Ch 7. PARAMETER ESTIMATION

Time Series Analysis
For an observed series \( \{Z_t\} \), assuming that we have specified the model ARIMA\((p,d,q)\) by Chapter 6, and taken the \( d \)-th difference to get an stationary ARMA\((p,q)\) series \( \{Y_t = \nabla^d Z_t\} \), we will estimate the parameters of the ARMA\((p,q)\) model \( \{Y_t\} \) in this chapter.

7.1 The Method of Moments (MM)

The method of moments is one of the easiest ways to estimate the parameters. We equate sample moments to corresponding theoretical moments and solve the equations to obtain estimates of unknown parameters.
For an observed series \( \{Z_t\} \), assuming that we have specified the model ARIMA\((p,d,q)\) by Chapter 6, and taken the \(d\)-th difference to get an stationary ARMA\((p,q)\) series \( \{Y_t = \nabla^d Z_t\} \), we will estimate the parameters of the ARMA\((p,q)\) model \( \{Y_t\} \) in this chapter.

7.1 The Method of Moments (MM)

The method of moments is one of the easiest ways to estimate the parameters. We equate sample moments to corresponding theoretical moments and solve the equations to obtain estimates of unknown parameters.
7.1.1. AR(p) Models

Examples.

1. **AR(1).** Theoretically $\rho_1 = \phi$. Then $\rho_1$ is estimated by $r_1$ in the method of moments. So $\phi$ can be estimated by:

$$\hat{\phi} = r_1.$$

2. **AR(2).** Replace $\rho_k$ by $r_k$ in Yule-Walker equations:

$$r_1 = \phi_1 + r_1 \phi_2, \quad r_2 = r_1 \phi_1 + \phi_2.$$

Solve the system and we get the estimation

$$\hat{\phi}_1 = \frac{r_1 (1 - r_2)}{1 - r_1^2}, \quad \hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}.$$
7.1.1. AR(p) Models

Examples.

1. **AR(1).** Theoretically $\rho_1 = \phi$. Then $\rho_1$ is estimated by $r_1$ in the method of moments. So $\phi$ can be estimated by:

\[ \hat{\phi} = r_1. \]

2. **AR(2).** Replace $\rho_k$ by $r_k$ in Yule-Walker equations:

\[ r_1 = \phi_1 + r_1 \phi_2, \quad r_2 = r_1 \phi_1 + \phi_2. \]

Solve the system and we get the estimation

\[ \hat{\phi}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2}, \quad \hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}. \]
**AR(p).** For the general AR(p) case, we replace $\rho_k$ by $r_k$ in Yule-Walker equations to get

$$
\begin{align*}
\phi_1 + r_1 \phi_2 + r_2 \phi_3 + \cdots + r_{p-1} \phi_p &= r_1 \\
r_1 \phi_1 + \phi_2 + r_1 \phi_3 + \cdots + r_{p-2} \phi_p &= r_2 \\
r_2 \phi_1 + r_1 \phi_2 + \phi_3 + \cdots + r_{p-3} \phi_p &= r_3 \\
\vdots & \quad \vdots \\
r_{p-1} \phi_1 + r_{p-2} \phi_2 + r_{p-3} \phi_3 + \cdots + \phi_p &= r_p
\end{align*}
$$

Then solve the linear system to get the **Yule-Walker estimates** $\hat{\phi}_1, \cdots, \hat{\phi}_p$. 
7.1.2 MA(q) Models and ARMA(p,q) Models

Examples

1. **MA(1).** Theoretically $\rho_1 = -\frac{\theta}{\theta^2 + 1}$. We replace $\rho_1$ by $r_1$ and solve the quadratic equation for $\theta$. The only invertible solution is $\hat{\theta} = \frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}$.

2. **ARMA(1,1).** Theoretically $\rho_k = \frac{(1 - \theta \phi)(\phi - \theta)}{1 - 2\theta \phi + \theta^2} \phi^{k-1}$ for $k \geq 1$. So $\phi$ can be estimated by $\hat{\phi} = r_2/r_1$. Then we use $r_1 = \frac{(1 - \theta \hat{\phi})(\hat{\phi} - \theta)}{1 - 2\theta \hat{\phi} + \theta^2}$ to find an invertible solution $\hat{\theta}$.

For ARMA(p,q) models, the method of moments results in solving nonlinear numerical equations.

