Chapter 7

Parameter Estimation

This chapter deals with the problem of estimating the parameters of an ARIMA model based on the observed time series $Y_1, Y_2,\ldots, Y_n$. We assume that a model has already been specified; that is, we have specified values for p, d, and q using the methods of Chapter 6. With regard to nonstationarity, since the dth difference of the observed series is assumed to be a stationary ARMA(p,q) process, we need only concern ourselves with the problem of estimating the parameters in such stationary models. In practice, then we treat the dth difference of the original time series as the time series from which we estimate the parameters of the complete model. For simplicity, we shall let $Y_1, Y_2,\ldots, Y_n$ denote our observed *stationary* process even though it may be an appropriate difference of the original series. We first discuss the method-of-moments estimators, then the least squares estimators, and finally full maximum likelihood estimators.

7.1 The Method of Moments

The method of moments is frequently one of the easiest, if not the most efficient, methods for obtaining parameter estimates. The method consists of equating sample moments to corresponding theoretical moments and solving the resulting equations to obtain estimates of any unknown parameters. The simplest example of the method is to estimate a stationary process mean by a sample mean. The properties of this estimator were studied extensively in Chapter 3.

Autoregressive Models

Consider first the AR(1) case. For this process, we have the simple relationship $\rho_1 = \phi$. In the method of moments, ρ_1 is equated to r_1, the lag 1 sample autocorrelation. Thus we can estimate ϕ by

$$\hat{\phi} = r_1 \tag{7.1.1}$$

Now consider the AR(2) case. The relationships between the parameters ϕ_1 and ϕ_2 and various moments are given by the Yule-Walker equations (4.3.13) on page 72:

$$\rho_1 = \phi_1 + \rho_1\phi_2 \quad \text{and} \quad \rho_2 = \rho_1\phi_1 + \phi_2$$

The method of moments replaces ρ_1 by r_1 and ρ_2 by r_2 to obtain

$$r_1 = \phi_1 + r_1\phi_2 \quad \text{and} \quad r_2 = r_1\phi_1 + \phi_2$$

which are then solved to obtain

$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2} \quad \text{and} \quad \hat{\phi}_2 = \frac{r_2-r_1^2}{1-r_1^2} \tag{7.1.2}$$

The general AR(p) case proceeds similarly. Replace ρ_k by r_k throughout the Yule-Walker equations on page 79 (or page 114) to obtain

$$\left.\begin{array}{rcl} \phi_1 + \quad r_1\phi_2 + \quad r_2\phi_3 \;+ \cdots +\; r_{p-1}\phi_p &=& r_1 \\ r_1\phi_1 + \quad \phi_2 + \quad r_1\phi_3 \;+ \cdots +\; r_{p-2}\phi_p &=& r_2 \\ & \vdots & \\ r_{p-1}\phi_1 + r_{p-2}\phi_2 + r_{p-3}\phi_3 \;+ \cdots +\; \phi_p &=& r_p \end{array}\right\} \tag{7.1.3}$$

These linear equations are then solved for $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$. The Durbin-Levinson recursion of Equation (6.2.9) on page 115 provides a convenient method of solution but is subject to substantial round-off errors if the solution is close to the boundary of the stationarity region. The estimates obtained in this way are also called **Yule-Walker estimates**.

Moving Average Models

Surprisingly, the method of moments is not nearly as convenient when applied to moving average models. Consider the simple MA(1) case. From Equations (4.2.2) on page 57, we know that

$$\rho_1 = -\frac{\theta}{1+\theta^2}$$

Equating ρ_1 to r_1, we are led to solve a quadratic equation in θ. If $|r_1| < 0.5$, then the two real roots are given by

$$-\frac{1}{2r_1} \pm \sqrt{\frac{1}{4r_1^2} - 1}$$

As can be easily checked, the product of the two solutions is always equal to 1; therefore, only one of the solutions satisfies the invertibility condition $|\theta| < 1$.

After further algebraic manipulation, we see that the invertible solution can be written as

$$\hat{\theta} = \frac{-1 + \sqrt{1-4r_1^2}}{2r_1} \tag{7.1.4}$$

If $r_1 = \pm 0.5$, unique, real solutions exist, namely ∓ 1, but neither is invertible. If $|r_1| > 0.5$ (which is certainly possible even though $|\rho_1| < 0.5$), no real solutions exist, and so the method of moments fails to yield an estimator of θ. Of course, if $|r_1| > 0.5$, the specification of an MA(1) model would be in considerable doubt.

For higher-order MA models, the method of moments quickly gets complicated. We can use Equations (4.2.5) on page 65 and replace ρ_k by r_k for $k = 1, 2,\ldots, q$, to obtain q equations in q unknowns $\theta_1, \theta_2,\ldots, \theta_q$. The resulting equations are highly nonlinear in the θ's, however, and their solution would of necessity be numerical. In addition, there will be multiple solutions, of which only one is invertible. We shall not pursue this further since we shall see in Section 7.4 that, for MA models, the method of moments generally produces poor estimates.

Mixed Models

We consider only the ARMA(1,1) case. Recall Equation (4.4.5) on page 78,

$$\rho_k = \frac{(1-\theta\phi)(\phi-\theta)}{1-2\theta\phi+\theta^2}\phi^{k-1} \quad \text{for } k \geq 1$$

Noting that $\rho_2/\rho_1 = \phi$, we can first estimate ϕ as

$$\hat{\phi} = \frac{r_2}{r_1} \tag{7.1.5}$$

Having done so, we can then use

$$r_1 = \frac{(1-\theta\hat{\phi})(\hat{\phi}-\theta)}{1-2\theta\hat{\phi}+\theta^2} \tag{7.1.6}$$

to solve for $\hat{\theta}$. Note again that a quadratic equation must be solved and only the invertible solution, if any, retained.

Estimates of the Noise Variance

The final parameter to be estimated is the noise variance, σ_e^2. In all cases, we can first estimate the process variance, $\gamma_0 = Var(Y_t)$, by the sample variance

$$s^2 = \frac{1}{n-1}\sum_{t=1}^{n}(Y_t-\bar{Y})^2 \tag{7.1.7}$$

and use known relationships from Chapter 4 among γ_0, σ_e^2, and the θ's and ϕ's to estimate σ_e^2.

For the AR(p) models, Equation (4.3.31) on page 77 yields

$$\hat{\sigma}_e^2 = (1-\hat{\phi}_1 r_1-\hat{\phi}_2 r_2-\cdots-\hat{\phi}_p r_p)s^2 \tag{7.1.8}$$

In particular, for an AR(1) process,

$$\hat{\sigma}_e^2 = (1-r_1^2)s^2$$

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

For the MA(q) case, we have, using Equation (4.2.4) on page 65,

$$\hat{\sigma}_e^2 = \frac{s^2}{1+\hat{\theta}_1^2+\hat{\theta}_2^2+\cdots+\hat{\theta}_q^2} \tag{7.1.9}$$

For the ARMA(1,1) process, Equation (4.4.4) on page 78 yields

$$\hat{\sigma}_e^2 = \frac{1-\hat{\phi}^2}{1-2\hat{\phi}\hat{\theta}+\hat{\theta}^2}s^2 \tag{7.1.10}$$

Numerical Examples

The table in Exhibit 7.1 displays method-of-moments estimates for the parameters from several simulated time series. Generally speaking, the estimates for all the autoregressive models are fairly good but the estimates for the moving average models are not acceptable. It can be shown that theory confirms this observation—method-of-moments estimators are very inefficient for models containing moving average terms.

