Ch. 7: Parameter estimation

- Suppose we want to fit an ARMA model to a time series of interest: how to estimate the parameters of the ARMA model?
- Material part I: Ch.7.1-7.3 (7.2 is optional).

Overview of parameter estimation methods

- Given an observed time series $Y_1, \ldots, Y_n$, and an assumed ARMA model for the data, e.g.

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t \\ & \quad - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}, \end{aligned}$$

  how to estimate the ϕ's and the θ's?
- Various approaches:
  - Method of moments (MoM) (Ch. 7.1)
  - Least-squares (LS) estimation (Ch. 7.2, optional)
  - Maximum likelihood (ML) estimation (Ch. 7.3)
  - Bayesian inference
    Unfortunately not covered in this book, maybe covered if we have sufficient time.
- Easy example for illustration:
  For an AR(1) model $Y_t = \phi Y_{t-1} + e_t$, there are two unknown parameters ϕ and σ_e^2. How can we estimate those parameters?

Method of moments (MoM)

- A very general/simple estimation method.
- As the name suggests: Estimate parameters by setting the sample moments equal to theoretical moments.
- Given a time series, we can estimate
  - the variance γ_0 by the sample mean s^2:

$$\hat{\gamma}_0 = s^2 = \frac{1}{n-1}\sum_{t=1}^{n}(Y_t - \bar{Y})^2,$$

  - the ACF ρ_k by the sample ACF r_k,

  and use those estimates to estimate model parameters.
- Example:
  - For an AR(1) model $Y_t = \phi Y_{t-1} + e_t$, the unknown parameters ϕ and σ_e^2 relate to the variance and ACF as follows:
    $\rho_1 = \phi$ and $\sigma_e^2 = (1 - \rho^2)\gamma_0$.
  - MoM estimation for the AR(1) model:
    Set $\hat{\phi} = \hat{\rho}_1 = r_1$ and $\hat{\gamma}_0^2 = s^2$ to obtain $\hat{\sigma}_e^2 = (1 - r_1^2)s^2$.

MoM for an AR(p) model

- Remember the Yule-Walker equations from Ch.4:

$$\left.\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \cdots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \cdots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \cdots + \phi_p
\end{aligned}\right\}$$

- Using $\hat{\rho}_k = r_k$, we can find estimates $\hat{\phi}_1, \hat{\phi}_3, \ldots, \hat{\phi}_p$ by solving

$$\left.\begin{array}{rrrcrcl}
\phi_1 + & r_1\phi_2 + & r_2\phi_3 & + \cdots + & r_{p-1}\phi_p & = & r_1 \\
r_1\phi_1 + & \phi_2 + & r_1\phi_3 & + \cdots + & r_{p-2}\phi_p & = & r_2 \\
 & & & & & \vdots & \\
r_{p-1}\phi_1 + & r_{p-2}\phi_2 + & r_{p-3}\phi_3 & + \cdots + & \phi_p & = & r_p
\end{array}\right\}$$

- For σ_e^2 we know that

$$\sigma_e^2 = (1 - \phi_1\rho_1 - \phi_2\rho_2 - \ldots - \phi_p\rho_p)\gamma_0,$$

thus how to obtain $\hat{\sigma}_e^2$?

MoM for MA-models

- For MA-models, it's more work to derive MoM estimates.
- Also, for a given sample autocorrelation function, you may end up with a non-invertible solution, so you will have to figure out which one is the invertible solution and use that one.
- Main drawback: the MoM estimator is not very *efficient* for MA models (it has a larger sampling variance than other estimators).

Illustration of performance of MoM estimators

- Table below shows for a number of AR and MA models
  - the true parameters that were used in simulating a time series from that model,
  - the sample size n used in the simulation,
  - the parameter estimates obtained with MoM estimation.
- Conclusion?
- Note: For the MA(1), we discussed that $|\rho_1| < 0.5$. For an observed time series with $|r_1| > 0.5$, we cannot find (real) parameter estimates for the MA(1) model.

