Times Series Analysis – Parameter Estimation (I)

Jianyong Sun
School of Mathematics and Statistics
Xi'an Jiaotong University

19th Oct., 2017

Recap

- Model specification – how to choose appropriate values for p, d, and q for a given series
- Non-stationarity case – Augmented Dickey-Fuller Test
- Some real time sereis.

Parameter Estimation

- We assume that a model has already been specified
- 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 model
- In practice, we treat the dth difference of the original time series as the time series from which we estimate the parameters of the complete model.

The Methods of Moments

- It is 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.

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 ϕ can be estimated by

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

The Methods of Moments

Consider AR(2) case. The relationships between the parameters ϕ_1 and ϕ_2 and various moments are given by the Yule-Walker equations:

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

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

$$r_1 = \phi_1 + \rho_1\phi_2 \qquad \text{and} \qquad r_2 = \rho_1\phi_1 + \phi_2$$

which are then solved to obtain

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

The Methods of Moments

Yule-Walker Estimates:

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

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

The Durbin-Levinson recursion 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 Methods of Moments

Moving Average Models:
The method of moments is not nearly as convenient when applied to moving average models.

Consider the simple MA(1) case, we know that $\rho_1 = -\frac{\theta}{1+\theta^2}$. Now equating ρ_1 to r_1, we are solve a quadratic equation in θ.

If $|r_1| < 0.5$, then the two real roots are given by

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

The product of the two solutions is always equal to 1. Only one of the solutions satisfies the invertibility condition $|\theta| < 1$:

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

- 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.
- 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. Note that

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k-1} + \cdots + \theta_{q-k}\theta_q}{1+\theta_1^2+\cdots+\theta_q^2} & k = 1, 2, \cdots, q \\ 0 & k > q \end{cases}$$

The resulting equations are highly nonlinear in the θ's, however, and their solution would of necessity be numerical.

There will be multiple solutions, of which only one is invertible.

The Methods of Moments

<u>Mixed Models</u>:

We consider only the ARMA(1,1) case. Recall

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

for $k \geq 1$. Noting that $\rho_2/\rho_1 = \phi$, we can first estimate ϕ as

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

We can then use

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

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

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

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

The Methods of Moments

For the AR(p) models, note that

$$\gamma_0 = \frac{\sigma_e^2}{1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p}$$

we get

$$\hat{\sigma_e}^2 = (1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p)s^2$$

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

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_e^2,$$

so

$$\hat{\sigma_e}^2 = \frac{s^2}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2}$$

For the ARMA(1,1) process, we have:

$$\gamma_0 = \frac{(1 - 2\phi\theta + \theta^2)}{1 - \phi^2}\sigma_e^2$$

thus

$$\hat{\sigma_e}^2 = \frac{1 - \hat{\phi}^2}{1 - 2\hat{\phi}\hat{\theta} + \hat{\theta}^2}s^2$$

Numerical Examples

Model	True Parameters θ	ϕ_1	ϕ_2	Method-of-Moments Estimates 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.

Figure: Method-of-Moments Parameter Estimates for Simulated Series

The estimates for all the autoregressive models are fairly good but the estimates for the moving average models are not acceptable.

Numerical Examples

Consider the Canadian hare abundance series with a square root transformation. Consider an AR(2) model with the hare data. The two sample ACFs are $r_1 = 0.736$ and $r_2 = 0.304$. Using previous equations, the method-of-moments estimates of ϕ_1 and ϕ_2 are

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

$$\hat{\phi}_2 = \frac{r_2 - r_1^2}{1-r_1^2} = -0.519$$

The sample mean and variance of this series are found to be 5.82 and 5.88. Then the noise variance can be estimated as

$$\hat{\sigma_e}^2 = (1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2) s^2 = 1.97$$

Numerical Examples

The estimated model is then

$$\sqrt{Y}_t - 5.82 = 1.1178\left(\sqrt{Y_{t-1}} - 5.82\right) - 0.519\left(\sqrt{Y_{t-2}} - 5.82\right) + e_t$$

or

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

with estimated noise variance of 1.97.