In general, the estimators are very inefficient for models containing MA terms.
7.1.2 MA(q) Models and ARMA(p,q) Models

Examples

1. **MA(1)**. Theoretically $\rho_1 = -\frac{\theta}{\theta^2 + 1}$. We replace $\rho_1$ by $r_1$ and solve the quadratic equation for $\theta$. The only invertible solution is $\hat{\theta} = \frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}$.

2. **ARMA(1,1)**. Theoretically $\rho_k = \frac{(1 - \theta \phi)(\phi - \theta)}{1 - 2\theta \phi + \theta^2} \phi^{k-1}$ for $k \geq 1$. So $\phi$ can be estimated by $\hat{\phi} = r_2/r_1$. Then we use $r_1 = \frac{(1 - \theta \hat{\phi})(\hat{\phi} - \theta)}{1 - 2\theta \hat{\phi} + \theta^2}$ to find an invertible solution $\hat{\theta}$.

- For ARMA(p,q) models, the method of moments results in solving nonlinear numerical equations.
- In general, the estimators are very inefficient for models containing MA terms.
7.1.2 MA(q) Models and ARMA(p,q) Models

Examples

1. **MA(1).** Theoretically $\rho_1 = -\frac{\theta}{\theta^2 + 1}$. We replace $\rho_1$ by $r_1$ and solve the quadratic equation for $\theta$. The only invertible solution is $\hat{\theta} = \frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}$.

2. **ARMA(1,1).** Theoretically $\rho_k = \frac{(1 - \theta \phi)(\phi - \theta)}{1 - 2\theta \phi + \theta^2} \phi^{k-1}$ for $k \geq 1$. So $\phi$ can be estimated by $\hat{\phi} = r_2/r_1$. Then we use $r_1 = \frac{(1 - \theta \hat{\phi})(\hat{\phi} - \theta)}{1 - 2\theta \hat{\phi} + \theta^2}$ to find an invertible solution $\hat{\theta}$.

- For ARMA(p,q) models, the method of moments results in solving nonlinear numerical equations.
- In general, the estimators are very inefficient for models containing MA terms.
7.1.2 MA(q) Models and ARMA(p,q) Models

Examples

1. **MA(1).** Theoretically $\rho_1 = -\frac{\theta}{\theta^2 + 1}$. We replace $\rho_1$ by $r_1$ and solve the quadratic equation for $\theta$. The only invertible solution is $\hat{\theta} = \frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}$.

2. **ARMA(1,1).** Theoretically $\rho_k = \frac{(1 - \theta \phi)(\phi - \theta)}{1 - 2\theta \phi + \theta^2} \phi^{k-1}$ for $k \geq 1$. So $\phi$ can be estimated by $\hat{\phi} = r_2/r_1$. Then we use $r_1 = \frac{(1 - \theta \hat{\phi})(\hat{\phi} - \theta)}{1 - 2\theta \hat{\phi} + \theta^2}$ to find an invertible solution $\hat{\theta}$.

For ARMA(p,q) models, the method of moments results in solving nonlinear numerical equations.

In general, the estimators are very inefficient for models containing MA terms.
7.1.3 The Noise Variance

We estimate $\gamma_0 = \text{Var}(Y_t)$ by the sample variance

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

and use the relationship among $\gamma_0$, $\sigma_e^2$, $\phi$'s and $\theta$'s to estimate $\sigma_e^2$.

1. AR(p) Models: $\hat{\sigma}_e^2 = (1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2 - \cdots - \hat{\phi}_p r_p) s^2$.

2. MA(q) Models: $\hat{\sigma}_e^2 = \frac{s^2}{1 + \hat{\theta}_1^2 + \cdots + \hat{\theta}_q^2}$.
7.1.3 The Noise Variance

We estimate $\gamma_0 = \text{Var}(Y_t)$ by the sample variance

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

and use the relationship among $\gamma_0$, $\sigma_e^2$, $\phi$'s and $\theta$'s to estimate $\sigma_e^2$.

1. **AR(p) Models**: $\hat{\sigma}_e^2 = (1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2 - \cdots - \hat{\phi}_p r_p) s^2$.

2. **MA(q) Models**: $\hat{\sigma}_e^2 = \frac{s^2}{1 + \hat{\theta}_1^2 + \cdots + \hat{\theta}_q^2}$.
7.1.3 The Noise Variance

We estimate $\gamma_0 = \text{Var} (Y_t)$ by the sample variance

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

and use the relationship among $\gamma_0$, $\sigma_e^2$, $\phi$’s and $\theta$’s to estimate $\sigma_e^2$.

1. AR(p) Models: $\hat{\sigma}_e^2 = (1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2 - \cdots - \hat{\phi}_p r_p) s^2$.