Exhibit 7.1 Method-of-Moments Parameter Estimates for Simulated Series

	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
AR(2)		1.5	−0.75		1.472	−0.767	120

[†] No method-of-moments estimate exists since $r_1 = 0.544$ for this simulation.

```
> data(ma1.2.s); data(ma1.1.s); data(ma1.3.s); data(ma1.4.s)
> estimate.ma1.mom(ma1.2.s); estimate.ma1.mom(ma1.1.s)
> estimate.ma1.mom(ma1.3.s); estimate.ma1.mom(ma1.4.s)
> arima(ma1.4.s,order=c(0,0,1),method='CSS',include.mean=F)
> data(ar1.s); data(ar1.2.s)
> ar(ar1.s,order.max=1,AIC=F,method='yw')
> ar(ar1.2.s,order.max=1,AIC=F,method='yw')
> data(ar2.s)
> ar(ar2.s,order.max=2,AIC=F,method='yw')
```

Consider now some actual time series. We start with the Canadian hare abundance series. Since we found in Exhibit 6.27 on page 136 that a square root transformation was appropriate here, we base all modeling on the square root of the original abundance numbers. We illustrate the estimation of an AR(2) model with the hare data, even though we shall show later that an AR(3) model provides a better fit to the data. The first two sample autocorrelations displayed in Exhibit 6.28 on page 137 are $r_1 = 0.736$ and $r_2 = 0.304$. Using Equations (7.1.2), the method-of-moments estimates of ϕ_1 and ϕ_2 are

$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2} = \frac{0.736(1-0.304)}{1-(0.736)^2} = 1.1178 \tag{7.1.11}$$

and

$$\hat{\phi}_2 = \frac{r_2 - r_1^2}{1-r_1^2} = \frac{0.304-(0.736)^2}{1-(0.736)^2} = -0.519 \tag{7.1.12}$$

The sample mean and variance of this series (after taking the square root) are found to be 5.82 and 5.88, respectively. Then, using Equation (7.1.8), we estimate the noise variance as

$$\begin{aligned}\hat{\sigma}_e^2 &= (1-\hat{\phi}_1 r_1 - \hat{\phi}_2 r_2)s^2 \\ &= [1-(1.1178)(0.736)-(-0.519)(0.304)](5.88) \\ &= 1.97\end{aligned} \tag{7.1.13}$$

The estimated model (in original terms) is then

$$\sqrt{Y_t} - 5.82 = 1.1178(\sqrt{Y_{t-1}} - 5.82) - 0.519(\sqrt{Y_{t-2}} - 5.82) + e_t \tag{7.1.14}$$

or

$$\sqrt{Y_t} = 2.335 + 1.1178\sqrt{Y_{t-1}} - 0.519\sqrt{Y_{t-2}} + e_t \tag{7.1.15}$$

with estimated noise variance of 1.97.

Consider now the oil price series. Exhibit 6.32 on page 140 suggested that we specify an MA(1) model for the first differences of the logarithms of the series. The lag 1 sample autocorrelation in that exhibit is 0.212, so the method-of-moments estimate of θ is

$$\hat{\theta} = \frac{-1+\sqrt{1-4(0.212)^2}}{2(0.212)} = -0.222 \tag{7.1.16}$$

The mean of the differences of the logs is 0.004 and the variance is 0.0072. The estimated model is

$$\nabla \log(Y_t) = 0.004 + e_t + 0.222e_{t-1} \tag{7.1.17}$$

or

$$\log(Y_t) = \log(Y_{t-1}) + 0.004 + e_t + 0.222e_{t-1} \tag{7.1.18}$$

with estimated noise variance of

$$\hat{\sigma}_e^2 = \frac{s^2}{1+\hat{\theta}^2} = \frac{0.0072}{1+(-0.222)^2} = 0.00686 \tag{7.1.19}$$

Using Equation (3.2.3) on page 28 with estimated parameters yields a standard error of the sample mean of 0.0060. Thus, the observed sample mean of 0.004 is not significantly different from zero and we would remove the constant term from the model, giving a final model of

$$\log(Y_t) = \log(Y_{t-1}) + e_t + 0.222e_{t-1} \tag{7.1.20}$$

7.2 Least Squares Estimation

Because the method of moments is unsatisfactory for many models, we must consider other methods of estimation. We begin with least squares. For autoregressive models, the ideas are quite straightforward. At this point, we introduce a possibly nonzero mean, μ, into our stationary models and treat it as another parameter to be estimated by least squares.

Autoregressive Models

Consider the first-order case where

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t \tag{7.2.1}$$

We can view this 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

$$(Y_t - \mu) - \phi(Y_{t-1} - \mu)$$

Since only $Y_1, Y_2, \ldots, Y_n$ are observed, we can only sum from $t = 2$ to $t = n$. Let

$$S_c(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 \tag{7.2.2}$$

This is usually called the **conditional sum-of-squares function**. (The reason for the term *conditional* will become apparent later on.) According to the principle of least squares, we estimate ϕ and μ by the respective values that minimize $S_c(\phi,\mu)$ given the observed values of $Y_1, Y_2, \ldots, Y_n$.

Consider the equation $\partial S_c / \partial\mu = 0$. We have

$$\frac{\partial S_c}{\partial\mu} = \sum_{t=2}^{n} 2[(Y_t - \mu) - \phi(Y_{t-1} - \mu)](-1 + \phi) = 0$$

or, simplifying and solving for μ,

$$\mu = \frac{1}{(n-1)(1-\phi)}\left[\sum_{t=2}^{n} Y_t - \phi \sum_{t=2}^{n} Y_{t-1}\right] \tag{7.2.3}$$

Now, for large n,

$$\frac{1}{n-1}\sum_{t=2}^{n} Y_t \approx \frac{1}{n-1}\sum_{t=2}^{n} Y_{t-1} \approx \bar{Y}$$

Thus, regardless of the value of ϕ, Equation (7.2.3) reduces to

$$\hat{\mu} \approx \frac{1}{1-\phi}(\bar{Y} - \phi\bar{Y}) = \bar{Y} \tag{7.2.4}$$

We sometimes say, except for end effects, $\hat{\mu} = \bar{Y}$.

Consider now the minimization of $S_c(\phi, \bar{Y})$ with respect to ϕ. We have

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

Setting this equal to zero and solving for ϕ yields

$$\hat{\phi} = \frac{\sum_{t=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=2}^{n}(Y_{t-1} - \bar{Y})^2}$$

Except for one term missing in the denominator, namely $(Y_n - \bar{Y})^2$, this is the same as r_1. The lone missing term is negligible for stationary processes, and thus the least squares and method-of-moments estimators are nearly identical, especially for large samples.

For the general AR(p) process, the methods used to obtain Equations (7.2.3) and (7.2.4) can easily be extended to yield the same result, namely

$$\hat{\mu} = \bar{Y} \tag{7.2.5}$$

To generalize the estimation of the ϕ's, we consider the second-order model. In accordance with Equation (7.2.5), we replace μ by $\bar{Y}$ in the conditional sum-of-squares function, so

$$S_c(\phi_1, \phi_2, \bar{Y}) = \sum_{t=3}^{n}[(Y_t - \bar{Y}) - \phi_1(Y_{t-1} - \bar{Y}) - \phi_2(Y_{t-2} - \bar{Y})]^2 \tag{7.2.6}$$

Setting $\partial S_c / \partial \phi_1 = 0$, we have

$$-2\sum_{t=3}^{n}[(Y_t - \bar{Y}) - \phi_1(Y_{t-1} - \bar{Y}) - \phi_2(Y_{t-2} - \bar{Y})](Y_{t-1} - \bar{Y}) = 0 \tag{7.2.7}$$

which we can rewrite as

$$\sum_{t=3}^{n}(Y_t-\overline{Y})(Y_{t-1}-\overline{Y}) = \left(\sum_{t=3}^{n}(Y_{t-1}-\overline{Y})^2\right)\phi_1 + \left(\sum_{t=3}^{n}(Y_{t-1}-\overline{Y})(Y_{t-2}-\overline{Y})\right)\phi_2 \tag{7.2.8}$$

