Parameter Estimation:

The Method of Moments

Week XI: Video 29

STAT 485/685, Fall 2020, SFU

Sonja Isberg

Our Roadmap

1. **Key Ideas:** Fundamental concepts (Ch. 1-2), Estimating trends (Ch. 3), Models for stationary time series (Ch. 4), Models for non-stationary time series (Ch. 5)
2. **Building a Model:**
   - **Model specification (Ch. 6):** How do we choose between the different models that we know?
   - **Parameter estimation (Ch. 7):** Now that we've chosen a model, there will be parameters whose values are unknown. How do we estimate these parameter values?
   - **Model diagnostics (Ch. 8):** How good is our chosen model? Should we be using a different model?
3. **Forecasting (Ch. 9)**
4. Other topics, as time permits.

Parameter Estimation: Introduction

In Videos 23-25 & 28, we learned about the first step of the Model-Building Process: *model specification*:

- Choosing between MA, AR, ARMA or ARIMA
- Choosing the values of p, d and/or q

Parameter Estimation: Introduction

In Videos 23-25 & 28, we learned about the first step of the Model-Building Process: *model specification*:

- Choosing between MA, AR, ARMA or ARIMA
- Choosing the values of p, d and/or q

The second step is *parameter estimation*, or *model fitting*. We need to estimate:

- $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots \theta_q$
- The white noise variance, σ_e^2
- The constant mean term μ, if there is one

Parameter Estimation: Introduction (cont'd)

We will learn about three different approaches to parameter estimation:

1. The Method of Moments (MOM)
2. Least Squares Estimation (LSE)
3. Maximum Likelihood Estimation (MLE)

Important Comment: Throughout this chapter, we will denote the stationary ARMA(p,q) series by $\{Y_1, \ldots, Y_n\}$. This may actually be a difference or some other transformation of the original dataset! We will ignore this fact for now.

Video 29 Learning Objectives

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

- Describe the general Method of Moments (MOM) approach to parameter estimation
- Apply the MOM approach to estimate $\phi_1, \ldots, \phi_p$ and/or $\theta_1, \ldots, \theta_q$ in an AR/MA/ARMA model
- Apply the MOM approach, and the above estimates, to estimate the process variance γ_0, and the white noise variance σ_e^2
- (If applicable) estimate the non-constant mean term μ and the corresponding "drift term" θ_0
- Understand why the MOM approach is not very useful for models with MA terms

Moment: Definition

Definition: For a random variable X with pdf $f(x)$, the $\boldsymbol{m}^{\text{th}}$ **moment of** $\boldsymbol{f(x)}$ **about a constant value** $\boldsymbol{c}$ is defined as:

$$E[(X-c)^m]$$

Moment: Definition

Definition: For a random variable X with pdf $f(x)$, the $\boldsymbol{m}^{\text{th}}$ **moment of $\boldsymbol{f(x)}$ about a constant value $\boldsymbol{c}$** is defined as:

$$E[(X-c)^m]$$

The expected value of X can be referred to as the *first moment of $f(x)$, about zero*.

The variance of X is the *second moment of $f(x)$, about the mean μ*:

$$Var(X) = E[(X-\mu)^2]$$

Moments about zero are sometimes referred to as **raw moments**, while moments about the mean μ are **central moments**.

Moment: Definition

Definition: For a random variable X with pdf $f(x)$, the $\boldsymbol{m}^{\text{th}}$ **moment of** $\boldsymbol{f(x)}$ **about a constant value** $\boldsymbol{c}$ is defined as:

$$E[(X-c)^m]$$

The expected value of X can be referred to as the *first moment of* $f(x)$*, about zero*.

The variance of X is the *second moment of* $f(x)$*, about the mean* μ:

$$Var(X) = E[(X-\mu)^2]$$

Moments about zero are sometimes referred to as **raw moments**, while moments about the mean μ are **central moments**.

Moments are often useful for describing the shape of the distribution. For example, *skewness* is a third moment, while *kurtosis* (a measure of how heavy the tails are) is a fourth moment.