2. MA(q) Models: $\hat{\sigma}_e^2 = \frac{s^2}{1 + \hat{\theta}_1^2 + \cdots + \hat{\theta}_q^2}$.
7.1.4 Numerical Examples

The Exhibit shows that Method-of-Moments parameter estimates are good for AR models, but very poor for models involving MA terms. (See the codes in chap7.R)
Now consider the Canadian hare abundance series.

Exhibit 6.27 showed that hare$^5$ is close to a stationary series.
Exhibit 6.29 showed that hare^\frac{5}{2} may be modeled by AR(2) or AR(3). Let us model it as AR(2) here.

```r
> pacf(hare^.5)
```

Exhibit 6.29 showed that hare^{\frac{5}{2}} may be modeled by AR(2) or AR(3). Let us model it as AR(2) here.
Exhibit 6.28 showed that the first two sample ACF of hare^{.5} are $r_1 = 0.736$ and $r_2 = 0.304$. 

```R
> acf(hare^.5)
```

Exhibit 6.28 showed that the first two sample ACF of hare^{.5} are $r_1 = 0.736$ and $r_2 = 0.304$. 

```R
> acf(hare^.5)
```
Exhibit 6.28 showed that the first two sample ACF of hare \(^5\) are \(r_1 = 0.736\) and \(r_2 = 0.304\).

The method-of-moments estimates of \(\phi_1\) and \(\phi_2\) are

\[
\hat{\phi}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2} = 1.1178, \quad \hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2} = -0.519.
\]

The sample mean is 5.82; the sample variance is \(s^2 = 5.88\). So the noise variance is

\[
\hat{\sigma}_{e}^2 = (1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2)s^2 = 1.97.
\]

The estimated model of hare is then

\[
\sqrt{Y_t} - 5.82 = 1.1178(\sqrt{Y_{t-1}} - 5.82) - 0.519(\sqrt{Y_{t-2}} - 5.82) + e_t
\]

with \(\sigma_{e}^2 = 1.97\).
The Oil Price series $\text{oil\_price} \sim \{Y_t\}$

Exhibit 6.32 showed that $\text{diff}(\log(\text{oil\_price})) \approx \text{MA}(1)$. The sample ACF value $r_1 = 0.212$. So the method-of-moments estimate of $\theta$ is

$$\hat{\theta} = \frac{-1 + \sqrt{1 - 4(0.212)^2}}{2(0.212)} = -0.222.$$
The sample mean and the sample variance of diff(log(oil.price)) is 0.004 and 0.0072, respectively. Therefore, the estimated model is

$$\nabla \log(Y_t) = 0.004 + e_t + 0.222e_{t-1},$$

with estimated noise variance

$$\hat{\sigma}_e^2 = \frac{s^2}{1 + \hat{\theta}^2} = \frac{0.0072}{1 + (-0.222)^2} = 0.00686.$$

The standard error of the sample mean of diff(log(oil.price)) can be estimated by

$$\frac{\hat{\gamma}_0}{n} [1 + 2(1 - 1/n)r_1] \approx 0.0065.$$ 

So the observed sample mean 0.004 is not significantly different from 0. We remove the intercept term and get the final model:

$$\log(Y_t) = \log(Y_{t-1}) + e_t + 0.222e_{t-1}.$$
7.2 Least Squares Estimation (CSS)

We introduce a possibly nonzero mean, $\mu$, into our stationary models and treat it as another parameter to be estimated by least squares.
7.2.1 AR(p) Model

Ex. The AR(1) case becomes $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$. View it as a regression model with predictor variable $Y_{t-1}$ and response variable $Y_t$. Least squares estimation then proceeds by minimizing the sum of squares of the differences $Y_t - \mu - \phi(Y_{t-1} - \mu)$. We make the conditional sum-of-squares (CSS) function

$$S_c(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2$$

and estimate $\phi$ and $\mu$ to minimize $S_c(\phi, \mu)$. After computing $\partial S_c / \partial \phi$, $\partial S_c / \partial \mu$, and doing approximation, we get about the same estimation as in method of moments:

$$\hat{\mu} \approx \bar{Y}, \quad \hat{\phi} \approx r_1.$$
Similar process works for stationary AR(p) models. The least squares estimates are about the same as those obtained by the method of moments — $\hat{\mu} \approx \overline{Y}$ and the conditional least squares estimates of the $\phi$’s are approximately obtained by solving the sample Yule-Walker equations.
7.2.2 MA(q) Models