The sum of the lagged products $\sum_{t=3}^{n}(Y_t-\overline{Y})(Y_{t-1}-\overline{Y})$ is very nearly the numerator of r_1—we are missing one product, $(Y_2-\overline{Y})(Y_1-\overline{Y})$. A similar situation exists for $\sum_{t=3}^{n}(Y_{t-1}-\overline{Y})(Y_{t-2}-\overline{Y})$, but here we are missing $(Y_n-\overline{Y})(Y_{n-1}-\overline{Y})$. If we divide both sides of Equation (7.2.8) by $\sum_{t=3}^{n}(Y_t-\overline{Y})^2$, then, except for end effects, which are negligible under the stationarity assumption, we obtain

$$r_1 = \phi_1 + r_1\phi_2 \tag{7.2.9}$$

Approximating in a similar way with the equation $\partial S_c/\partial\phi_2 = 0$ leads to

$$r_2 = r_1\phi_1 + \phi_2 \tag{7.2.10}$$

But Equations (7.2.9) and (7.2.10) are just the sample Yule-Walker equations for an AR(2) model.

Entirely analogous results follow for the general stationary AR(p) case: To an excellent approximation, the conditional least squares estimates of the ϕ's are obtained by solving the sample Yule-Walker equations (7.1.3).†

Moving Average Models

Consider now the least-squares estimation of θ in the MA(1) model:

$$Y_t = e_t - \theta e_{t-1} \tag{7.2.11}$$

At first glance, it is not apparent how a least squares or regression method can be applied to such models. However, recall from Equation (4.4.2) on page 77 that invertible MA(1) models can be expressed as

$$Y_t = -\theta Y_{t-1} - \theta^2 Y_{t-2} - \theta^3 Y_{t-3} - \cdots + e_t$$

an autoregressive model but of infinite order. Thus least squares can be meaningfully carried out by choosing a value of θ that minimizes

† We note that Lai and Wei (1983) established that the conditional least squares estimators are consistent even for nonstationary autoregressive models where the Yule-Walker equations do not apply.

$$S_c(\theta) = \sum (e_t)^2 = \sum [Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \theta^3 Y_{t-3} + \cdots]^2 \tag{7.2.12}$$

where, implicitly, $e_t = e_t(\theta)$ is a function of the observed series and the unknown parameter θ.

It is clear from Equation (7.2.12) that the least squares problem is *nonlinear* in the parameters. We will not be able to minimize $S_c(\theta)$ by taking a derivative with respect to θ, setting it to zero, and solving. Thus, even for the simple MA(1) model, we must resort to techniques of numerical optimization. Other problems exist in this case: We have not shown explicit limits on the summation in Equation (7.2.12) nor have we said how to deal with the infinite series under the summation sign.

To address these issues, consider evaluating $S_c(\theta)$ for a *single given value* of θ. The only Y's we have available are our observed series, $Y_1, Y_2, \ldots, Y_n$. Rewrite Equation (7.2.11) as

$$e_t = Y_t + \theta e_{t-1} \tag{7.2.13}$$

Using this equation, $e_1, e_2, \ldots, e_n$ can be calculated recursively if we have the initial value e_0. A common approximation is to set $e_0 = 0$—its expected value. Then, *conditional on* $e_0 = 0$, we can obtain

$$\left.\begin{aligned} e_1 &= Y_1 \\ e_2 &= Y_2 + \theta e_1 \\ e_3 &= Y_3 + \theta e_2 \\ &\vdots \\ e_n &= Y_n + \theta e_{n-1} \end{aligned}\right\} \tag{7.2.14}$$

and thus calculate $S_c(\theta) = \sum (e_t)^2$, conditional on $e_0 = 0$, for that single given value of θ.

For the simple case of one parameter, we could carry out a grid search over the invertible range $(-1,+1)$ for θ to find the minimum sum of squares. For more general MA(q) models, a numerical optimization algorithm, such as Gauss-Newton or Nelder-Mead, will be needed.

For higher-order moving average models, the ideas are analogous and no new difficulties arise. We compute $e_t = e_t(\theta_1, \theta_2, \ldots, \theta_q)$ recursively from

$$e_t = Y_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \tag{7.2.15}$$

with $e_0 = e_{-1} = \cdots = e_{-q} = 0$. The sum of squares is minimized jointly in $\theta_1, \theta_2, \ldots, \theta_q$ using a multivariate numerical method.

Mixed Models

Consider the ARMA(1,1) case

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1} \tag{7.2.16}$$

As in the pure MA case, we consider $e_t = e_t(\phi,\theta)$ and wish to minimize $S_c(\phi, \theta) = \sum e_t^2$. We can rewrite Equation (7.2.16) as

$$e_t = Y_t - \phi Y_{t-1} + \theta e_{t-1} \tag{7.2.17}$$

To obtain e_1, we now have an additional "startup" problem, namely Y_0. One approach is to set $Y_0 = 0$ or to $\bar{Y}$ if our model contains a nonzero mean. However, a better approach is to begin the recursion at $t = 2$, thus avoiding Y_0 altogether, and simply minimize

$$S_c(\phi, \theta) = \sum_{t=2}^{n} e_t^2$$

For the general ARMA(p,q) model, we compute

$$\begin{aligned} e_t = {} & Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p} \\ & + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \end{aligned} \tag{7.2.18}$$

with $e_p = e_{p-1} = \cdots = e_{p+1-q} = 0$ and then minimize $S_c(\phi_1,\phi_2,\ldots,\phi_p,\theta_1,\theta_2,\ldots,\theta_q)$ numerically to obtain the conditional least squares estimates of all the parameters.

For parameter sets $\theta_1, \theta_2,\ldots, \theta_q$ corresponding to invertible models, the start-up values $e_p, e_{p-1},\ldots, e_{p+1-q}$ will have very little influence on the final estimates of the parameters for large samples.

7.3 Maximum Likelihood and Unconditional Least Squares

For series of moderate length and also for stochastic seasonal models to be discussed in Chapter 10, the start-up values $e_p = e_{p-1} = \cdots = e_{p+1-q} = 0$ will have a more pronounced effect on the final estimates for the parameters. Thus we are led to consider the more difficult problem of 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, as is the case with least squares. Another advantage is that many large-sample results are known under very general conditions. One disadvantage is that we must for the first time work specifically with the joint probability density function of the process.

Maximum Likelihood Estimation

For any set of observations, $Y_1, Y_2,\ldots, Y_n$, time series or not, the likelihood function L is defined to be the joint probability density of obtaining the data actually observed. However, it is considered as a function of the unknown parameters in the model with the observed data held fixed. For ARIMA models, L will be a function of the ϕ's, θ's, μ, and σ_e^2 given the observations $Y_1, Y_2,\ldots, Y_n$. The maximum likelihood estimators are then defined as those values of the parameters for which the data actually observed are *most likely*, that is, the values that maximize the likelihood function.

