglik + 5 * log(5)
[1] 2886.125
> -2 * ar4.mle$loglik + 6 * log(6)
[1] 2887.999
```