For an invertible MA(q) model:

\[ Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \]

assuming

\[ e_0 = e_{-1} = \cdots = e_{-q} = 0, \]

we compute \( e_t = e_t(\theta_1, \cdots, \theta_q) \) recursively by

\[ e_t = Y_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}, \quad t = 1, 2, \cdots, n. \]

Then we use a multivariate numerical method (e.g. grid search for MA(1) models) to minimize

\[ S_c(\theta_1, \cdots, \theta_q) = \sum_{t=1}^{n} (e_t)^2. \]
7.2.3 ARMA(p,q) Models

Similarly to the MA(q) case, for the ARMA(p,q) model:

\[ Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \]

assuming \( e_p = e_{p-1} = \cdots = e_{p+1-q} = 0 \), we compute

\[ e_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}, \quad t = p+1, \cdots, n, \]

then minimize

\[
S_c(\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q) = \sum_{t=p+1}^{n} e_t^2
\]

numerically to obtain the conditional least squares estimates of all the parameters.

For stationary invertible models, the start-up values \( e_p, e_{p-1}, \cdots, e_{p+1-q} \) will have very little influence on the final estimates of the parameters for large samples.
7.3 Maximum Likelihood (MLE) and Unconditional Least Squares (Unconditional SS)

For series of moderate length and for stochastic seasonal models, the start-up values $e_p = e_{p-1} = \cdots = e_{p+1-q} = 0$ will have great impact on the final estimates for the parameters. So we consider the maximum likelihood estimation. The advantage of the method of maximum likelihood is that all of the information in the data is used rather than just the first and second moments.

**Definition 1**

For any fixed observation $Y_1, \cdots, Y_n$, the **likelihood function** $L$ is the joint probability density (pdf) of obtaining the observed data, considered as a function of the unknown parameters.
Ex. Consider the AR(1) model. Assume that the i.i.d. r.v.s $e_t \sim N(0, \sigma_e^2)$. The pdf of each $e_t$ is

$$(2\pi\sigma_e^2)^{-1/2} \exp \left(-\frac{e_t^2}{2\sigma_e^2}\right), \quad -\infty < e_t < \infty.$$ 

By independence, the joint pdf of $e_2, \ldots, e_n$ is

$$(2\pi\sigma_e^2)^{-(n-1)/2} \exp \left(-\frac{1}{2\sigma_e^2} \sum_{t=2}^{n} e_t^2\right). \quad (2)$$

Given $Y_1 = y_1$, the linear transformation between $e_2, \ldots, e_n$ and $Y_2, \ldots, Y_n$ is

$$Y_2 - \mu = \phi(Y_1 - \mu) + e_2$$
$$Y_3 - \mu = \phi(Y_2 - \mu) + e_2$$
$$\vdots$$
$$Y_n - \mu = \phi(Y_{n-1} - \mu) + e_2$$
The Jacobian of this transformation equals to 1. Thus the joint pdf of $Y_2, Y_3, \ldots, Y_n$ given $Y_1 = y_1$ can be obtained by substituting $e_t = (y_t - \mu) - \phi(y_{t-1} - \mu)$ in (2), namely

$$f(y_2, \ldots, y_n \mid y_1) =$$

$$(2\pi\sigma_e^2)^{-(n-1)/2} \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{t=2}^{n} [(y_t - \mu) - \phi(y_{t-1} - \mu)]^2 \right\}.$$

Now consider the (marginal) distribution of $Y_1$. By the linear process representation $Y_1 - \mu = \sum_{k=0}^{\infty} \phi^k e_{1-k}$, we have

$$\text{Var}(Y_1) = \sum_{k=0}^{\infty} \phi^{2k} \sigma_e^2 = \frac{\sigma_e^2}{1 - \phi^2} \text{ and thus } Y_1 \sim N\left(\mu, \frac{\sigma_e^2}{1 - \phi^2}\right).$$