We begin by looking in detail at the AR(1) model. The most common assumption is that the white noise terms are independent, normally distributed random variables with zero means and common standard deviation σ_e. The probability density function (pdf) of each e_t is then

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

and, by independence, the joint pdf for $e_2, e_3, \ldots, e_n$ is

$$(2\pi\sigma_e^2)^{-(n-1)/2}\exp\left(-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n} e_t^2\right) \tag{7.3.1}$$

Now consider

$$\left.\begin{aligned} Y_2 - \mu &= \phi(Y_1 - \mu) + e_2 \\ Y_3 - \mu &= \phi(Y_2 - \mu) + e_3 \\ &\vdots \\ Y_n - \mu &= \phi(Y_{n-1} - \mu) + e_n \end{aligned}\right\} \tag{7.3.2}$$

If we condition on $Y_1 = y_1$, Equation (7.3.2) defines a linear transformation between $e_2, e_3, \ldots, e_n$ and $Y_2, Y_3, \ldots, Y_n$ (with Jacobian equal to 1). Thus the joint pdf of $Y_2, Y_3, \ldots, Y_n$ given $Y_1 = y_1$ can be obtained by using Equation (7.3.2) to substitute for the e's in terms of the Y's in Equation (7.3.1). Thus we get

$$\begin{aligned} f(y_2, y_3, \ldots, y_n | y_1) &= (2\pi\sigma_e^2)^{-(n-1)/2} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n}[(y_t - \mu) - \phi(y_{t-1} - \mu)]^2\right\} \end{aligned} \tag{7.3.3}$$

Now consider the (marginal) distribution of Y_1. It follows from the linear process representation of the AR(1) process (Equation (4.3.8) on page 70) that Y_1 will have a normal distribution with mean μ and variance $\sigma_e^2/(1-\phi^2)$. Multiplying the conditional pdf in Equation (7.3.3) by the marginal pdf of Y_1 gives us the joint pdf of $Y_1, Y_2, \ldots, Y_n$ that we require. Interpreted as a function of the parameters ϕ, μ, and σ_e^2, the likelihood function for an AR(1) model is given by

$$L(\phi, \mu, \sigma_e^2) = (2\pi\sigma_e^2)^{-n/2}(1-\phi^2)^{1/2}\exp\left[-\frac{1}{2\sigma_e^2}S(\phi, \mu)\right] \tag{7.3.4}$$

where

$$S(\phi, \mu) = \sum_{t=2}^{n}[(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1-\phi^2)(Y_1 - \mu) \tag{7.3.5}$$

The function $S(\phi, \mu)$ is called the **unconditional sum-of-squares function**.

As a general rule, the logarithm of the likelihood function is more convenient to work with than the likelihood itself. For the AR(1) case, the **log-likelihood function**, denoted $\ell(\phi, \mu, \sigma_e^2)$, is given by

$$\ell(\phi, \mu, \sigma_e^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_e^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma_e^2}S(\phi, \mu) \tag{7.3.6}$$

For given values of ϕ and μ, $\ell(\phi, \mu, \sigma_e^2)$ can be maximized analytically with respect to σ_e^2 in terms of the yet-to-be-determined estimators of ϕ and μ. We obtain

$$\hat{\sigma}_e^2 = \frac{S(\hat{\phi}, \hat{\mu})}{n} \tag{7.3.7}$$

As in many other similar contexts, we usually divide by $n - 2$ rather than n (since we are estimating *two* parameters, ϕ and μ) to obtain an estimator with less bias. For typical time series sample sizes, there will be very little difference.

Consider now the estimation of ϕ and μ. A comparison of the unconditional sum-of-squares function $S(\phi,\mu)$ with the earlier conditional sum-of-squares function $S_c(\phi,\mu)$ of Equation (7.2.2) on page 154, reveals one simple difference:

$$S(\phi, \mu) = S_c(\phi, \mu) + (1-\phi^2)(Y_1 - \mu)^2 \tag{7.3.8}$$

Since $S_c(\phi,\mu)$ involves a sum of $n - 1$ components, whereas $(1-\phi^2)(Y_1-\mu)^2$ does not involve n, we shall have $S(\phi, \mu) \approx S_c(\phi, \mu)$. Thus the values of ϕ and μ that minimize $S(\phi,\mu)$ or $S_c(\phi,\mu)$ should be very similar, at least for larger sample sizes. The effect of the rightmost term in Equation (7.3.8) will be more substantial when the minimum for ϕ occurs near the stationarity boundary of ± 1.

Unconditional Least Squares

As a compromise between conditional least squares estimates and full maximum likelihood estimates, we might consider obtaining unconditional least squares estimates; that is, estimates minimizing $S(\phi,\mu)$. Unfortunately, the term $(1-\phi^2)(Y_1-\mu)^2$ causes the equations $\partial S/\partial\phi = 0$ and $\partial S/\partial\mu = 0$ to be nonlinear in ϕ and μ, and reparameterization to a constant term $\theta_0 = \mu(1 - \phi)$ does not improve the situation substantially. Thus minimization must be carried out numerically. The resulting estimates are called **unconditional least squares estimates**.

The derivation of the likelihood function for more general ARMA models is considerably more involved. One derivation may be found in Appendix H: State Space Models on page 222. We refer the reader to Brockwell and Davis (1991) or Shumway and Stoffer (2006) for even more details.

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. Details can be found in Shumway and Stoffer (2006, pp. 125–129). We shall look at the results and their implications for simple ARMA models.

For large n, the estimators are approximately unbiased and normally distributed. The variances and correlations are as follows:

$$\text{AR(1):}\ Var(\hat{\phi}) \approx \frac{1-\phi^2}{n} \tag{7.4.9}$$

$$\text{AR(2):} \begin{cases} Var(\hat{\phi}_1) \approx Var(\hat{\phi}_2) \approx \dfrac{1-\phi_2^2}{n} \\ Corr(\hat{\phi}_1, \hat{\phi}_2) \approx -\dfrac{\phi_1}{1-\phi_2} = -\rho_1 \end{cases} \tag{7.4.10}$$

$$\text{MA(1):}\ Var(\hat{\theta}) \approx \frac{1-\theta^2}{n} \tag{7.4.11}$$

$$\text{MA(2):} \begin{cases} Var(\hat{\theta}_1) \approx Var(\hat{\theta}_2) \approx \dfrac{1-\theta_2^2}{n} \\ Corr(\hat{\theta}_1, \hat{\theta}_2) \approx -\dfrac{\theta_1}{1-\theta_2} \end{cases} \tag{7.4.12}$$

$$\text{ARMA(1,1):} \begin{cases} Var(\hat{\phi}) \approx \left[\dfrac{1-\phi^2}{n}\right]\left[\dfrac{1-\phi\theta}{\phi-\theta}\right]^2 \\ Var(\hat{\theta}) \approx \left[\dfrac{1-\theta^2}{n}\right]\left[\dfrac{1-\phi\theta}{\phi-\theta}\right]^2 \\ Corr(\hat{\phi}, \hat{\theta}) \approx \dfrac{\sqrt{(1-\phi^2)(1-\theta^2)}}{1-\phi\theta} \end{cases} \tag{7.4.13}$$

Notice that, in the AR(1) case, the variance of the estimator of ϕ decreases as ϕ approaches ± 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 ϕ_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 ϕ and θ are nearly equal, the variability in the estimators of ϕ and θ 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.

The table shown in Exhibit 7.2 gives numerical values for the large-sample approximate standard deviations of the estimates of ϕ in an AR(1) model for several values of ϕ and several sample sizes. Since the values in the table are equal to $\sqrt{(1-\phi^2)/n}$, they apply equally well to standard deviations computed according to Equations (7.4.10), (7.4.11), and (7.4.12).

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 ϕ is in error by no more than $\pm 2(0.07) = \pm 0.14$.

