Unit 13: ARMA Estimation

Taylor Brown

Department of Statistics, University of Virginia

Fall 2020

Readings for Unit 13

Textbook chapter 3.5 (pages 115 to 121).

Last Unit

1. ARMA forecasting.
2. Prediction error.
3. Prediction interval.

This Unit

1. Method of Moments Estimation.
2. Maximum Likelihood Estimation.

Motivation

In this unit, we explore a couple of ways to estimate the parameters for ARMA models: Method of Moments (MOM) estimation and Maximum Likelihood (ML) estimation.

1. Method of Moments Estimation

2. Maximum Likelihood Estimation

3. Comparison of MOM and MLE

Method of Moments

Let's start with the method of moments (MOM) estimation. The idea behind this is to equate population moments to sample moments and then solve for the parameters in terms of the sample moments.

We re-use a lot of the same equations from the previous section!

AR Estimation

Let's first assume that we have a causal AR(p) model

$$\phi(B)(x_t - \mu) = w_t,$$

where the white noise w_t has variance σ_w^2,

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p.$$

Given n observations $x_1, x_2, \ldots, x_n$, we are interested in estimating the parameters $\phi_1, \ldots, \phi_p$ and σ_w^2. Initially we assume that the order p is known.

AR Estimation

We'll assume again without loss of generality (WLOG) that $\mu = 0$. Why?

$E[x_t] = \mu$ can always be estimated with the first sample moment $\bar{x}$.

If $\mu \neq 0$, then transform your data before estimating as follows: $\tilde{x}_t = x_t - \bar{x}$.

Yule-Walker Estimation for AR(p)

The method of moments works well when estimating causal AR(p) models. We consider the causal AR(p) model

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t. \tag{1}$$

For $h = 1, \ldots, p$, multiply both sides of (1) by x_{t-h}, and take expectations:

$$\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p) \tag{2}$$

When $h = 0$, we do the same thing and get

$$\sigma_w^2 = \gamma(0) - \phi_1 \gamma(1) + \cdots + \phi_p \gamma(p) \tag{3}$$

Yule-Walker Estimation for AR(p)

We call these the **Yule-Walker equations**

$$\begin{aligned} \gamma(h) &= \phi_1\gamma(h-1) + \phi_2\gamma(h-2) + \cdots + \phi_p\gamma(h-p), \qquad h = 1, \ldots, p \\ \sigma_w^2 &= \gamma(0) - \phi_1\gamma(1) - \phi_2\gamma(2) - \cdots - \phi_p\gamma(p). \end{aligned}$$

We can also write them in matrix notation that should look familiar:

$$\Gamma_p\phi = \gamma_p \qquad \sigma_w^2 = \gamma(0) - \phi'\gamma_p$$

where $\Gamma_p = \{\gamma(k-j)\}_{j,k=1}^{p}$ is a $p \times p$ matrix, $\phi = (\phi_1, \cdots, \phi_p)'$ is a $p \times 1$ vector, and $\gamma_p = (\gamma(1), \cdots, \gamma(p))'$ is a $p \times 1$ vector.

Yule-Walker Estimation for AR(p)

Using method of moments, put hat signs on everything, and then solve for the desired parameters:

$$\Gamma_p \phi = \gamma_p \qquad \sigma_w^2 = \gamma(0) - \phi' \gamma_p$$

yields

$$\hat{\phi} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p$$

and

$$\hat{\sigma}_w^2 = \hat{\gamma}(0) - \hat{\gamma}_p' \hat{\Gamma}_p^{-1} \hat{\gamma}_p.$$

Yule-Walker Estimation for AR(p)

One more small move: divide through by $\hat{\gamma}(0)$ before solving, so that we have formulas in terms of ACFs.

$$\frac{1}{\gamma(0)}\Gamma_p\phi = \frac{1}{\gamma(0)}\gamma_p \qquad \sigma_w^2 = \gamma(0) - \phi'\gamma_p$$

gives us

$$\hat{\phi} = \hat{R}_p^{-1}\hat{\rho}_p \tag{4}$$

and

$$\hat{\sigma}_w^2 \quad = \quad \hat{\gamma}(0)\left[1 - \hat{\rho}_p'\hat{R}_p^{-1}\hat{\rho}_p\right]$$

where $\hat{R}_p = \{\hat{\rho}(k-j)\}_{j,k=1}^{p}$ is a $p \times p$ matrix and $\hat{\rho}_p = (\hat{\rho}(1), \cdots, \hat{\rho}(p))'$ is a $p \times 1$ vector.