The marginal pdf of $Y_1$ is

$$f(y_1) = \left(\frac{2\pi\sigma_e^2}{1 - \phi^2}\right)^{-1/2} \exp \left( -\frac{(1 - \phi^2)(y_1 - \mu)^2}{2\sigma_e^2} \right).$$
(Ex. cont.) Multiplying the conditional pdf \( f(y_2, \cdots, y_n | y_1) \) by the marginal pdf of \( Y_1 \) gives us the joint pdf of \( Y_1, Y_2, \cdots, Y_n \), that is, the likelihood function for an AR(1) model (interpreted as a function of \( \phi, \mu \) and \( \sigma_e^2 \)):

\[
L(\phi, \mu, \sigma_e^2) = (2\pi\sigma_e^2)^{-n/2}(1 - \phi^2)^{1/2} \exp \left[ -\frac{1}{2\sigma_e^2} S(\phi, \mu) \right], \tag{3}
\]

where \( S(\phi, \mu) \) is called the unconditional sum-of-squares function and is given by

\[
S(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1 - \phi^2)(Y_1 - \mu)^2.
\] (4)
(Ex. cont.) The difference between $S(\phi, \mu)$ and $S_c(\phi, \mu)$ is only the rightmost term $(1 - \phi^2)(Y_1 - \mu)^2$. Since $S_c(\phi, \mu)$ involves a sum of $n - 1$ components, we have $S(\phi, \mu) \approx S_c(\phi, \mu)$ for large sample size $n$. The effect of the rightmost term in estimating $\phi$ and $\mu$ will be more substantial when the minimum for $\phi$ occurs near the stationarity boundary of $\pm 1$.

In general, it is more convenient to work with the log-likelihood function

$$\ell(\phi, \mu, \sigma_e^2) = \log L(\phi, \mu, \sigma_e^2)$$

$$= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_e^2) + \frac{1}{2} \log(1 - \phi^2) - \frac{1}{2\sigma_e^2} S(\phi, \mu)$$

(5)

To maximize $\ell$, we take partial derivative of $\ell(\phi, \mu, \sigma_e^2)$ w.r.t. $\sigma_e^2$ and get $\sigma_e^2 = \frac{S(\phi, \mu)}{n}$. Since we are estimating two parameters $\phi$ and $\mu$, we obtain a less biased estimator

$$\hat{\sigma}_e^2 = \frac{S(\hat{\phi}, \hat{\mu})}{n - 2}.$$
The derivation of the likelihood function for more general ARMA models is considerably more involved. The estimations are often done by numerical methods.

A compromise between conditional least squares estimates and full maximum likelihood estimates is the unconditional least squares estimates, that is, estimates minimizing $S(\phi, \mu)$ (instead of $L(\phi, \mu, \sigma^2_e)$).
7.4. Properties of the Estimates

The large-sample properties of the maximum likelihood and least squares (conditional or unconditional) estimators are identical and can be obtained by modifying standard maximum likelihood theory.
For large $n$, the estimators are approximately unbiased and normally distributed. The variances and correlations are as follows:

**AR(1):**

$$\text{Var}(\hat{\phi}) \approx \frac{1 - \phi^2}{n}$$  \hspace{1cm} (7.4.9)

**AR(2):**

$$\begin{align*}
\text{Var}(\hat{\phi}_1) & \approx \text{Var}(\hat{\phi}_2) \approx \frac{1 - \phi_2^2}{n} \\
\text{Corr}(\hat{\phi}_1, \hat{\phi}_2) & \approx -\frac{\phi_1}{1 - \phi_2} = -\rho_1
\end{align*}$$  \hspace{1cm} (7.4.10)

**MA(1):**

$$\text{Var}(\hat{\theta}) \approx \frac{1 - \theta^2}{n}$$  \hspace{1cm} (7.4.11)

**MA(2):**

$$\begin{align*}
\text{Var}(\hat{\theta}_1) & \approx \text{Var}(\hat{\theta}_2) \approx \frac{1 - \theta_2^2}{n} \\
\text{Corr}(\hat{\theta}_1, \hat{\theta}_2) & \approx -\frac{\theta_1}{1 - \theta_2}
\end{align*}$$  \hspace{1cm} (7.4.12)

**ARMA(1,1):**

$$\begin{align*}
\text{Var}(\hat{\phi}) & \approx \left[\frac{1 - \phi^2}{n}\right] \left[\frac{1 - \phi \theta}{\phi - \theta}\right]^2 \\
\text{Var}(\hat{\theta}) & \approx \left[\frac{1 - \theta^2}{n}\right] \left[\frac{1 - \phi \theta}{\phi - \theta}\right]^2 \\
\text{Corr}(\hat{\phi}, \hat{\theta}) & \approx \frac{\sqrt{(1 - \phi^2)(1 - \theta^2)}}{1 - \phi \theta}
\end{align*}$$  \hspace{1cm} (7.4.13)
Notice that, in the AR(1) case, the variance of the estimator of \( \phi \) decreases as \( \phi \) approaches \( \pm 1 \). Also notice that even though an AR(1) model is a special case of an AR(2) model, the variance of \( \hat{\phi}_1 \) shown in Equations (7.4.10) shows that our estimation of \( \phi_1 \) will generally suffer if we erroneously fit an AR(2) model when, in fact, \( \phi_2 = 0 \). Similar comments could be made about fitting an MA(2) model when an MA(1) would suffice or fitting an ARMA(1,1) when an AR(1) or an MA(1) is adequate.