	True Parameters			Method-of-Moments Estimates			
Model	θ	ϕ_1	ϕ_2	θ	ϕ_1	ϕ_2	***n***
MA(1)	−0.9			−0.554			120
MA(1)	0.9			0.719			120
MA(1)	−0.9			NA†			60
MA(1)	0.5			−0.314			60
AR(1)		0.9			0.831		60
AR(1)		0.4			0.470		60

Maximum likelihood (ML) estimation

- Note: so far we did not make any assumptions about the probability distribution of Y_t; we only used information on its moments.
- The advantages of ML estimators are:
  - We use all information in the data (not just the information on the first and second moments),
  - We can use large sample results for ML estimators,
  - ML estimators turn out to perform better than MoM estimators for MA-processes, and have the same large sample properties for AR-processes.
- To use ML estimation, we do need to specify probability distributions.
- We will assume that white noise and the Y_t's are normally distributed unless otherwise specified.

Brief review on ML estimation

- Definition of a likelihood function L given a data set:
  $L =$ Joint probability density function for obtaining the data actually observed.
  - E.g., if $Z_i \sim N(\mu, \sigma^2)$, independent for $i = 1, \ldots, n$, then the likelihood function based on observded $z_1, z_2, \ldots, z_n$ is given by

$$
\begin{aligned}
f(z_1, \ldots, z_n) &= \prod_{i=1}^{n} f(z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(z_i - \mu)^2\right) \quad (1) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}^{\,n}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i - \mu)^2\right), \quad (2)
\end{aligned}
$$

    where $f(z_1, \ldots, z_n)$ refers to the joint probability density function for all the Z's combined, and $f(z_i)$ to the density function for Z_i.
- For parameter estimation, we consider L as a function of the unknown parameters, given the observed data,
  $L(\mu, \sigma^2) = f(z_1, \ldots, z_n)$.

Brief review on ML estimation (ctd)

- For parameter estimation, we consider L as a function of the unknown parameters, given the observed data, $L(\mu, \sigma^2) = f(z_1, \dots, z_n)$, and the values for the unknown parameters that maximize $L(\mu, \sigma^2)$ are called the ML estimates.
- Interpretation: ML estimates are those values of the parameters for which the data actually observed are most likely.

ML estimation for time series

- The idea is simple: given $Y_1, \ldots, Y_t$, we want to find those parameters of the ARMA(p, q) model that maximize the likelihood function $f(y_1, \ldots, y_n)$.
- However, dealing with the Y_t's directly is complicated because of the autocorrelation: $f(y_1, \ldots, y_n) \neq \prod_{i=1}^{n} f(y_i)$.
- We can always write down the joint distribution for $(Y_1, Y_2, \ldots, Y_n)$ but it turns out that calculating the covariance matrix and its inverse can be avoided; simpler forms for the likelihood function are available.
- We discuss one approach for the AR(1) model with mean $E(Y_t) = \mu$.

ML estimation for the AR(1)-model

- Some info to start with: remember that we can always write

$$
\begin{aligned}
f(y_2, y_1) &= f(y_2|y_1)f(y_1), \\
f(y_2, y_2, y_1) &= f(y_3, y_2|y_1)f(y_1), \\
&= f(y_3|y_2, y_1)f(y_2|y_1)f(y_1), \\
\ldots
\end{aligned}
$$

  etc., even if the Y's are dependent.
  - E.g., when $n = 4$:

$$
\begin{aligned}
f(y_4, \ldots, y_1) &= f(y_4, \ldots, y_2|y_1)f(y_1), \\
&= f(y_4, y_3|y_2, y_1)f(y_2|y_1)f(y_1), \\
&= f(y_4, y_3|y_2, y_1)f(y_2|y_1)f(y_1), \\
&= f(y_4|y_3, y_2, y_1)f(y_3|y_2, y_1)f(y_2|y_1)f(y_1).
\end{aligned}
$$

- How does that help us to write down the likelihood function for the AR(1) model?

