Parameter Estimation (Theory)

Al Nosedal
University of Toronto

April 15, 2019

METHOD OF MOMENTS

The Method of Moments

The method of moments is frequently one of the easiest. The method consists of equating sample moments to corresponding theoretical moments and solving the resulting equations to obtain estimates of any unknown parameters.

Example. Autoregressive Models

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.

$$\begin{aligned} \rho(1) &= \phi_1 + \phi_2 \rho(1) \\ \rho(2) &= \phi_1 \rho(1) + \phi_2 \end{aligned}$$

Example. Autoregressive Models (cont.)

The method of moments replaces $\rho(1)$ by r_1 and $\rho(2)$ by r_2 to obtain

$$\begin{aligned} r_1 &= \phi_1 + r_1\phi_2 \\ r_2 &= \phi_1 r_1 + \phi_2 \end{aligned}$$

Example. Autoregressive Models (cont.)

Solving for ϕ_1 and ϕ_2 yields

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

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

Example. ARMA(1,1)

Consider the ARMA(1,1) case.
$Y_t = \phi Y_{t-1} - \theta W_{t-1} + W_t$

$$\begin{aligned}
\gamma(0) &= \frac{(1 - 2\phi\theta + \theta^2)\sigma_W^2}{1 - \phi^2} \\
\gamma(1) &= \phi\gamma(0) - \theta\sigma_W^2 \\
\gamma(2) &= \phi\gamma(1)
\end{aligned}$$

Example. ARMA(1,1) (cont.)

$$\begin{aligned}\rho(0) &= 1 \\ \rho(1) &= \frac{(\phi-\theta)(1-\phi\theta)}{1-2\phi\theta+\theta^2} \\ \rho(2) &= \phi\rho(1)\end{aligned}$$

Example. ARMA(1,1) (cont.)

Replacing $\rho(1)$ by r_1, $\rho(2)$ by r_2, and solving for ϕ yields

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

Example. ARMA(1,1) (cont.)

To obtain $\hat{\theta}$, the following quadratic equation must be solved (and only the invertible solution retained)

$$\theta^2(r_1 - \hat{\phi}) + \theta(1 - 2r_1\hat{\phi} + \phi^2) + (r_1 - \hat{\phi}) = 0$$

Estimates of the Noise Variance

The final parameter to be estimated is the noise variance, σ_W^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 $\gamma(0)$.

Example. AR(2) process

For an AR(2) process,

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma_W^2$$

Dividing by $\gamma(0)$ and solving for σ_W^2 yields

$$\hat{\sigma}_W^2 = s^2[1 - \hat{\phi} r_1 - \hat{\phi}_2 r_2]$$

LEAST SQUARES ESTIMATION

Autoregressive Models

Consider the first-order case where

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

Note that

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

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

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

This is usually called the conditional sum-of-squares function.

Now, consider the equation $\frac{\partial S_C}{\partial \mu} = 0$

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

Solving it for μ yields

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

Now, for large n,

$$\hat{\mu} \approx \bar{Y}$$

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

$$\frac{\partial S_C(\phi, \bar{Y})}{\partial \phi} \sum_{t=2}^{n} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y}) - \phi \sum_{t=2}^{n} (Y_{t-1} - \bar{Y})^2 = 0$$

Solving for ϕ gives

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

Except for one term missing in the denominator, this is the same as r_1.

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.

Moving Average Models

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

$$Y_t = W_t - W_{t-1}$$

Recall that an invertible MA(1) can be expressed as

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

an autoregressive model but of infinite order.

In this case,

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

It is clear from this equation that the least squares problem is nonlinear in the parameters.