Exhibit 7.2 AR(1) Model Large-Sample Standard Deviations of $\hat{\phi}$

ϕ	n = 50	100	200
0.4	0.13	0.09	0.06
0.7	0.10	0.07	0.05
0.9	0.06	0.04	0.03

For stationary autoregressive models, the method of moments yields estimators equivalent to least squares and maximum likelihood, at least for large samples. For models containing moving average terms, such is not the case. For an MA(1) model, it can be shown that the large-sample variance of the method-of-moments estimator of θ is equal to

$$Var(\hat{\theta}) \approx \frac{1 + \theta^2 + 4\theta^4 + \theta^6 + \theta^8}{n(1 - \theta^2)^2} \tag{7.4.14}$$

Comparing Equation (7.4.14) with that of Equation (7.4.11), we see that the variance for the method-of-moments estimator is always larger than the variance of the maximum likelihood estimator. The table in Exhibit 7.3 displays the ratio of the large-sample standard deviations for the two methods for several values of θ. For example, if θ is 0.5, the method-of-moments estimator has a large-sample standard deviation that is 42% larger than the standard deviation of the estimator obtained using maximum likelihood. It is clear from these ratios that the method-of-moments estimator should not be used for the MA(1) model. This same advice applies to all models that contain moving average terms.

Exhibit 7.3 Method of Moments (MM) vs. Maximum Likelihood (MLE) in MA(1) Models

θ	SD_{MM}/SD_{MLE}
0.25	1.07
0.50	1.42
0.75	2.66
0.90	5.33

7.5 Illustrations of Parameter Estimation

Consider the simulated MA(1) series with $\theta = -0.9$. The series was displayed in Exhibit 4.2 on page 59, and we found the method-of-moments estimate of θ to be a rather poor -0.554; see Exhibit 7.1 on page 152. In contrast, the maximum likelihood estimate is -0.915, the unconditional sum-of-squares estimate is -0.923, and the conditional least squares estimate is -0.879. For this series, the maximum likelihood estimate of -0.915 is closest to the true value used in the simulation. Using Equation (7.4.11) on page 161 and replacing θ by its estimate, we have a standard error of about

$$\sqrt{Va\hat{r}(\hat{\theta})} \approx \sqrt{\frac{1-\hat{\theta}^2}{n}} = \sqrt{\frac{1-(0.91)^2}{120}} \approx 0.04$$

so none of the maximum likelihood, conditional sum-of-squares, or unconditional sum-of-squares estimates are significantly far from the true value of -0.9.

The second MA(1) simulation with $\theta = 0.9$ produced the method-of-moments estimate of 0.719 shown in Exhibit 7.1. The conditional sum-of-squares estimate is 0.958, the unconditional sum-of-squares estimate is 0.983, and the maximum likelihood estimate is 1.000. These all have a standard error of about 0.04 as above. Here the maximum likelihood estimate of $\hat{\theta} = 1$ is a little disconcerting since it corresponds to a noninvertible model.

The third MA(1) simulation with $\theta = -0.9$ produced a method-of-moments estimate of -0.719 (see Exhibit 7.1). The maximum likelihood estimate here is -0.894 with a standard error of about

$$\sqrt{Va\hat{r}(\hat{\theta})} \approx \sqrt{\frac{1-(0.894)^2}{60}} \approx 0.06$$

For these data, the conditional sum-of-squares estimate is -0.979 and the unconditional sum-of-squares estimate is -0.961. Of course, with a standard error of this magnitude, it is unwise to report digits in the estimates of θ beyond the tenths place.

For our simulated autoregressive models, the results are reported in Exhibits 7.4 and 7.5.

Exhibit 7.4 Parameter Estimation for Simulated AR(1) Models

Parameter ϕ	Method-of-Moments Estimate	Conditional SS Estimate	Unconditional SS Estimate	Maximum Likelihood Estimate	n
0.9	0.831	0.857	0.911	0.892	60
0.4	0.470	0.473	0.473	0.465	60

```
> 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')
```

From Equation (7.4.9) on page 161, the standard errors for the estimates are

$$\sqrt{V\hat{a}r(\hat{\phi})} \approx \sqrt{\frac{1-\hat{\phi}^2}{n}} = \sqrt{\frac{1-(0.831)^2}{60}} \approx 0.07$$

and

$$\sqrt{V\hat{a}r(\hat{\phi})} = \sqrt{\frac{1-(0.470)^2}{60}} \approx 0.11$$

respectively. Considering the magnitude of these standard errors, all four methods estimate reasonably well for AR(1) models.

Exhibit 7.5 Parameter Estimation for a Simulated AR(2) Model

Parameters	Method-of-Moments Estimates	Conditional SS Estimates	Unconditional SS Estimates	Maximum Likelihood Estimate	*n*
$\phi_1 = 1.5$	1.472	1.5137	1.5183	1.5061	120
$\phi_2 = -0.75$	−0.767	−0.8050	−0.8093	−0.7965	120

```
> data(ar2.s)
> ar(ar2.s,order.max=2,AIC=F,method='yw')
> ar(ar2.s,order.max=2,AIC=F,method='ols')
> ar(ar2.s,order.max=2,AIC=F,method='mle')
```

From Equation (7.4.10) on page 161, the standard errors for the estimates are

$$\sqrt{V\hat{a}r(\hat{\phi}_1)} \approx \sqrt{V\hat{a}r(\hat{\phi}_2)} \approx \sqrt{\frac{1-\phi_2^2}{n}} = \sqrt{\frac{1-(0.75)^2}{120}} \approx 0.06$$

Again, considering the size of the standard errors, all four methods estimate reasonably well for AR(2) models.

As a final example using simulated data, consider the ARMA(1,1) shown in Exhibit 6.14 on page 123. Here $\phi = 0.6$, $\theta = -0.3$, and $n = 100$. Estimates using the various methods are shown in Exhibit 7.6.

Exhibit 7.6 Parameter Estimation for a Simulated ARMA(1,1) Model

Parameters	Method-of-Moments Estimates	Conditional SS Estimates	Unconditional SS Estimates	Maximum Likelihood Estimate	n
$\phi = 0.6$	0.637	0.5586	0.5691	0.5647	100
$\theta = -0.3$	−0.2066	−0.3669	−0.3618	−0.3557	100

```
> data(arma11.s)
> arima(arma11.s, order=c(1,0,1),method='CSS')
> arima(arma11.s, order=c(1,0,1),method='ML')
```

Now let's look at some real time series. The industrial chemical property time series was first shown in Exhibit 1.3 on page 3. The sample PACF displayed in Exhibit 6.26 on page 135, strongly suggested an AR(1) model for this series. Exhibit 7.7 shows the various estimates of the ϕ parameter using four different methods of estimation.

Exhibit 7.7 Parameter Estimation for the Color Property Series

Parameter	Method-of-Moments Estimate	Conditional SS Estimate	Unconditional SS Estimate	Maximum Likelihood Estimate	n
ϕ	0.5282	0.5549	0.5890	0.5703	35

```
> data(color)
> ar(color,order.max=1,AIC=F,method='yw')
> ar(color,order.max=1,AIC=F,method='ols')
> ar(color,order.max=1,AIC=F,method='mle')
```

Here the standard error of the estimates is about

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

so all of the estimates are comparable.

As a second example, consider again the Canadian hare abundance series. As before, we base all modeling on the square root of the original abundance numbers. Based on the partial autocorrelation function shown in Exhibit 6.29 on page 137, we will estimate an AR(3) model. For this illustration, we use maximum likelihood estimation and show the results obtained from the R software in Exhibit 7.8.

Exhibit 7.8 Maximum Likelihood Estimates from R Software: Hare Series

Coefficients:	ar1	ar2	ar3	Intercept†
	1.0519	–0.2292	–0.3931	5.6923
s.e.	0.1877	0.2942	0.1915	0.3371

```
sigma^2 estimated as 1.066: log-likelihood = -46.54, AIC = 101.08
```

† The intercept here is the estimate of the process mean μ—not of θ_0.

```
> data(hare)
> arima(sqrt(hare),order=c(3,0,0))
```

Here we see that $\hat{\phi}_1 = 1.0519$, $\hat{\phi}_2 = -0.2292$, and $\hat{\phi}_3 = -0.3930$. We also see that the estimated noise variance is $\hat{\sigma}_e^2 = 1.066$. Noting the standard errors, the estimates of the lag 1 and lag 3 autoregressive coefficients are significantly different from zero, as is the intercept term, but the lag 2 autoregressive parameter estimate is not significant.