For the ARMA(1,1) case, note the denominator of \( \phi - \theta \) in the variances in Equations (7.4.13). If \( \phi \) and \( \theta \) are nearly equal, the variability in the estimators of \( \phi \) and \( \theta \) can be extremely large.

Note that in all of the two-parameter models, the estimates can be highly correlated, even for very large sample sizes.
Exhibit 7.2 gives numerical values for the large-sample approximate standard deviations of the estimates of $\phi$ in an AR(1) model for some $\phi$ and $n$.

Thus, in estimating an AR(1) model with, for example, $n = 100$ and $\phi = 0.7$, we can be about 95% confident that our estimate of $\phi$ is in error by no more than $\pm 2(0.07) = \pm 0.14$.

<table>
<thead>
<tr>
<th>Exhibit 7.2</th>
<th>AR(1) Model Large-Sample Standard Deviations of $\hat{\phi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>$n = 50$</td>
</tr>
<tr>
<td>0.4</td>
<td>0.13</td>
</tr>
<tr>
<td>0.7</td>
<td>0.10</td>
</tr>
<tr>
<td>0.9</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Comparision of Parameter Estimation Methods:

1. For stationary AR(p) models with large samples, the method of moments yields estimators equivalent to least squares and maximum likelihood.

2. For ARMA(p,q) models, the variance for the method-of-moments estimator is always larger than the variance of the maximum likelihood estimator.
Exhibit 7.3 displays the ratio of the large-sample standard deviations for Method of Moments (MM) vs. Maximum Likelihood (MLE) for some $\theta$. These ratios indicate that the method-of-moments estimator should not be used for the MA(1) model. The same advice applies to all models that contain moving average terms.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$SD_{\text{MM}} / SD_{\text{MLE}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>1.07</td>
</tr>
<tr>
<td>0.50</td>
<td>1.42</td>
</tr>
<tr>
<td>0.75</td>
<td>2.66</td>
</tr>
<tr>
<td>0.90</td>
<td>5.33</td>
</tr>
</tbody>
</table>
7.5. Illustrations of Parameter Estimation

Ex. (TS-ch7.R) Consider the simulated MA(1) series ma1.2.s with $\theta = -0.9$:

- MM: $\hat{\theta} = -0.554$ (poor, see Exhibit 7.1);
- conditional SS: $-0.879$;
- unconditional SS: $-0.923$;
- MLE: $-0.915$ (closest).

By Equation (7.4.11), the estimated standard error of $\hat{\theta}$ is

$$\text{Var}\left(\hat{\theta}\right) = \sqrt{\frac{1 - \hat{\theta}^2}{n}} = \sqrt{\frac{1 - (0.915)^2}{120}} \approx 0.04.$$ 

Both the maximum likelihood and conditional/unconditional sum-of-squares estimates are not significantly far from the true value of $-0.9$. 

The arima function estimates an ARIMA(p,d,q) model for the time series passed to it as the first argument.

1. order=c(p,d,q): The ARIMA order is specified by the order argument.

2. method= ’CSS’: Estimate by conditional sum-of-squares method.

3. method=’ML’: Estimate by maximum likelihood method. The default estimation method is maximum likelihood, with initial values determined by the CSS method.

4. The intercept term reported in the output of the arima function is in fact the mean \( \mu \) (instead of \( \theta_0 \)).

5. We may fix the values of some elements by fixed argument. For example, for an ARMA(1,2) model, fixed=c(NA,0.2,NA,0) sets \( ma1 = 0.2 \) (or \( \theta_1 = -0.2 \)) and intercept \( \mu = 0 \).
The output of the arima function is a list structure. Try the following commands:

```
arima(ma1.2.s, order=c(0,0,1),
       method='ML', fixed=c(NA,0))
ma1.2.f=arima(ma1.2.s, order=c(0,0,1),
              method='ML')
str(ma1.2.f)  # show the content of ma1.2.f
fitted(ma1.2.f)  # values of the fitted model
residuals(ma1.2.f)  # residuals for the fitted model
```
Ex. (TS-ch7.R) For the MA(1) simulation ma1.1.s with $\theta = 0.9$:

- MM: $\hat{\theta} = 0.719$ (Exhibit 7.1);
- conditional SS: 0.958;
- unconditional SS: 0.983;
- MLE is 1.