Yule-Walker Estimation for AR(p)

The asymptotic behavior of the Yule-Walker estimators for causal AR(p) processes is

$$\sqrt{n}\left(\hat{\phi}-\phi\right) \xrightarrow{d} N\left(0, \sigma_w^2 \mathbf{\Gamma}_p^{-1}\right) \tag{5}$$

and

$$\hat{\sigma}_w^2 \xrightarrow{p} \sigma_w^2.$$

Yule-Walker Estimation for AR(p)

The variance-covariance matrix for $\hat{\phi}$ is

$$\begin{aligned} Var(\hat{\phi}) &= \frac{\sigma^2}{n}\mathbf{\Gamma}_p^{-1} \\ &= \frac{\sigma^2}{n\hat{\gamma}(0)}\boldsymbol{R}_p^{-1} \end{aligned} \tag{6}$$

Simulation Example

Using the `armasim()` function, I simulated $n = 1000$ observations from the following AR(2) process

$$x_t = 1.5x_{t-1} - 0.75x_{t-2} + w_t$$

where $\sigma_w^2 = 1$. For the sample, $\hat{\gamma}(0) = 7.69697$, $\hat{\rho}(1) = 0.8456375$, and $\hat{\rho}(2) = 0.5054795$.

Simulation Example

The data had $\hat{\rho}(1) = .849$, $\hat{\rho}(2) = .519$ and $\hat{\gamma}(0) = 8.903$

$$\begin{aligned}
\hat{\phi} &= \hat{R}_p^{-1} \hat{\rho}_p \\
&= \begin{bmatrix} 1 & .849 \\ .849 & 1 \end{bmatrix}^{-1} \begin{bmatrix} .849 \\ .519 \end{bmatrix} \\
&= \begin{bmatrix} 1.463 \\ -.723 \end{bmatrix}
\end{aligned}$$

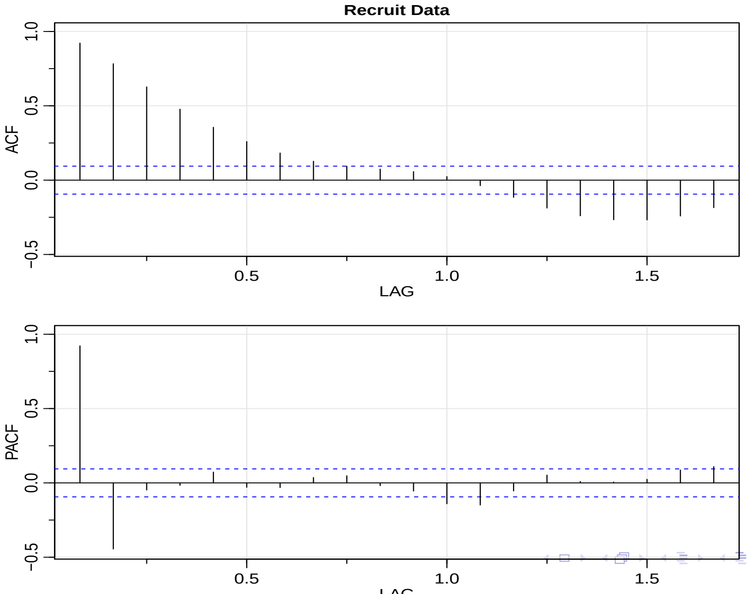
Also,

$$\begin{aligned}
\hat{\sigma}_w^2 &= \hat{\gamma}(0) \left[1 - \hat{\rho}_p' \hat{R}_p^{-1} \hat{\rho}_p \right] \\
&= \hat{\gamma}(0) \left[1 - \hat{\rho}_p' \hat{\phi} \right] \\
&= 8.903 * (1 - \begin{bmatrix} .849 & .519 \end{bmatrix} \begin{bmatrix} 1.463 \\ -.723 \end{bmatrix}) = 1.187
\end{aligned}$$