The estimated model would be written

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

or

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

where Y_t is the hare abundance in year t in original terms. Since the lag 2 autoregressive term is insignificant, we might drop that term (that is, set $\phi_2 = 0$) and obtain new estimates of ϕ_1 and ϕ_3 with this subset model.

As a last example, we return to the oil price series. The sample ACF shown in Exhibit 6.32 on page 140, suggested an MA(1) model on the differences of the logs of the prices. Exhibit 7.9 gives the estimates of θ by the various methods and, as we have seen earlier, 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.

Exhibit 7.9 Estimation for the Difference of Logs of the Oil Price Series

Parameter	Method-of-Moments Estimate	Conditional SS Estimate	Unconditional SS Estimate	Maximum Likelihood Estimate	*n*
θ	–0.2225	–0.2731	–0.2954	–0.2956	241

```
> 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')
```

7.6 Bootstrapping ARIMA Models

In Section 7.4, we summarized some approximate normal distribution results for the estimator $\hat{\gamma}$, where γ is the vector consisting of all the ARMA parameters. These normal approximations are accurate for large samples, and statistical software generally uses those results in calculating and reporting standard errors. The standard error of some complex function of the model parameters, for example the quasi-period of the model, if it exists, is then usually obtained by the delta method. However, the general theory provides no practical guidance on how large the sample size should be for the normal approximation to be reliable. Bootstrap methods (Efron and Tibshirani, 1993; Davison and Hinkley, 2003) provide an alternative approach to assessing the uncertainty of an estimator and may be more accurate for small samples. There are several variants of the bootstrap method for dependent data—see Politis (2003). We shall confine our discussion to the parametric bootstrap that generates the bootstrap time series $Y_1^*, Y_2^*, \ldots, Y_n^*$ by simulation from the fitted ARIMA(p,d,q) model. (The bootstrap may be done by fixing the first $p + d$ initial values of Y^* to those of the observed data. For stationary models, an alternative procedure is to simulate stationary realizations from the fitted model, which can be done approximately by simulating a long time series from the fitted model and then deleting the transient initial segment of the simulated data—the so-called burn-in.) If the errors are assumed to be normally distributed, the errors may be drawn randomly and with replacement from $N(0, \hat{\sigma}_e^2)$. For the case of an unknown error distribution, the errors can be drawn randomly and with replacement from the residuals of the fitted model. For each bootstrap series, let $\hat{\gamma}^*$ be the estimator computed based on the bootstrap time series data using the method of full maximum likelihood estimation assuming stationarity. (Other estimation methods may be used.) The bootstrap is replicated, say, B times. (For example, $B = 1000$.) From the B bootstrap parameter estimates, we can form an empirical distribution and use it to calibrate the uncertainty in $\hat{\gamma}$. Suppose we are interested in estimating some function of γ, say $h(\gamma)$—for example, the AR(1) coefficient. Using the percentile method, a 95% bootstrap confidence interval for $h(\gamma)$ can be obtained as the interval from the 2.5 percentile to the 97.5 percentile of the bootstrap distribution of $h(\hat{\gamma}^*)$.

We illustrate the bootstrap method with the hare data. The bootstrap 95% confidence intervals reported in the first row of the table in Exhibit 7.10 are based on the bootstrap obtained by conditioning on the initial three observations and assuming normal errors. Those in the second row are obtained using the same method except that the errors are drawn from the residuals. The third and fourth rows report the confidence intervals based on the stationary bootstrap with a normal error distribution for the third row and the empirical residual distribution for the fourth row. The fifth row in the table shows the theoretical 95% confidence intervals based on the large-sample distribution results for the estimators. In particular, the bootstrap time series for the first bootstrap method is generated recursively using the equation

$$Y_t^* - \hat{\phi}_1 Y_{t-1}^* - \hat{\phi}_2 Y_{t-2}^* - \hat{\phi}_3 Y_{t-3}^* = \hat{\theta}_0 + e_t^* \tag{7.6.1}$$

for $t = 4, 5, \ldots, 31$, where the e_t^* are chosen independently from $N(0, \hat{\sigma}_e^2)$, $Y_1^* = Y_1$, $Y_2^* = Y_2$, $Y_3^* = Y_3$; and the parameters are set to be the estimates from the AR(3) model fitted to the (square root transformed) hare data with $\hat{\theta}_0 = \hat{\mu}(1 - \hat{\phi}_1 - \hat{\phi}_2 - \hat{\phi}_3)$. All results are based on about 1000 bootstrap replications, but full maximum likelihood estimation fails for 6.3%, 6.3%, 3.8%, and 4.8% of 1000 cases for the four bootstrap methods I, II, III, and IV, respectively.

Exhibit 7.10 Bootstrap and Theoretical Confidence Intervals for the AR(3) Model Fitted to the Hare Data

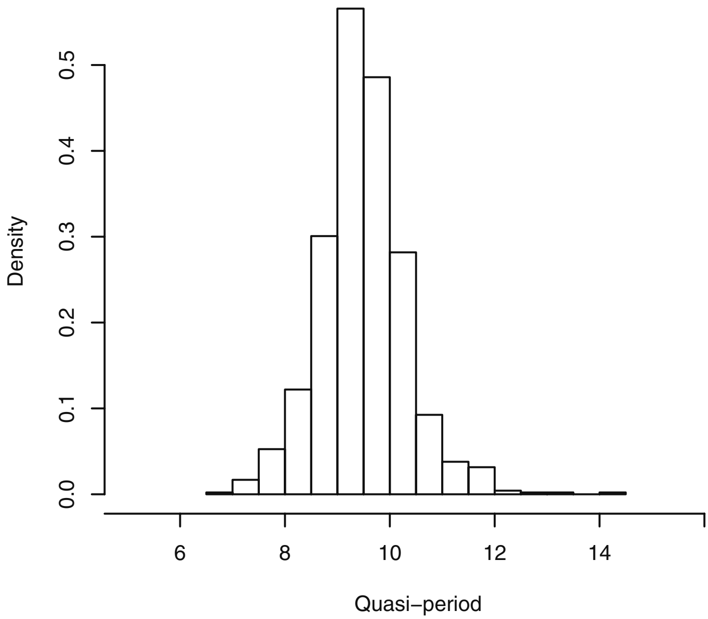
Method	ar1	ar2	ar3	intercept	noise var.
I	(0.593, 1.269)	(−0.655, 0.237)	(−0.666, −0.018)	(5.115, 6.394)	(0.551, 1.546)
II	(0.612, 1.296)	(−0.702, 0.243)	(−0.669, −0.026)	(5.004, 6.324)	(0.510, 1.510)
III	(0.699, 1.369)	(−0.746, 0.195)	(−0.666, −0.021)	(5.056, 6.379)	(0.499, 1.515)
IV	(0.674, 1.389)	(−0.769, 0.194)	(−0.665, −0.002)	(4.995, 6.312)	(0.477, 1.530)
Theoretical	(0.684, 1.42)	(−0.8058, 0.3474)	(−0.7684, −0.01776)	(5.032, 6.353)	(0.536, 1.597)

```
>   See the Chapter 7 R scripts file for the extensive code
    required to generate these results.
```

All four methods yield similar bootstrap confidence intervals, although the conditional bootstrap approach generally yields slightly narrower confidence intervals. This is expected, as the conditional bootstrap time series bear more resemblance to each other because all are subject to identical initial conditions. The bootstrap confidence intervals are generally slightly wider than their theoretical counterparts that are derived from the large-sample results. Overall, we can draw the inference that the ϕ_2 coefficient estimate is insignificant, whereas both the ϕ_1 and ϕ_3 coefficient estimates are significant at the 5% significance level.