ML estimation for the AR(1)-model

- For AR(1) with mean μ:

$$Y_{t+1} = \mu + \phi(Y_t - \mu) + e_{t+1}, \tag{3}$$

$$Y_{t+1}|(Y_t = y_t, Y_{t-1} = y_{t-1}, \ldots) = \mu + \phi(y_t - \mu) + e_{t+1}, \tag{4}$$

$$Y_{t+1}|(Y_t = y_t, Y_{t-1} = y_{t-1}, \ldots) \sim N(\mu + \phi(y_t - \mu), \sigma_e^2), \tag{5}$$

thus for any $t > 1$:

$$f(y_{t+1}|y_t, y_{t-1}, \ldots) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2\sigma_e^2}(y_{t+1} - \mu - \phi(y_t - \mu))^2\right). \tag{6}$$

For Y_1, we know

$$Y_1 \sim N(\mu, \sigma_e^2/(1 - \phi^2)), \tag{7}$$

$$f(y_1) = \frac{1}{\sqrt{2\pi\sigma_e^2/(1-\phi^2)}} \exp\left(\frac{-1}{2\sigma_e^2/(1-\phi^2)}(y_1 - \mu)^2\right). \tag{8}$$

ML estimation for the AR(1)-model

- Putting the equations together:

$$f(y_t,\ldots,y_1) = f(y_t|y_{t-1},\ldots,y_1)\cdots f(y_2|y_1)f(y_1),$$

$$f(y_{t+1}|y_t,y_{t-1},\ldots,y_1) = \tfrac{1}{\sqrt{2\pi\sigma_e^2}}\exp\left(-\tfrac{1}{2\sigma_e^2}(y_{t+1}-\mu-\phi(y_t-\mu))^2\right)$$

$$f(y_1) = \tfrac{1}{\sqrt{2\pi\sigma_e^2/(1-\phi^2)}}\exp\left(\tfrac{-1}{2\sigma_e^2/(1-\phi^2)}(y_1-\mu)^2\right),$$

gives

$$\begin{aligned}
L(\phi,\mu,\sigma_e^2) &= \prod_{t=2}^{n} f(y_t|y_{t-1},y_{t-2},\ldots,y_1)f(y_1),\\
&= (2\pi\sigma_e^2)^{-n/2}(1-\phi^2)^{1/2}\exp\left(-\frac{1}{2\sigma_e^2}S(\phi,\mu)\right),\ \text{with}\\
S(\phi,\mu) &= (1-\phi^2)(y_1-\mu)^2+\sum_{t=2}^{n}((y_t-\mu)-\phi(y_{t-1}-\mu))^2.
\end{aligned}$$

ML estimation for the AR(1)-model (ctd)

- From previous slide:

$$\begin{aligned} L(\phi, \mu, \sigma_e^2) &= (2\pi\sigma_e^2)^{-n/2}(1-\phi^2)^{1/2} \exp\left(-\frac{1}{2\sigma_e^2} S(\phi, \mu)\right), \text{ with} \\ S(\phi, \mu) &= (1-\phi^2)(y_1-\mu)^2 + \sum_{t=2}^{n}((y_t-\mu) - \phi(y_{t-1}-\mu))^2. \end{aligned}$$

- $S(\phi, \mu)$ is called the sum-of-squares function (specifically, the unconditional sum-of-squares function because it is not conditional on Y_1).
  - Note typo in book Eq 7.3.5 on p.159 (misses a square).
- The ML estimates $\hat{\phi}, \hat{\sigma}_e$ and $\hat{\mu}$ can be found by maximizing the log-likelihood function. There is no-closed form expression for $\hat{\phi}$ and $\hat{\mu}$ so numeric optimization techniques are used.

Summary Ch.7 - part I

- We discussed the MoM approach to estimate the parameters of ARMA models, and the ML estimation method for the AR(1) model.
- You can obtain the likelihood function for AR(p) models in a similar way, by conditioning on $Y_0, Y_{-1}, \dots, Y_{p-1}$. However, obtaining the likelihood function for an ARMA(p, q) process is a bit more challenging (see Ch.9).
- Outlook part II:
  - Standard errors of parameter estimates.
  - How to obtain parameter estimates in R.