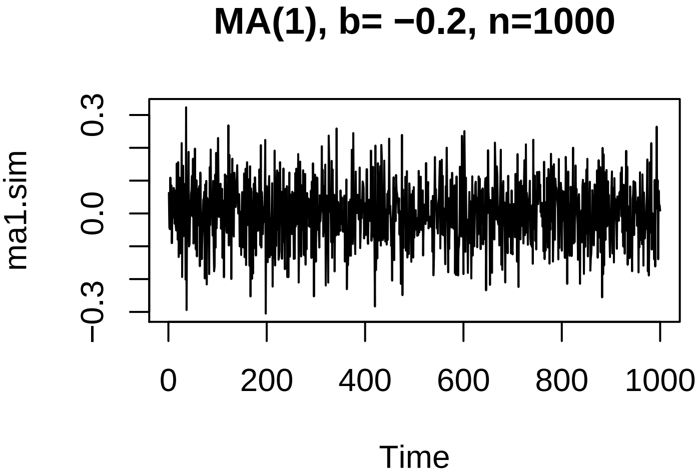
We will not be able to minimize $S_C(\theta)$ by taking a derivative with respect to θ, setting it to zero, and solving. We must resort to techniques of numerical optimization.

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 will be needed.

Example. MA(1) simulation

```
set.seed(19)

# simulating MA(1);
ma1.sim<-arima.sim(list(ma = c( -0.2)),
n = 1000, sd=0.1);
```

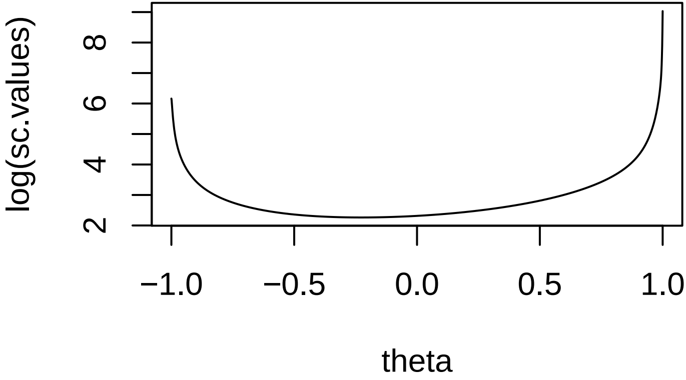
Example. MA(1) time series plot



```
## SC = conditional sum of squares function for MA(1)

SC<-function(y,theta){
        n<-length(y);
        error<-numeric(n);
        error[1]<-y[1];

        for (i in 2:n){
                error[i]<-y[i]-theta*error[i-1]
                          }

        SSE<-sum(error^2)

        return(SSE)
        }
```

```
SC.VEC<-function(y,theta){

                    N<-length(theta);
                    SSE.VEC<-numeric(N);

                    for (j in 1:N){

                         SSE.VEC[j]<-SC(y,theta[j])

                                       }

                     return(SSE.VEC)
}
```

```
## Testing function SC

theta=seq(-1,1, by=0.001)

sc.values=SC.VEC(ma1.sim,theta);

plot(theta,log(sc.values),type="l")

index=seq(1,length(sc.values),by=1);

theta.hat.one=theta[index[sc.values==min(sc.values)]];

theta.hat.one
```




```
## [1] -0.228
```

EXACT LIKELIHOOD FUNCTIONS

To illustrate the derivation of the exact likelihood function for a time series model, consider the AR(1) process

$$(1 - \phi B)\dot{Y}_t = W_t$$

or

$$\dot{Y}_t = \phi \dot{Y}_{t-1} + W_t$$

where $\dot{Y}_t = Y_t - \mu$, $|\phi| < 1$ and the W_t are iid $N(0, \sigma_W^2)$.

Rewriting the process in the moving average representation, we have

$$\dot{Y}_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$$

$$E[\dot{Y}_t] = 0$$

$$V[\dot{Y}_t] = \frac{\sigma_W^2}{1 - \phi^2}$$

Clearly, $\dot{Y}_t$ will be distributed as $N\left(0, \frac{\sigma_W^2}{1-\phi^2}\right)$.