Fish Population Example

In Unit 11, we looked at the ACF and PACF of the time series from "recruit.dat", which contains data on fish population in the central Pacific Ocean. The numbers represent the number of new fish for each month in the years 1950-1987.

Fish Population Example

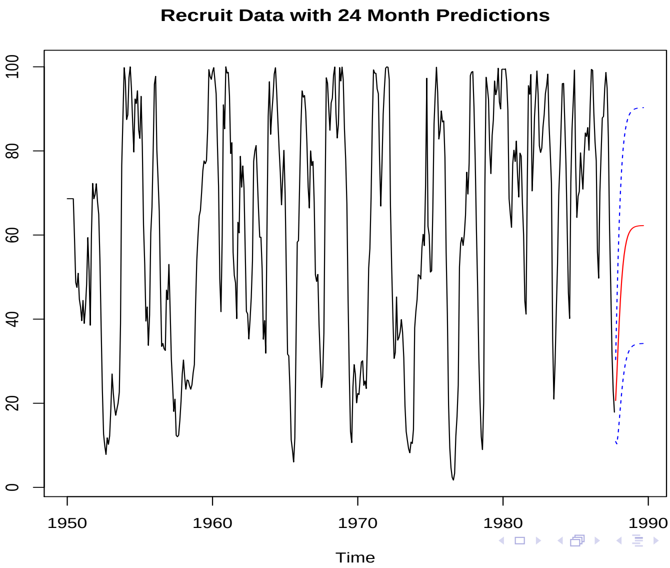


Fish Population Example

Let's check the results of fitting an AR(2) model using Yule-Walker estimation in R.

```
> rec.yw<-ar.yw(rec, order=2)
> rec.yw$x.mean
[1] 62.26278
> rec.yw$ar
[1]  1.3315874 -0.4445447
> sqrt(diag(rec.yw$asy.var.coef))
[1] 0.04222637 0.04222637
```

Fish Population Example



Fish Population Example

```
rec.pred <- predict(rec.yw, n.ahead=24)
ts.plot(rec, rec.pred$pred, col=1:2)
lines(rec.pred$pred - rec.pred$s, col=4, lty=2)
lines(rec.pred$pred + rec.pred$s, col=4, lty=2)
```

Method of Moments Estimation for MA(q)

Consider an invertible MA(1) process $x_t = w_t + \theta w_{t-1}$, with $|\theta| < 1$. We know that

$$\rho(1) = \frac{\theta}{1 + \theta^2}.$$

Using method of moments, we equate $\hat{\rho}(1)$ to $\rho(1)$ and solve a quadratic equation in θ.

Method of Moments Estimation for MA(q)

The invertible solution(s) is(are)

$$\hat{\theta} = \frac{1 \pm \sqrt{1 - 4\hat{\rho}(1)^2}}{2\hat{\rho}(1)}.$$

- If $|\hat{\rho}(1)| < 0.5$, two solutions exist, so we pick the invertible one.
- If $\hat{\rho}(1) = \pm 0.5$, $\hat{\theta} = \pm 1$. No invertible solution.
- If $\hat{\rho}(1) > 0.5$, no real solutions exist: the method of moments fails to yield an estimator of θ.

Method of Moments Estimation for MA(q)

For higher order MA(q) models, the method of moments quickly gets complicated. The equations are non-linear in $\theta_1, \cdots, \theta_q$, so numerical methods need to be used.

