Parameter Estimation:

Least Squares

Week XI: Video 30

STAT 485/685, Fall 2020, SFU

Sonja Isberg

Video 30 Learning Objectives

By the end of this video, we should be able to:

- Describe the (conditional) Least Squares approach to parameter estimation for time series datasets
- Apply the LSE approach to estimate $\phi_1, \ldots, \phi_p$ and/or $\theta_1, \ldots, \theta_q$ in several AR/MA/ARMA models
- Understand why numerical optimization is often necessary with the LSE approach

An Introduction to Least Squares Estimation

We saw in Video 29 that Method of Moments estimation is not always very effective (e.g., for MA models, or for ARMA models with MA terms).

Another approach to parameter estimation is (conditional) **least squares estimation**.

An Introduction to Least Squares Estimation

We saw in Video 29 that Method of Moments estimation is not always very effective (e.g., for MA models, or for ARMA models with MA terms).

Another approach to parameter estimation is (conditional) **least squares estimation**.

Consider the multiple linear regression problem:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i \quad \text{for } i = 1, \ldots, n$$

The least-squares estimates of $\{\beta_0, \beta_1, \ldots \beta_k\}$ are the values that minimize the **sum-of-squares error function**:

$$S(\beta_0, \beta_1, \ldots, \beta_k) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \cdots - \beta_k x_{ki})^2 = \sum_{i=1}^{n} \hat{\epsilon}_i^2$$

We will use a similar approach for our time series datasets.

Least Squares Estimation for AR(1)

Recall the AR(1) model with a non-zero constant mean:

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

This can be thought of as a "regression model", in which we are predicting Y_t using the variable Y_{t-1}.

The white noise process terms e_t can be thought of as the residuals in our model, i.e. the "errors" in our prediction.

Then, the **conditional sum-of-squares function** is:

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

We would like to minimize this value with respect to ϕ and μ.

Least Squares Estimation for AR(1) (cont'd)

Estimating μ:

Take the partial derivative $\partial S_c / \partial \mu$, and set it to zero:

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

Least Squares Estimation for AR(1) (cont'd)

Estimating μ:

Take the partial derivative $\partial S_c / \partial \mu$, and set it to zero:

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

Solve for μ:

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

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

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

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

Least Squares Estimation for AR(1) (cont'd)

Estimating μ: *(cont'd)*

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

This equation contains the unknown parameter ϕ!

However, note that for large n:

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

Therefore, for large n:

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

So, for "large enough" sample size n, the LSE of μ is the same as the MOM estimator!

Least Squares Estimation for AR(1) (cont'd)

Now that we have an estimate of μ, we can plug it into S_c:

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

Least Squares Estimation for AR(1) (cont'd)

Now that we have an estimate of μ, we can plug it into S_c:

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

Estimating ϕ:

Minimize $S_c(\phi, \bar{Y})$ with respect to ϕ:

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

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

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

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

Least Squares Estimation for AR(1) (cont'd)

Estimating ϕ: *(cont'd)*

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

Recall the equation for the sample correlation at lag k:

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

For lag $k = 1$:

$$r_1 = \frac{\sum_{t=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

Least Squares Estimation for AR(1) (cont'd)

Estimating ϕ: *(cont'd)*

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

Recall the equation for the sample correlation at lag k:

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

For lag $k = 1$:

$$r_1 = \frac{\sum_{t=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

Therefore, the estimator of ϕ is almost exactly equal to r_1. It only lacks one term in the denominator: $(Y_n - \bar{Y})^2$. (Check this for yourself!)

For large n, this "missing term" is negligible, therefore $\hat{\phi} \approx r_1$. So, for "large enough" sample size n, the LSE of ϕ is the same as the MOM estimator!

Least Squares Estimation for AR(2)

Recall the AR(2) model with a non-zero constant mean:

$$(Y_t - \mu) = \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{t-2} - \mu) + e_t$$

The conditional sum-of-squares function for this model is:

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

We would like to minimize this value with respect to ϕ_1, ϕ_2 and μ.

Least Squares Estimation for AR(2)

Recall the AR(2) model with a non-zero constant mean:

$$(Y_t - \mu) = \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{t-2} - \mu) + e_t$$

The conditional sum-of-squares function for this model is:

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

We would like to minimize this value with respect to ϕ_1, ϕ_2 and μ.

Estimating μ:

It can be shown (using methods similar to AR(1)) that, for large n:

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

So, for "large enough" sample size n, the LSE of μ is the same as the MOM estimator.

Least Squares Estimation for AR(2) (cont'd)

Estimating ϕ:

Take the partial derivatives of $S_c(\phi_1, \phi_2, \bar{Y})$ with respect to ϕ_1 and ϕ_2 and set to zero.

After several "large-sample" approximations, we obtain the following:

$$r_1 \approx \phi_1 + r_1\phi_2$$
$$r_2 \approx r_1\phi_1 + \phi_2$$

Least Squares Estimation for AR(2) (cont'd)

Estimating ϕ:

Take the partial derivatives of $S_c(\phi_1, \phi_2, \bar{Y})$ with respect to ϕ_1 and ϕ_2 and set to zero.

After several "large-sample" approximations, we obtain the following:

$$r_1 \approx \phi_1 + r_1\phi_2$$
$$r_2 \approx r_1\phi_1 + \phi_2$$

These are just the sample Yule-Walker equations for AR(2). We solved them in Video 29:

$$\hat{\phi}_1 \approx \frac{r_1(1 - r_2)}{1 - r_1^2}$$
$$\hat{\phi}_2 \approx \frac{r_2 - r_1^2}{1 - r_1^2}$$

Once again, for "large enough" sample size n, the LSEs of ϕ_1 and ϕ_2 are the same as the MOM estimators.