The Method of Moments: Introduction

The **method of moments** is an approach to parameter estimation that works as follows:

1. Write all parameters of interest in terms of their relationship to theoretical moments. If you are estimating k parameters, you will need k equations.

2. Estimate all theoretical moments using their corresponding sample moments:

$$\begin{aligned}(\text{first } \widehat{\text{moment}}) &= (\text{first sample moment}) \\ (\text{second } \widehat{\text{moment}}) &= (\text{second sample moment}) \\ &\vdots \\ &\text{etc.}\end{aligned}$$

3. Solve for the parameters of interest.

The Method of Moments: Sample Moments

The m^{th} raw sample moment is:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^m$$

The Method of Moments: Sample Moments

The m^{th} raw sample moment is:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^m$$

Mean:

MOM estimator for a mean μ: $\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

The Method of Moments: Sample Moments

The m^{th} raw sample moment is:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^m$$

Mean:

MOM estimator for a mean μ: $\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

Variance:

Recall: $\sigma^2 = E[(X-\mu)^2] = E(X^2) - \mu^2$

So:

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

We often use the *unbiased MOM estimator of the variance*:

$$\hat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

The Method of Moments: Sample Moments (cont'd)

Correlation:

Recall:

$$\rho_{X,Y} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

$$= \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

$$= \frac{E(XY) - \mu_X \mu_Y}{\sigma_X \sigma_Y}$$

Therefore:

$$\hat{\rho}_{X,Y} = \frac{\frac{1}{n}\sum_{i=1}^n X_i Y_i - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)\left(\frac{1}{n}\sum_{i=1}^n Y_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2}\sqrt{\frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

$$= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

$$= r_{X,Y}$$

MOM Estimation for AR(1)

Recall the AR(1) model:

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

We will only focus on estimating ϕ for now.

MOM Estimation for AR(1)

Recall the AR(1) model:

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

We will only focus on estimating ϕ for now.

The Method of Moments:

1. Write all parameters of interest in terms of their relationship to theoretical moments. If you are estimating k parameters, you will need k equations:

$$\phi = \rho_1$$

2. Estimate all theoretical moments using their corresponding sample moments: $\hat{\rho}_1 = r_1$

3. Solve for the parameter(s) of interest:

$$\hat{\phi} = \hat{\rho}_1 = r_1$$

MOM Estimation for AR(2)

We will only focus on estimating ϕ_1 and ϕ_2 for now.

1. Write all parameters of interest in terms of their relationship to theoretical moments.

   We will need 2 equations. We know from Video 16 that $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$ for all $k \geq 1$. Therefore:

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_{-1} = \phi_1 + \phi_2 \rho_1 \tag{1}$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \rho_0 = \phi_1 \rho_1 + \phi_2 \tag{2}$$

2. Estimate all theoretical moments using their corresponding sample moments: $\hat{\rho}_1 = r_1$ and $\hat{\rho}_2 = r_2$

3. Solve (1) & (2) for the parameters of interest.

MOM Estimation for AR(2) (cont'd)

3. Solve (1) & (2) for the parameters of interest:

$$r_1 = \hat{\phi}_1 + \hat{\phi}_2 r_1 \tag{1}$$
$$r_2 = \hat{\phi}_1 r_1 + \hat{\phi}_2 \tag{2}$$

Solving (1) for $\hat{\phi}_1$:

$$\hat{\phi}_1 = r_1 - \hat{\phi}_2 r_1$$

Plugging this into (2), and solving for $\hat{\phi}_2$:

$$r_2 = (r_1 - \hat{\phi}_2 r_1) r_1 + \hat{\phi}_2$$
$$\vdots$$
$$\hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$$

Therefore:

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

MOM Estimation for AR(p)

1. We will need p equations relating $\phi_1, \ldots, \phi_p$ to theoretical moments. We know from Video 16 that $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p}$. Therefore, the Yule-Walker equations are:

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \cdots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \cdots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \cdots + \phi_p
\end{aligned}$$

2. Estimate all theoretical moments using their corresponding sample moments: $\hat{\rho}_1 = r_1, \ldots, \hat{\rho}_p = r_p$
3. Replace all ρ's with $\hat{\rho}$'s, and solve the Yule-Walker equations to obtain $\hat{\phi}_1, \ldots, \hat{\phi}_p$

MOM Estimation for MA(1)

Method of moments estimation is not nearly as convenient for MA models!