1. Method of Moments Estimation
2. Maximum Likelihood Estimation
3. Comparison of MOM and MLE

Maximum Likelihood Estimation

To illustrate the main concept with maximum likelihood estimation, we consider the AR(1) model with nonzero mean

$$x_t = \mu + \phi(x_{t-1} - \mu) + w_t, \tag{7}$$

where $|\phi| < 1$ and $w_t \sim iid\ \mathrm{N}(0,\sigma_w^2)$.

Maximum Likelihood Estimation

We seek the likelihood

$$L(\mu, \phi, \sigma_w^2) = f_{\mu,\phi,\sigma_w^2}(x_1, x_2, \ldots, x_n). \tag{8}$$

The likelihood function (8) $L(\mu, \phi, \sigma_w^2)$ is functionally equivalent to the joint probability distribution of the observed data $x_1, x_2, \ldots, x_n$.

Maximum Likelihood Estimation

For a given data set, think of the likelihood as a function of the parameters (not the data). Since we've already observed the data $x_1, x_2, \ldots, x_n$, we can find parameters (μ, ϕ, σ_w^2) to maximize the likelihood $L(\mu, \phi, \sigma_w^2)$. This is the basic idea behind maximum likelihood estimation.

Likelihood Function

We will use the following:

$$
\begin{aligned}
L(\mu, \phi, \sigma_w^2) &= f(x_1, \ldots, x_t) \\
&= f(x_1)f(x_2|x_1)f(x_3|x_2, x_1) \cdots f(x_t|x_{t-1}, x_{t-2}, \ldots, x_1) \\
&= f(x_1)f(x_2|x_1)f(x_3|x_2) \cdots f(x_t|x_{t-1})
\end{aligned}
$$

Likelihood Function

These are all the same:

$$x_t = \mu + \phi(x_{t-1} - \mu) + w_t \qquad w_t \sim \text{Normal}(0, \sigma_w^2)$$

$$x_t | x_{t-1} \sim \text{Normal}(\mu + \phi(x_{t-1} - \mu), \sigma_w^2)$$

and

$$f_{x_t|x_{t-1}}(x_t|x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\Big\{ - \frac{[x_t - \mu - \phi(x_{t-1} - \mu)]^2}{2\sigma_w^2} \Big\}.$$

Likelihood Function

We have

$$L(\mu, \phi, \sigma_w^2) = f_{x_1}(x_1) \times f_{x_2|x_1}(x_2|x_1) \times \cdots \times f_{x_n|x_{n-1}}(x_n|x_{n-1})$$

$$= f_{x_1}(x_1)(2\pi\sigma_w^2)^{-(n-1)/2} \times \exp\left\{ -\frac{\sum_{t=2}^{n}[x_t - \mu - \phi(x_{t-1} - \mu)]^2}{2\sigma_w^2} \right\}.$$

Likelihood Function: what is $f_{x_1}(x_1)$?

In midterm 1, we assumed

$$x_1 \sim \text{Normal}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right)$$

because that would allow all other time points to have the same marginal distribution.

Here's another rationalization: assume this model is causal ($|\phi| < 1$), and pretend you have an infinite history of data (impossible in practice).

The causal representation $x_1 = \mu + \sum_{j=0}^{\infty} \phi^j w_{1-j}$ is. Take expectations on both sides, and take the variance on both sides.

Since w_t are iid normal, x_1 is a normal with mean μ and variance $\sigma_w^2/(1-\phi^2)$.