Numerical Examples

Consider now the oil price series. It is 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

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

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

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

with estimated noise variance of

$$\hat{\sigma_e}^2 = \frac{s^2}{1 + \hat{\theta}^2} = 0.00686.$$

The standard error of the sample mean is 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

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

Least Squares Estimation

Because the method of moments is unsatisfactory for many models, we must consider other methods of estimation. First, the least squares.

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.

Least Squares Estimation

Autoregressive Models:

Consider the first-order case where:

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

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

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

Least Squares Estimation

Since only $Y_1, Y_2, \cdots, Y_n$ are observed, we can only sum from $t = 2$ to $t = n$:

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

which is usually called the **conditional sum-of-squares function**. 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 $Y_1, Y_2, \cdots, Y_n$.

Least Squares Estimation

Take derivatives of S_c w.r.t μ, and zeroing it, we obtain

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

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, we have:

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

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

Least Squares Estimation

Consider now the minimization of $S_c(\phi, \hat{Y})$ w.r.t. ϕ. We have

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

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.

Least Squares Estimation

For the general AR(p) process, the same methods can be used to obtain the same result, namely $\hat{\mu} = \bar{Y}$.

To generalize the estimation of the ϕ's, we consider the second-order model. If we replace μ by $\bar{Y}$ in the conditional sum-of-squares function, so

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

Take S_c w.r.t. ϕ_1, zeroing it, we have:

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

which we can rewrite as

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

Least Squares Estimation

The sum of the lagged products $\sum_{t=3}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})$ is very nearly the numerator of r_1, just one product $(Y_2 - \bar{Y})(Y_1 - \bar{Y})$ is missing.

A similar situation exists for $\sum_{t=3}^{n}(Y_{t-1} - \bar{Y})(Y_{t-2} - \bar{Y})$, but $(Y_n - \bar{Y})(Y_{n-1} - \bar{Y})$ is missing.

If we divide both sides of $\sum_{t=3}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})$ by $\sum_{t=3}^{n}(Y_t - \bar{Y})^2$, then except for end effects, which are negligible under the stationarity assumption, we obtain

$$r_1 = \phi_1 + r_1\phi_2$$

Least Squares Estimation

Take derivatives S_c w.r.t. ϕ_2, and zeroing it, analyze similarly as previous, we have:

$$r_2 = r_1\phi_1 + \phi_2$$

We obtain the sample Yule-Walker equations for an AR(2) model.

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.

Least Squares Estimation

Consider now the least-squares estimation of θ in the MA(1) model: $Y_t = e_t - \theta e_{t-1}$.

Note that invertible MA(1) models can be expressed as AR(∞)

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

Least Squares Estimation

So, least squares can be carried out by choosing a value θ that minimizes

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

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

Least Squares Estimation

- It is clear 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 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. Rewrite MA(1):

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

Using this equation, $e_1, \cdots, e_n$ can be calculated recursively if we have the initial value e_0.

A common approximation is to set $e_0 = 0$ (this is its expected value). Then, conditional on $e_0 = 0$, we have:

Least Squares Estimation

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

Thus calculate $S_c(\theta)$ conditional on $e_0 = 0$ for that single given value of θ.

Least Squares Estimation

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 MA models, the ideas are the same. We compute $e_t = e_t(\theta_1, \cdots, \theta_q)$ recursively from

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

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

<u>Mixed Models:</u>

Consider the ARMA(1,1) model:

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

As in the pure MA case, we wish to minimize $S_c(\theta) = \sum e_t^2$. We can rewrite ARMA(1,1) as

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

To obtain e_1, we have the so-called 'startup' problem, namely Y_0.

Least Squares Estimation

One approach is to set $Y_0 = 0$ or to $\bar{Y}$ if our model contains a nonzero mean. 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), we compute

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

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

Least Squares Estimation

For parameters $\theta_1, \cdots, \theta_q$ corresponding to 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.

For series of moderate length and also for stochastic seasonal models, the start-up values will have a more pronounced effect on the final estimates for the parameters.

Maximum Likelihood and Unconditional Least Squares

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, $Y_1, Y_2, \cdots, Y_n$, time series or not, the likelihood function L is defined to be the joint probability density of obtaining the data actually observed.

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 $Y_1, Y_2, \cdots, Y_n$

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.