To derive the joint pdf of $(\dot{Y}_1, \dot{Y}_2, \cdots, \dot{Y}_n)$ and hence the likelihood function for the parameters, we consider

$$
\begin{aligned}
e_1 &= \sum_{j=0}^{\infty} \phi^j W_{1-j} = \dot{Y}_1 \\
W_2 &= \dot{Y}_2 - \phi \dot{Y}_1 \\
W_3 &= \dot{Y}_3 - \phi \dot{Y}_2 \\
&\vdots \\
W_n &= \dot{Y}_n - \phi \dot{Y}_{n-1}
\end{aligned}
$$

Note that e_1 follows the Normal distribution $N\left(0, \frac{\sigma_W^2}{1-\phi^2}\right)$.
W_t, for $2 \leq t \leq n$, follows the Normal distribution $N(0, \sigma_W^2)$ and they are all independent of one another. Hence, the joint probability density of $(e_1, W_2, W_3, \cdots, W_n)$ is

$$\left[\frac{1-\phi^2}{2\pi\sigma_W^2}\right]^{1/2} exp\left[-\frac{e_1^2(1-\phi^2)}{2\sigma_W^2}\right] \times \left[\frac{1}{2\pi\sigma_W^2}\right]^{(n-1)/2} exp\left[-\frac{1}{2\sigma_W^2}\sum_{t=2}^{n} W_t^2\right]$$

Now consider the following transformation

$$\begin{aligned} \dot{Y}_1 &= e_1 \\ \dot{Y}_2 &= \phi \dot{Y}_1 + W_2 \\ \dot{Y}_3 &= \phi \dot{Y}_2 + W_3 \\ &\vdots \\ \dot{Y}_n &= \phi \dot{Y}_{n-1} + W_n \end{aligned}$$

(Note that the Jacobian of this transformation and its inverse is one).

It follows that the joint probability density of $(\dot{Y}_1, \dot{Y}_2, \dot{Y}_3, \cdots, \dot{Y}_n)$ is

$$\left[\frac{1-\phi^2}{2\pi\sigma_W^2}\right]^{1/2} exp\left[-\frac{\dot{Y}_1^2(1-\phi^2)}{2\sigma_W^2}\right] \times$$

$$\left[\frac{1}{2\pi\sigma_W^2}\right]^{(n-1)/2} exp\left[-\frac{1}{2\sigma_W^2}\sum_{t=2}^{n}(\dot{Y}_t - \phi\dot{Y}_{t-1})^2\right]$$

Hence for a given series $(\dot{Y}_1, \dot{Y}_2, \dot{Y}_3, \cdots, \dot{Y}_n)$ we have the following **exact log-likelihood function**:

$$\lambda(\phi, \mu, \sigma_W^2) = -\frac{n}{2} ln(2\pi) + \frac{1}{2} ln(1-\phi^2) - \frac{n}{2} ln\sigma_W^2 - \frac{S(\phi,\mu)}{2\sigma_W^2}$$

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

For given values of ϕ and μ, $\lambda(\phi, \mu, \sigma_W^2)$ can be maximized analytically with respect to σ_W^2 in terms of yet to be determined estimators of ϕ and μ.

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

(Just find $\frac{\partial \lambda}{\partial \sigma^2}$, set it equal to zero, and solve for σ_W^2.)

Summary

Starting with the most precise method and continuing in decreasing order of precision, we can summarize the various methods of estimating an AR(1) model as follows:

1. Exact likelihood method. Find parameters ϕ, μ such that $\lambda(\phi, \mu, \sigma_W^2)$ is maximized. This is usually nonlinear and requires numerical routines.
2. Unconditional least squares. Find parameters such that $S(\phi, \mu)$ is minimized. Again, nonlinearity dictates the use of numerical routines.
3. Conditional least squares. Find ϕ such that $S_C(\phi, \mu)$ is minimized. This is the simplest case since $\hat{\phi}$ can be solved analytically.