Likelihood Function

The likelihood function is

$$
\begin{aligned}
L(\mu, \phi, \sigma_w^2) &= f_{x_1}(x_1)(2\pi\sigma_w^2)^{-(n-1)/2} \exp\left\{ - \frac{\sum_{t=2}^{n}[x_t - \mu - \phi(x_{t-1} - \mu)}{2\sigma_w^2} \right. \\
&= (2\pi\sigma_w^2)^{-n/2}(1-\phi^2)^{1/2} \exp\left\{ - \frac{S(\mu, \phi)}{2\sigma_w^2} \right\},
\end{aligned}
$$

where

$$
S(\mu, \phi) = (1-\phi^2)(x_1 - \mu)^2 \sum_{t=2}^{n}[x_t - \mu - \phi(x_{t-1} - \mu)]^2.
$$

The Log-likelihood Function

It is worth pointing out that it is more common to consider the log-likelihood

$$
\begin{aligned}
\ell(\mu, \phi, \sigma^2) &= \log L(\mu, \phi, \sigma^2) \\
&= \log \left[(2\pi\sigma_w^2)^{-n/2} (1 - \phi^2)^{1/2} \exp\left\{ -\frac{S(\mu, \phi)}{2\sigma_w^2} \right\} \right] \\
&= -\frac{n}{2} \log(2\pi\sigma_w^2) + \frac{1}{2} \log(1 - \phi^2) - \frac{S(\mu, \phi)}{2\sigma_w^2}
\end{aligned}
$$

Numerically more stable, and the derivatives are easier to calculate.

The variance estimator

The variance estimator can be obtained after you get the other estimators:

$$\frac{d}{d\sigma^2}L(\mu,\phi,\sigma_w^2) = \frac{d}{d\sigma^2}\left[-\frac{n}{2}\log(2\pi\sigma_w^2) + \frac{1}{2}\log(1-\phi^2) - \frac{S(\mu,\phi)}{2\sigma_w^2}\right]$$
$$= -\frac{n}{2\sigma_w^2} + S(\mu,\phi)\left(\frac{1}{\sigma_w^2}\right)^2$$

Setting that equal to 0, replacing σ_w^2 with $\hat{\sigma}_w^2$, and solving for $\hat{\sigma}_w^2$ gives us

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

The variance estimator

The estimates for μ and σ^2 is more complicated. Taking the derivative with respect to these, and setting the equations equal to 0 yields something that cannot be solved analytically. It's usually accomplished with a numerical procedure (e.g. Newton-Raphson or Fisher scoring).

1. Method of Moments Estimation

2. Maximum Likelihood Estimation

3. Comparison of MOM and MLE

Properties of ML Estimators

Property 3.10 Large Sample Distribution of the Estimators

Under appropriate conditions, for causal and invertible ARMA processes, the maximum likelihood, the unconditional least squares, and the conditional least squares estimators, each initialized by the method of moments estimator, all provide optimal estimators of σ_w^2 and β, in the sense that $\hat{\sigma}_w^2$ is consistent, and the asymptotic distribution of $\hat{\beta}$ is the best asymptotic normal distribution. In particular, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma_w^2\, \Gamma_{p,q}^{-1}\right). \tag{3.132}$$

The asymptotic variance–covariance matrix of the estimator $\hat{\beta}$ is the inverse of the information *matrix. In particular, the $(p+q) \times (p+q)$ matrix $\Gamma_{p,q}$, has the form*

$$\Gamma_{p,q} = \begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta} \end{pmatrix}. \tag{3.133}$$

The $p \times p$ matrix $\Gamma_{\phi\phi}$ is given by (3.100), that is, the ij-th element of $\Gamma_{\phi\phi}$, for $i, j = 1, \ldots, p$, is $\gamma_x(i-j)$ from an AR(p) process, $\phi(B)x_t = w_t$. Similarly, $\Gamma_{\theta\theta}$ is a $q \times q$ matrix with the ij-th element, for $i, j = 1, \ldots, q$, equal to $\gamma_y(i-j)$ from an AR(q) process, $\theta(B)y_t = w_t$. The $p \times q$ matrix $\Gamma_{\phi\theta} = \{\gamma_{xy}(i-j)\}$, for $i = 1, \ldots, p$; $j = 1, \ldots, q$; that is, the ij-th element is the cross-covariance between the two AR processes given by $\phi(B)x_t = w_t$ and $\theta(B)y_t = w_t$. Finally, $\Gamma_{\theta\phi} = \Gamma'_{\phi\theta}$ is $q \times p$.