The bootstrap method has the advantage of allowing easy construction of confidence intervals for a model characteristic that is a nonlinear function of the model parameters. For example, the characteristic AR polynomial of the fitted AR(3) model for the hare data admits a pair of complex roots. Indeed, the roots are $0.84 \pm 0.647i$ and -2.26, where $i = \sqrt{-1}$. The two complex roots can be written in polar form: $1.06\exp(\pm 0.657i)$. As in the discussion of the quasi-period for the AR(2) model on page 74, the quasi-period of the fitted AR(3) model can be defined as $2\pi/0.657 = 9.57$. Thus, the fitted model suggests that the hare abundance underwent cyclical fluctuation with a period of about 9.57 years. The interesting question of constructing a 95% confidence interval for the quasi-period could be studied using the delta method. However, this will be quite complex, as the quasi-period is a complicated function of the parameters. But the bootstrap provides a simple solution: For each set of bootstrap parameter estimates, we can compute the quasi-period and hence obtain the bootstrap distribution of the quasi-period. Confidence intervals for the quasi-period can then be constructed using the percentile method, and the shape of the distribution can be explored via the histogram of the bootstrap quasi-period estimates. (Note that the quasi-period will be undefined whenever the roots of the AR characteristic equation are all real numbers.) Among the 1000 stationary bootstrap time series obtained by simulating from the fitted model with the errors drawn randomly from the residuals with replacement, 952 series lead to successful full maximum likelihood estimation. All but one of the 952 series have well-defined quasi-periods, and the histogram of these is shown in Exhibit 7.11. The histogram shows that the sampling distribution of the quasi-period estimate is slightly skewed to the right.† The Q-Q normal plot (Exhibit 7.12) suggests that the quasi-period estimator has, furthermore, a thick-tailed distribution. Thus, the delta method and the corresponding normal distribution approximation may be inappropriate for approximating the sampling distribution of the quasi-period estimator. Finally, using the percentile method, a 95% confidence interval of the quasi-period is found to be (7.84,11.34).

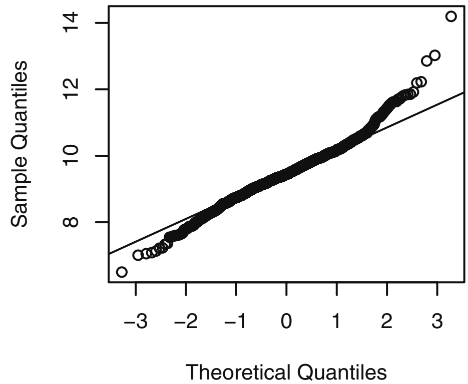
Exhibit 7.11 Histogram of Bootstrap Quasi-period Estimates




```
> win.graph(width=3.9,height=3.8,pointsize=8)
> hist(period.replace,prob=T,xlab='Quasi-period',axes=F,
    xlim=c(5,16))
> axis(2); axis(1,c(4,6,8,10,12,14,16),c(4,6,8,10,12,14,NA))
```

† However, see the discussion below Equation (13.5.9) on page 338 where it is argued that, from the perspective of frequency domain, there is a small parametric region corresponding to complex roots and yet the associated quasi-period may not be physically meaningful. This illustrates the subtlety of the concept of quasi-period.

Exhibit 7.12 Q-Q Normal Plot of Bootstrap Quasi-period Estimates




```
> win.graph(width=2.5,height=2.5,pointsize=8)
> qqnorm(period.replace); qqline(period.replace)
```

7.7 Summary

This chapter delved into the estimation of the parameters of ARIMA models. We considered estimation criteria based on the method of moments, various types of least squares, and maximizing the likelihood function. The properties of the various estimators were given, and the estimators were illustrated both with simulated and actual time series data. Bootstrapping with ARIMA models was also discussed and illustrated.

EXERCISES

7.1 From a series of length 100, we have computed $r_1 = 0.8$, $r_2 = 0.5$, $r_3 = 0.4$, $\overline{Y} = 2$, and a sample variance of 5. If we assume that an AR(2) model with a constant term is appropriate, how can we get (simple) estimates of ϕ_1, ϕ_2, θ_0, and σ_e^2 ?

7.2 Assuming that the following data arise from a stationary process, calculate method-of-moments estimates of μ, γ_0, and ρ_1: 6, 5, 4, 6, 4.

7.3 If $\{Y_t\}$ satisfies an AR(1) model with ϕ of about 0.7, how long of a series do we need to estimate $\phi = \rho_1$ with 95% confidence that our estimation error is no more than ± 0.1?

7.4 Consider an MA(1) process for which it is *known* that the process mean is zero. Based on a series of length $n = 3$, we observe $Y_1 = 0$, $Y_2 = -1$, and $Y_3 = ½$.

(a) Show that the conditional least-squares estimate of θ is ½.

(b) Find an estimate of the noise variance. (Hint: Iterative methods are not needed in this simple case.)

7.5 Given the data $Y_1 = 10$, $Y_2 = 9$, and $Y_3 = 9.5$, we wish to fit an IMA(1,1) model without a constant term.

(a) Find the conditional least squares estimate of θ. (Hint: Do Exercise 7.4 first.)

(b) Estimate σ_e^2.

7.6 Consider two different parameterizations of the AR(1) process with nonzero mean:

Model I. $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$.

Model II. $Y_t = \phi Y_{t-1} + \theta_0 + e_t$.

We want to estimate ϕ and μ or ϕ and θ_0 using conditional least squares conditional on Y_1. Show that with Model I we are led to solve nonlinear equations to obtain the estimates, while with Model II we need only solve linear equations.

7.7 Verify Equation (7.1.4) on page 150.

7.8 Consider an ARMA(1,1) model with $\phi = 0.5$ and $\theta = 0.45$.

(a) For $n = 48$, evaluate the variances and correlation of the maximum likelihood estimators of ϕ and θ using Equations (7.4.13) on page 161. Comment on the results.

(b) Repeat part (a) but now with $n = 120$. Comment on the new results.

7.9 Simulate an MA(1) series with $\theta = 0.8$ and $n = 48$.

(a) Find the method-of-moments estimate of θ.

(b) Find the conditional least squares estimate of θ and compare it with part (a).

(c) Find the maximum likelihood estimate of θ and compare it with parts (a) and (b).

(d) Repeat parts (a), (b), and (c) with a new simulated series using the same parameters and same sample size. Compare your results with your results from the first simulation.

7.10 Simulate an MA(1) series with $\theta = -0.6$ and $n = 36$.

(a) Find the method-of-moments estimate of θ.

(b) Find the conditional least squares estimate of θ and compare it with part (a).

(c) Find the maximum likelihood estimate of θ and compare it with parts (a) and (b).

(d) Repeat parts (a), (b), and (c) with a new simulated series using the same parameters and same sample size. Compare your results with your results from the first simulation.

7.11 Simulate an MA(1) series with $\theta = -0.6$ and $n = 48$.

(a) Find the maximum likelihood estimate of θ.

(b) If your software permits, repeat part (a) many times with a new simulated series using the same parameters and same sample size.

(c) Form the sampling distribution of the maximum likelihood estimates of θ.

(d) Are the estimates (approximately) unbiased?

(e) Calculate the variance of your sampling distribution and compare it with the large-sample result in Equation (7.4.11) on page 161.

7.12 Repeat Exercise 7.11 using a sample size of $n = 120$.

7.13 Simulate an AR(1) series with $\phi = 0.8$ and $n = 48$.
(a) Find the method-of-moments estimate of ϕ.
(b) Find the conditional least squares estimate of ϕ and compare it with part (a).
(c) Find the maximum likelihood estimate of ϕ and compare it with parts (a) and (b).
(d) Repeat parts (a), (b), and (c) with a new simulated series using the same parameters and same sample size. Compare your results with your results from the first simulation.