Least Squares Estimation for AR(p)

It can be shown that for large n in an AR(p) model:

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

Least Squares Estimation for AR(p)

It can be shown that for large n in an AR(p) model:

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

In order to estimate $\phi_1, \ldots, \phi_p$, we must solve

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

Therefore, the LSEs for μ, ϕ_1, ..., ϕ_p are approximately equal to their corresponding MOM estimators, for large n.

Least Squares Estimation for MA(1)

Recall the MA(1) model (for this model, we will only consider the zero-mean case):

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

How can this be thought of as a "regression"?

Least Squares Estimation for MA(1)

Recall the MA(1) model (for this model, we will only consider the zero-mean case):

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

How can this be thought of as a "regression"?

Recall from Video 17 (slide 19) that any invertible MA(q) process can also be expressed as an AR(∞) process.

Specifically, for an MA(1): (proof is on pg. 79-80 of the textbook)

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

So, this can be thought of as a "regression model", in which we are predicting Y_t using the variables Y_{t-1}, $Y_{t-2}, Y_{t-3}, \ldots$.

Least Squares Estimation for MA(1) (cont'd)

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

The white noise process terms e_t can be thought of as the residuals in our model, i.e. the "errors" in our prediction.

Then, the conditional sum-of-squares function is:

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

We would like to minimize this value with respect to θ.

Least Squares Estimation for MA(1) (cont'd)

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

The white noise process terms e_t can be thought of as the residuals in our model, i.e. the "errors" in our prediction.

Then, the conditional sum-of-squares function is:

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

We would like to minimize this value with respect to θ.

However, we see two problems arising here:

- Each term in the summation is an infinite series
- The equation is highly nonlinear in θ

Least Squares Estimation for MA(1) (cont'd)

Instead of using the analytical equation to minimize $S_c(\theta)$, we will approximate this as follows:

1. Search over a grid of θ-values in $(-1, +1)$. For each value of θ:
   a) Set $e_0 = 0$ (This is a fairly safe assumption, since $E(e_t) = 0$ for all t)
   b) Note that the process can be written as
   $$e_t = Y_t + \theta e_{t-1}$$
   c) Obtain $e_1, e_2, \ldots, e_n$ iteratively:
   $$\begin{aligned} e_1 &= Y_1 + \theta e_0 = Y_1 \\ e_2 &= Y_2 + \theta e_1 \\ e_3 &= Y_3 + \theta e_2 \\ &\vdots \\ e_n &= Y_n + \theta e_{n-1} \end{aligned}$$
   d) Calculate $S_c(\theta) \approx \sum_{t=1}^{n} (e_t)^2$
2. Choose the value of θ that minimizes $S_c(\theta)$

Least Squares Estimation for MA(q)

A similar process applies for an MA(q) model:

1. Use a numerical optimization algorithm to search for an optimal set of parameters $\{\theta_1, \ldots, \theta_q\}$. For each candidate set:

   a) Set $e_0 = e_{-1} = \cdots = e_{-q} = 0$

   b) Note that the process can be written as

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

   c) Obtain $e_1, e_2, \ldots, e_n$ iteratively, using the above identity

   d) Calculate $S_c(\theta_1, \ldots, \theta_q) \approx \sum_{t=1}^{n} (e_t)^2$

2. Choose the set of parameters $\{\theta_1, \ldots, \theta_q\}$ that minimizes $S_c(\theta_1, \ldots, \theta_q)$

Least Squares Estimation for ARMA(1,1)

Recall the ARMA(1,1) model:

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

A similar problem occurs here as with MA models.

Therefore, we seek to numerically minimize $S_c(\phi, \theta) = \sum_{t=1}^{n} (e_t)^2$, using the fact that the model can be written as

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

Least Squares Estimation for ARMA(1,1)

Recall the ARMA(1,1) model:

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

A similar problem occurs here as with MA models.

Therefore, we seek to numerically minimize $S_c(\phi, \theta) = \sum_{t=1}^{n} (e_t)^2$, using the fact that the model can be written as

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

However, now we also have the "start-up problem" of choosing Y_0 as well (which we will need in order to evaluate e_1).

Since this is merely an optimization problem, it may be easiest to simply remove this term from the objective function:

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

Least Squares Estimation for ARMA(1,1) (cont'd)

Then we can follow the usual approach:

1. Use a numerical optimization algorithm to search for an optimal set of parameters $\{\phi, \theta\}$. For each candidate set:

   a) Set $e_1 = 0$

   b) Recall that the process can be written as

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

   c) Obtain $e_2, e_3, \ldots, e_n$ iteratively, using the above identity

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

2. Choose the set of parameters $\{\phi, \theta\}$ that minimizes $S_c(\phi, \theta)$

Final Comments

That's all for now!

In this video, we've learned how to estimate ARMA parameters using the (conditional) Least Squares Estimation method.

We saw how this approach results in many of the same estimators as the MOM approach.

We also saw how this approach usually needs to be completed numerically for MA and ARMA models.

Coming Up Next: Maximum Likelihood Estimation (MLE).

Thank you!

References:

[1] Cryer, J. D., & Chan, K. S. (2008). *Time series analysis: with applications in R.* Springer Science and Business Media.

[2] Chan, K. S., & Ripley, B. (2020). TSA: Time Series Analysis. R package version 1.2.1.