The estimated standard error is about 0.04. The MLE $\hat{\theta} = 1$ corresponds to a noninvertible model. We should perform further investigation.
Ex. We simulate an MA(2) series `ma2.my.s` with \( \theta_1 = 0.4, \theta_2 = 0.21, \) and \( n = 200. \) Then we may work to specify the model, find the best subsets ARMA models, and estimate the parameters of this series. (See TS-ch7.R)

By equation (7.4.12), the standard error of the parameter estimations are

\[
\sqrt{\text{Var} \left( \hat{\theta}_1 \right)} \approx \sqrt{\text{Var} \left( \hat{\theta}_2 \right)} \approx \sqrt{\frac{1 - \theta_2^2}{n}} \approx 0.069.
\]

Both CSS and ML estimates of the parameters are within the confidence intervals.
Now we look at some examples of AR models.

**Ex.** (chap7.R) The dataset `ar1.s` is an AR(1) series with $\phi = 0.9$, and `ar1.2.s` is an AR(1) series with $\phi = 0.4$.

<table>
<thead>
<tr>
<th>Exhibit 7.4</th>
<th>Parameter Estimation for Simulated AR(1) Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter $\phi$</td>
<td>Method-of-Moments Estimate</td>
</tr>
<tr>
<td>0.9</td>
<td>0.831</td>
</tr>
<tr>
<td>0.4</td>
<td>0.470</td>
</tr>
</tbody>
</table>

```r
data(ar1.s); data(ar1.2.s)
ar(ar1.s, order.max=1, AIC=F, method='yw')
ar(ar1.s, order.max=1, AIC=F, method='ols')
ar(ar1.s, order.max=1, AIC=F, method='mle')
ar(ar1.2.s, order.max=1, AIC=F, method='yw')
ar(ar1.2.s, order.max=1, AIC=F, method='ols')
ar(ar1.2.s, order.max=1, AIC=F, method='mle')
```
Ex. (cont) The \texttt{ar} function estimates the AR model for the centered data (that is, mean-corrected data), so the intercept must be zero. Some arguments:

1. \texttt{order.max}: the maximum AR order, must be specified;
2. \texttt{AIC=T} (default): the AR order may be estimated by choosing the order, between 0 and the maximum order, whose model has the smallest AIC;
3. \texttt{AIC=F}: the AR order is set to the maximum AR order;
4. \texttt{method=’yw’}: the parameters are estimated by the Yule-Walker equations;
5. \texttt{method=’ols’}: the parameters are estimated by ordinary least squares;
6. \texttt{method=’mle’}: the parameters are estimated by maximum likelihood estimation (assuming normally distributed white noise error terms).
Ex. (cont) By Equation (7.4.9), the standard errors for the estimated parameter of $\text{ar1.s}$ is

$$\sqrt{\text{Var} (\hat{\phi})} \approx \sqrt{\frac{1 - \hat{\phi}^2}{n}} = \sqrt{\frac{1 - 0.831^2}{60}} \approx 0.07.$$ 

Similarly, the standard errors for the estimated parameter of $\text{ar1.2.s}$ is

$$\sqrt{\text{Var} (\hat{\phi})} \approx 0.11.$$ 

All four methods estimate reasonably well for AR(1) models.
Ex. *(chap7.R)*

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Method-of-Moments Estimates</th>
<th>Conditional SS Estimates</th>
<th>Unconditional SS Estimates</th>
<th>Maximum Likelihood Estimate</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_1 = 1.5$</td>
<td>1.472</td>
<td>1.5137</td>
<td>1.5183</td>
<td>1.5061</td>
<td>120</td>
</tr>
<tr>
<td>$\phi_2 = -0.75$</td>
<td>-0.767</td>
<td>-0.8050</td>
<td>-0.8093</td>
<td>-0.7965</td>
<td>120</td>
</tr>
</tbody>
</table>

By Equation (7.4.10), the standard errors for the estimates are

$$\sqrt{\text{Var}\left(\hat{\phi}_1\right)} \approx \sqrt{\text{Var}\left(\hat{\phi}_2\right)} \approx \sqrt{\frac{1 - \hat{\phi}_2^2}{n}} = \sqrt{\frac{1 - 0.75^2}{120}} \approx 0.06.$$ 

All four methods estimate reasonably well for AR(2) models.
Ex. (TS-ch7.R) We simulate an AR(2) series with $\phi_1 = -0.3, \phi_2 = 0.4$, and $n = 1000$. Then we specify the model, find the best subset ARMA models, and estimate the parameters.