First, let's start with 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:

$$p(e_t) = (2\pi\sigma_e^2)^{-1/2} \exp\left\{-\frac{e_t^2}{2\sigma_e^2}\right\}$$

By independence, the joint pdf for $e_2, e_3, \cdots, e_n$ is

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

Now consider

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

If we condition on $Y_1 = y_1$, this equation defines a linear transformation between $e_2, e_3, \cdots, e_n$ and $Y_2, Y_3, \cdots, Y_n$.

Maximum Likelihood and Unconditional Least Squares

Thus, we get

$$f(y_2, y_3, \cdots, y_n|y_1) = (2\pi\sigma_e^2)^{-(n-1)/2} \times \exp\left\{-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n}[(y_t-\mu)-\phi(y_{t-1}-\mu)]^2\right\}$$

Now consider the marginal distribution of Y_1.

It follows from the linear process representation of the AR(1) process that Y_1 will have a normal distribution with mean μ and variance $\sigma_e^2/(1-\phi^2)$ because

$$Y_t = \mu + e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \phi^3 e_{t-3} + \cdots$$

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 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

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

where

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

which is called the **unconditional sum-of-squares function**.

Maximum Likelihood and Unconditional Least Squares

For AR(1), the **log-likelihood function**, denoted $\ell(\phi, \mu, \sigma_e^2)$ is given by

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

For a given ϕ and μ, $\ell(\phi, \mu, \sigma_e^2)$ can be maximized analytically w.r.t. σ_e^2 in terms of ϕ and μ. We obtain

$$\sigma_e^2 = \frac{S(\hat{\phi}, \hat{\mu})}{n}$$

Usually divide by $n-2$ rather than n (since we are estimating two parameters, ϕ and μ) to obtain an estimator with less bias.

Maximum Likelihood and Unconditional Least Squares

Now estimate ϕ and μ. A comparison of the unconditional sum-of-squares function $S(\phi, \mu)$ with the earlier conditional sum-of-squares function $S_c(\phi, \mu)$ reveals one simple difference

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

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.

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**.

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.

Properties of the Estimates

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

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

$$\text{AR}(2): \quad Var(\hat{\phi}_1) \approx Var(\hat{\phi}_2) \approx \frac{1-\phi_2^2}{n}$$

$$Corr(\hat{\phi}_1, \hat{\phi}_2) \approx \frac{\phi_1}{1-\phi_2} = -\rho_1$$

Properties of the Estimates

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

$$\text{MA}(2): \quad Var(\hat{\theta}_1) \approx Var(\hat{\theta}_2) \approx \frac{1-\theta_2^2}{n}$$

$$Corr(\hat{\theta}_1, \hat{\theta}_2) \approx -\frac{\theta_1}{1-\theta_2}$$

$$\text{ARMA}(1,1): \quad Var(\hat{\phi}) \approx \left[\frac{1-\phi^2}{n}\right]\left[\frac{1-\phi\theta}{\phi-\theta}\right]^2$$

$$Var(\hat{\theta}) \approx \left[\frac{1-\theta^2}{n}\right]\left[\frac{1-\phi\theta}{\phi-\theta}\right]^2$$

$$Corr(\hat{\phi}, \hat{\theta}) \approx \frac{\sqrt{(1-\phi^2)(1-\theta^2)}}{1-\phi\theta}$$

Properties of the Estimates

- Notice that in AR(1), the variance of the estimator of ϕ decreases as ϕ approaches ± 1.
- Notice that even though an AR(1) model is a special case of an AR(2) model, the variance of $\hat{\phi}_1$ 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. If ϕ and θ are nearly equal, the variability in the estimators of ϕ and θ can be extremely large.
- In all of the two-parameter models, the estimates can be highly correlated, even for very large sample sizes.

Properties of the Estimates

	n		
ϕ	**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

Figure: AR(1) Model Large-Sample Standard Deviations of $\hat{\phi}$.

This table 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.

Properties of the Estimates

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

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

It is seen that the variance for the method-of-moments estimator is always larger than the variance of the maximum likelihood estimator.

Properties of the Estimates

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

Figure: Method of Moments (MM) vs. Maximum Likelihood (MLE) in MA(1) Models

This figure displays the ratio of the large-sample standard deviations for the two methods for several values of θ.

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.

Questions?