Recall the MA(1) model:

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

We will only focus on estimating θ for now.

1. Write all parameters of interest in terms of their relationship to theoretical moments.
$$\rho_1 = -\frac{\theta}{1+\theta^2}$$
2. Estimate all theoretical moments using their corresponding sample moments: $\hat{\rho}_1 = r_1$
3. Solve for the parameter(s) of interest:
$$r_1\hat{\theta}^2 + \hat{\theta} + r_1 = 0$$
$$\hat{\theta} = \frac{-1 \pm \sqrt{1-4r_1^2}}{2r_1}$$

MOM Estimation for MA(1) (cont'd)

3. Solve for the parameter(s) of interest:

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

We need to consider three possible scenarios:

A) **$|r_1| < 0.5$ (i.e., the value under the square root is positive):**

Then we have two unique, real solutions. However, notice that:

$$\left(\frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}\right)\left(\frac{-1 - \sqrt{1 - 4r_1^2}}{2r_1}\right) = 1$$

Therefore, only one of the solutions can satisfy the invertibility condition $|\theta| < 1$. The one that satisfies it is:

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

MOM Estimation for MA(1) (cont'd)

3. Solve for the parameter(s) of interest: *(cont'd)*

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

B) $|r_1| = 0.5$ **(i.e., the value under the square root is zero):**

Then we have one unique real solution:

$$\hat{\theta} = -\frac{1}{2r_1}$$

This is equal to -1 or 1, depending on the value of r_1. Therefore, this solution is not invertible.

MOM Estimation for MA(1) (cont'd)

3. Solve for the parameter(s) of interest: *(cont'd)*

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

B) $|r_1| = 0.5$ **(i.e., the value under the square root is zero):**

Then we have one unique real solution:

$$\hat{\theta} = -\frac{1}{2r_1}$$

This is equal to -1 or 1, depending on the value of r_1. Therefore, this solution is not invertible.

C) $|r_1| > 0.5$ **(i.e., the value under the square root is negative):**

Then no real solutions exist, so MOM fails to yield an estimator of θ.

Note: For MA(1), $|\rho_1| \leq 0.5$ (see pg. 58). So, if $|r_1| > 0.5$, then perhaps an MA(1) model does not make sense anyway.

MOM Estimation for MA(q)

1. We will need q equations relating $\theta_1, \ldots, \theta_q$ to theoretical moments. We know from Video 15 that

$$\rho_k = \frac{-\theta_k + \sum_{j=k+1}^{q} \theta_{j-k}\theta_j}{1 + \sum_{j=1}^{q} \theta_j^2} \quad \text{for } k = 1, 2, \ldots, q$$

   Therefore, we can plug in $k = 1, 2, \ldots, q$ to obtain q equations.

2. Estimate all theoretical moments using their corresponding sample moments: $\hat{\rho}_1 = r_1, \ \ldots, \ \hat{\rho}_q = r_q$

3. Replace all ρ's with $\hat{\rho}$'s, and solve the q equations above, to obtain $\hat{\theta}_1, \ldots, \hat{\theta}_q$

   Note: This is very complicated! The solution will usually be numerical.

   Note 2: There will be multiple solutions, only one of which is invertible.

Generally, MOM is quite complicated for MA models. It also often produces poor estimates.

MOM Estimation for ARMA(1,1)

Recall the ARMA(1,1) model:

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

Recall from Video 18:

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

MOM Estimation for ARMA(1,1)

Recall the ARMA(1,1) model:

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

Recall from Video 18:

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

We notice something interesting:

$$\phi = \frac{\rho_2}{\rho_1}$$

Therefore:

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

MOM Estimation for ARMA(1,1)

Recall the ARMA(1,1) model:

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

Recall from Video 18:

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

We notice something interesting:

$$\phi = \frac{\rho_2}{\rho_1}$$

Therefore:

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

Then, solve the following for θ to get $\hat{\theta}$:

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

Estimating the White Noise Variance, σ_e^2

Most expressions for σ_e^2 will include a term containing the process variance, $\gamma_0 = Var(Y_t)$.