Eq (7.4.10) shows that the standard errors for the estimates are

$$\sqrt{\text{Var} (\hat{\phi}_1)} \approx \sqrt{\text{Var} (\hat{\phi}_2)} \approx \sqrt{\frac{1 - \hat{\phi}_2^2}{n}} = \sqrt{\frac{1 - 0.4^2}{1000}} \approx 0.029.$$ 

All three estimates (MM, CSS, ML) of the parameters are within the confidence intervals.
Ex. (chap7.R) Exhibit 7.6 estimates the parameters for the simulated ARMA(1,1) Model arma11.s with $\phi = 0.6$, $\theta = -0.3$, $n = 100$.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Method-of-Moments Estimates</th>
<th>Conditional SS Estimates</th>
<th>Unconditional SS Estimates</th>
<th>Maximum Likelihood Estimate</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi = 0.6$</td>
<td>0.637</td>
<td>0.5586</td>
<td>0.5691</td>
<td>0.5647</td>
<td>100</td>
</tr>
<tr>
<td>$\theta = -0.3$</td>
<td>-0.2066</td>
<td>-0.3669</td>
<td>-0.3618</td>
<td>-0.3557</td>
<td>100</td>
</tr>
</tbody>
</table>

> data(arma11.s)
> arima(arma11.s, order=c(1,0,1), method='CSS')
> arima(arma11.s, order=c(1,0,1), method='ML')
Let discuss some real time series.

**Ex.** *Consider the industrial chemical property time series color.* The sample PACF strongly suggested an AR(1) model for this series. Here we show the various estimates of the parameter $\phi$ using four different methods of estimation.

<table>
<thead>
<tr>
<th>Exhibit 7.7</th>
<th>Parameter Estimation for the Color Property Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Method-of-Moments Estimate</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.5282</td>
</tr>
</tbody>
</table>

The standard error of the estimates is about

$$\sqrt{\text{Var} \left( \hat{\phi} \right)} \approx \sqrt{\frac{1 - (0.57)^2}{35}} \approx 0.14,$$

so all of the estimates are comparable.
Ex. Consider the Canadian hare abundance series $\text{hare}$. The sample PACF also suggested an AR$(3)$ model. We use MLE for the parameters:

<table>
<thead>
<tr>
<th>Exhibit 7.8 Maximum Likelihood Estimates from R Software: Hare Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients:</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>s.e.</td>
</tr>
</tbody>
</table>
| sigma$^2$ estimated as 1.066: log-likelihood = -46.54, AIC = 101.08

$^\dagger$ The intercept here is the estimate of the process mean $\mu$ — not of $\theta_0$. 

```R
> data(hare)
> arima(sqrt(hare), order=c(3,0,0))
```
Ex. (cont.) The estimated AR(3) model for hare = \{Y_t\} is

\[
\sqrt{Y_t} - 5.6923 = 1.0519(\sqrt{Y_{t-1}} - 5.6923) \\
-0.2292(\sqrt{Y_{t-2}} - 5.6923) \\
-0.3930(\sqrt{Y_{t-3}} - 5.6923) + e_t
\]

or

\[
\sqrt{Y_t} = 3.25 + 1.0519\sqrt{Y_{t-1}} - 0.2292\sqrt{Y_{t-2}} - 0.3930\sqrt{Y_{t-3}} + e_t.
\]

Since the lag 2 term \( \hat{\phi}_2 \) is insignificant from 0, we may drop the term and obtain new estimates of \( \phi_1 \) and \( \phi_3 \) with this subset model.
Ex. Consider the oil price series `oil.price`. The sample ACF suggested an MA(1) model on \( \text{diff}(\log(\text{oil.price})) = \{Y_t\} \). Here we estimate \( \theta \) by the various methods.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Method-of-Moments Estimate</th>
<th>Conditional SS Estimate</th>
<th>Unconditional SS Estimate</th>
<th>Maximum Likelihood Estimate</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta )</td>
<td>-0.2225</td>
<td>-0.2731</td>
<td>-0.2954</td>
<td>-0.2956</td>
<td>241</td>
</tr>
</tbody>
</table>

The method-of-moments estimate differs quite a bit from the others. The others are nearly equal given their standard errors of about 0.07.

```r
> data(oil.price)
> arima(log(oil.price), order=c(0,1,1), method='CSS')
> arima(log(oil.price), order=c(0,1,1), method='ML')
```