7.14 Simulate an AR(1) series with $\phi = -0.5$ and $n = 60$.
(a) Find the method-of-moments estimate of ϕ.
(b) Find the conditional least squares estimate of ϕ and compare it with part (a).
(c) Find the maximum likelihood estimate of ϕ and compare it with parts (a) and (b).
(d) Repeat parts (a), (b), and (c) with a new simulated series using the same parameters and same sample size. Compare your results with your results from the first simulation.

7.15 Simulate an AR(1) series with $\phi = 0.7$ and $n = 100$.
(a) Find the maximum likelihood estimate of ϕ.
(b) If your software permits, repeat part (a) many times with a new simulated series using the same parameters and same sample size.
(c) Form the sampling distribution of the maximum likelihood estimates of ϕ.
(d) Are the estimates (approximately) unbiased?
(e) Calculate the variance of your sampling distribution and compare it with the large-sample result in Equation (7.4.9) on page 161.

7.16 Simulate an AR(2) series with $\phi_1 = 0.6$, $\phi_2 = 0.3$, and $n = 60$.
(a) Find the method-of-moments estimates of ϕ_1 and ϕ_2.
(b) Find the conditional least squares estimates of ϕ_1 and ϕ_2 and compare them with part (a).
(c) Find the maximum likelihood estimates of ϕ_1 and ϕ_2 and compare them with parts (a) and (b).
(d) Repeat parts (a), (b), and (c) with a new simulated series using the same parameters and same sample size. Compare these results to your results from the first simulation.

7.17 Simulate an ARMA(1,1) series with $\phi = 0.7$, $\theta = 0.4$, and $n = 72$.
(a) Find the method-of-moments estimates of ϕ and θ.
(b) Find the conditional least squares estimates of ϕ and θ and compare them with part (a).
(c) Find the maximum likelihood estimates of ϕ and θ and compare them with parts (a) and (b).
(d) Repeat parts (a), (b), and (c) with a new simulated series using the same parameters and same sample size. Compare your new results with your results from the first simulation.

7.18 Simulate an AR(1) series with $\phi = 0.6$, $n = 36$ but with error terms from a *t*-distribution with 3 degrees of freedom.

(a) Display the sample PACF of the series. Is an AR(1) model suggested?
(b) Estimate ϕ from the series and comment on the results.
(c) Repeat parts (a) and (b) with a new simulated series under the same conditions.

7.19 Simulate an MA(1) series with $\theta = -0.8$, $n = 60$ but with error terms from a t-distribution with 4 degrees of freedom.
(a) Display the sample ACF of the series. Is an MA(1) model suggested?
(b) Estimate θ from the series and comment on the results.
(c) Repeat parts (a) and (b) with a new simulated series under the same conditions.

7.20 Simulate an AR(2) series with $\phi_1 = 1.0$, $\phi_2 = -0.6$, $n = 48$ but with error terms from a t-distribution with 5 degrees of freedom.
(a) Display the sample PACF of the series. Is an AR(2) model suggested?
(b) Estimate ϕ_1 and ϕ_2 from the series and comment on the results.
(c) Repeat parts (a) and (b) with a new simulated series under the same conditions.

7.21 Simulate an ARMA(1,1) series with $\phi = 0.7$, $\theta = -0.6$, $n = 48$ but with error terms from a t-distribution with 6 degrees of freedom.
(a) Display the sample EACF of the series. Is an ARMA(1,1) model suggested?
(b) Estimate ϕ and θ from the series and comment on the results.
(c) Repeat parts (a) and (b) with a new simulated series under the same conditions.

7.22 Simulate an AR(1) series with $\phi = 0.6$, $n = 36$ but with error terms from a chi-square distribution with 6 degrees of freedom.
(a) Display the sample PACF of the series. Is an AR(1) model suggested?
(b) Estimate ϕ from the series and comment on the results.
(c) Repeat parts (a) and (b) with a new simulated series under the same conditions.

7.23 Simulate an MA(1) series with $\theta = -0.8$, $n = 60$ but with error terms from a chi-square distribution with 7 degrees of freedom.
(a) Display the sample ACF of the series. Is an MA(1) model suggested?
(b) Estimate θ from the series and comment on the results.
(c) Repeat parts (a) and (b) with a new simulated series under the same conditions.

7.24 Simulate an AR(2) series with $\phi_1 = 1.0$, $\phi_2 = -0.6$, $n = 48$ but with error terms from a chi-square distribution with 8 degrees of freedom.
(a) Display the sample PACF of the series. Is an AR(2) model suggested?
(b) Estimate ϕ_1 and ϕ_2 from the series and comment on the results.
(c) Repeat parts (a) and (b) with a new simulated series under the same conditions.

7.25 Simulate an ARMA(1,1) series with $\phi = 0.7$, $\theta = -0.6$, $n = 48$ but with error terms from a chi-square distribution with 9 degrees of freedom.
(a) Display the sample EACF of the series. Is an ARMA(1,1) model suggested?
(b) Estimate ϕ and θ from the series and comment on the results.
(c) Repeat parts (a) and (b) with a new series under the same conditions.

7.26 Consider the AR(1) model specified for the color property time series displayed in Exhibit 1.3 on page 3. The data are in the file named color.

(a) Find the method-of-moments estimate of ϕ.

(b) Find the maximum likelihood estimate of ϕ and compare it with part (a).

7.27 Exhibit 6.31 on page 139 suggested specifying either an AR(1) or possibly an AR(4) model for the difference of the logarithms of the oil price series. The data are in the file named oil.price.

(a) Estimate both of these models using maximum likelihood and compare it with the results using the AIC criteria.

(b) Exhibit 6.32 on page 140 suggested specifying an MA(1) model for the difference of the logs. Estimate this model by maximum likelihood and compare to your results in part (a).

7.28 The data file named deere3 contains 57 consecutive values from a complex machine tool at Deere & Co. The values given are deviations from a target value in units of ten millionths of an inch. The process employs a control mechanism that resets some of the parameters of the machine tool depending on the magnitude of deviation from target of the last item produced.

(a) Estimate the parameters of an AR(1) model for this series.

(b) Estimate the parameters of an AR(2) model for this series and compare the results with those in part (a).

7.29 The data file named robot contains a time series obtained from an industrial robot. The robot was put through a sequence of maneuvers, and the distance from a desired ending point was recorded in inches. This was repeated 324 times to form the time series.

(a) Estimate the parameters of an AR(1) model for these data.

(b) Estimate the parameters of an IMA(1,1) model for these data.

(c) Compare the results from parts (a) and (b) in terms of AIC.

7.30 The data file named days contains accounting data from the Winegard Co. of Burlington, Iowa. The data are the number of days until Winegard receives payment for 130 consecutive orders from a particular distributor of Winegard products. (The name of the distributor must remain anonymous for confidentiality reasons.) The time series contains outliers that are quite obvious in the time series plot.

(a) Replace each of the unusual values with a value of 35 days, a much more typical value, and then estimate the parameters of an MA(2) model.

(b) Now assume an MA(5) model and estimate the parameters. Compare these results with those obtained in part (a).

7.31 Simulate a time series of length $n = 48$ from an AR(1) model with $\phi = 0.7$. Use that series as if it were real data. Now compare the theoretical asymptotic distribution of the estimator of ϕ with the distribution of the bootstrap estimator of ϕ.

7.32 The industrial color property time series was fitted quite well by an AR(1) model. However, the series is rather short, with $n = 35$. Compare the theoretical asymptotic distribution of the estimator of ϕ with the distribution of the bootstrap estimator of ϕ. The data are in the file named color.