We estimate γ_0 using the unbiased sample variance:

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

Then, we can use relationships between σ_e^2 and $\{\gamma_0, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q\}$, in order to estimate σ_e^2.

Estimating the White Noise Variance, σ_e^2

Most expressions for σ_e^2 will include a term containing the process variance, $\gamma_0 = Var(Y_t)$.

We estimate γ_0 using the unbiased sample variance:

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

Then, we can use relationships between σ_e^2 and $\{\gamma_0, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q\}$, in order to estimate σ_e^2.

Example: For AR(p): (from Video 16)

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

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

Estimating the White Noise Variance, σ_e^2 (cont'd)

Example: For MA(q): (from Video 15)

$$\gamma_0 = \sigma_e^2 \left(1 + \sum_{j=1}^{q} \theta_j^2\right)$$

$$\implies \hat{\sigma}_e^2 = \frac{s^2}{1 + \sum_{j=1}^{q} \hat{\theta}_j^2}$$

Estimating the White Noise Variance, σ_e^2 (cont'd)

Example: For MA(q): (from Video 15)

$$\gamma_0 = \sigma_e^2 \left(1 + \sum_{j=1}^{q} \theta_j^2 \right)$$

$$\implies \hat{\sigma}_e^2 = \frac{s^2}{1 + \sum_{j=1}^{q} \hat{\theta}_j^2}$$

Example: For ARMA(1,1): (from Video 18)

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

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

Example: Hare Data

The `hare` dataset in the `TSA` package gives the annual number of hares from 1905 to 1935.

In Ch. 6 (pg. 136), it was found that a square-root transformation would be most appropriate for this dataset.

On pg. 137, the sample PACF plot appeared to show that an AR(2) model might be appropriate for the square root series.

We find, for the square-root dataset, $r_1 = 0.736$ and $r_2 = 0.304$.

Then:

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

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

Example: Hare Data (cont'd)

The sample variance of this (square-root) series is $s^2 = 5.88$.

Therefore:

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

There may also be reason to believe that there is a constant mean in the data:

$$\hat{\mu} = \text{(sample mean of the square-root series)} = 5.82$$

Example: Hare Data (cont'd)

Therefore, if in this case we allow $Y_t =$ (original series):

$$\sqrt{Y_t} - \hat{\mu} = \hat{\phi}_1(\sqrt{Y_{t-1}} - \hat{\mu}) + \hat{\phi}_2(\sqrt{Y_{t-2}} - \hat{\mu}) + e_t$$

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

Example: Hare Data (cont'd)

Therefore, if in this case we allow $Y_t =$ (original series):

$$\sqrt{Y_t} - \hat{\mu} = \hat{\phi}_1(\sqrt{Y_{t-1}} - \hat{\mu}) + \hat{\phi}_2(\sqrt{Y_{t-2}} - \hat{\mu}) + e_t$$

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

Alternatively:

$$\sqrt{Y_t} = \hat{\theta}_0 + \hat{\phi}_1\sqrt{Y_{t-1}} + \hat{\phi}_2\sqrt{Y_{t-2}} + e_t$$

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

where (from Video 22)

$$\theta_0 = \mu(1 - \phi_1 - \cdots - \phi_p)$$

$$\implies \hat{\theta}_0 = \hat{\mu}(1 - \hat{\phi}_1 - \cdots - \hat{\phi}_p)$$

$$= 5.82(1 - 1.1178 + 0.519)$$

$$= 2.335$$

Other Examples

A few more examples can be found near the end of § 7.1 of the textbook, as well as § 7.5.

I encourage you to look through these examples, as they provide excellent practice on:

- MOM estimation
- Transformations
- Non-constant means in ARMA models
- All sorts of other things we've learned in this course so far

Final Comments

That's all for now!

In this video, we've learned how to estimate ARMA model parameters using the Method of Moments.

We also went through a brief example of this type of estimation. I encourage you to try more examples on your own!

Coming Up Next: Least Squares Estimation (LSE).

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.