Tutorial 10 - STAT 485/685

Trevor Thomson

Department of Statistics & Actuarial Science
Simon Fraser University, BC, Canada

November 23, 2020

Today's Plan

1. Recap of Tutorial 9
   - Nonstationarity
   - Other Specification Methods
   - Method of Moments
2. Parameter Estimation
   - Least Squares Estimation
   - Maximum Likelihood Estimation
   - Bootstrapping ARIMA Models
3. Examples
   - `hare` Dataset
   - `oil.price` Dataset

Recap of Tutorial 9

Nonstationarity

Question: What if $\{Y_t : t \in \mathcal{I}\}$ is non-stationary?

- $E(Y_t) \neq \mu$?
- $Cov(Y_t, Y_{t-k})$ depends on t?

Recall: If $\{Y_t : t \in \mathcal{I}\}$ is not stationary, derive a "new" stationary process $\{W_t : t \in \mathcal{I}\}$.

- **Approach 1**: Define $W_t = \nabla^d Y_t = \nabla(\nabla^{d-1} Y_t)$, where $d = 1$ or $d = 2$.
- **Approach 2**: Define $W_t = f(Y_t)$, for some function $f(.)$.

In terms or **Approach 1**, why do we only take $d = 1$ or $d = 2$?

$\Rightarrow$ If we *over-difference* the time series, we increase the model complexity!

$\Rightarrow$ Interpretation of ρ_k?

$\Rightarrow$ Asymptotic properties of r_k?

Let θ denote the model parameters.

- Akaike's Information Criterion (AIC)

$$AIC = -2\log\hat{\theta} + 2k,$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ, and k is the number of parameters in the model, and n is the number of observations.

- Corrected AIC (AIC_c)

$$AIC_c = AIC + \frac{2(k+1)(k+2)}{n-k-2}.$$

$\Rightarrow$ Corrects for the bias in AIC.

- Schwarz Bayesian Information Criterion (BIC)

$$BIC = -2\log\hat{\theta} + k\log n.$$

$\Rightarrow$ Which one to use?

- Use them all and proceed from there!

Recap of Tutorial 9

Other Specification Methods

Definition: A *subset* $ARMA(p,q)$ model is an $ARMA(p,q)$ model with a subset of its coefficients known to be zero.

Example:

$$Y_t = 0.8Y_{t-12} + e_t + 0.7e_{t-12}$$

$\Rightarrow$ this is an $ARMA(12,12)$ process with

$$\phi_1 = 0, \phi_2 = 0, \cdots, \phi_{11} = 0, \quad \phi_{12} = 0.8,$$
$$\theta_1 = 0, \theta_2 = 0, \cdots, \theta_{11} = 0, \quad \theta_{12} = -0.7.$$

Note that $\{Y_t : t \in \mathcal{I}\}$ satisfies the stationarity conditions.

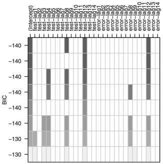
The textbook (see Section 6.5 page 132)...

1. ...fits a variety of models. Each row in Exhibit 6.22 corresponds to a model, where a variable is included if its cell is shaded.
2. ...sort the models in descending value with respect to some information criterion (e.g. BIC).

$\Rightarrow$ Summarize the results in a table.

$\Rightarrow$ Use the table to "identify" the model $Y_t = \phi_{12}Y_{t-12} + e_t - \theta_{12}e_{t-12}$.

Exhibit 6.22 Best Subset ARMA Selection Based on BIC



Basic Idea: For a random variable X, suppose that we want to estimate $g(\mu_1, \mu_2, \cdots, \mu_r)$, where $g(.)$ is a known function, and $\mu_k = E(X^k)$, for $k = 1, \cdots, r$.

$\Rightarrow$ Use $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ to estimate μ_k.

$\Rightarrow$ The *method of moment estimate* for $g(\mu_1, \cdots, \mu_k)$ is, then $g(\hat{\mu}_1, \cdots, \hat{\mu}_k)$.

Note that

$$r_k = \hat{\rho}_k = \widehat{Corr}(Y_t, Y_{t-k}) = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

is a method of moment estimator for ρ_k.

$\Rightarrow$ If $\{Y_t : t \in \mathcal{I}\}$ is an $ARMA(p, q)$ process, the goal is to use r_k to estimate $\phi_1, \cdots, \phi_p$ and $\theta_1, \cdots, \theta_q$.

Recap of Tutorial 9

Method of Moments

Autoregressive Process: Suppose that $\{Y_t : t \in \mathcal{I}\}$ is an $AR(p)$ process.

Recall the Yule-Walker equations

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

$\Rightarrow$ Replace ρ_k with r_k:

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

$\Rightarrow$ Solve the p equations for $\phi_1, \phi_2, \cdots, \phi_p$

$\Rightarrow$ Obtain the method of moment estimates $\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p$.

$\Rightarrow$ The method of moment estimator for σ_e^2 is

$$\hat{\sigma}_e^2 = s^2(1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2 - \cdots - \hat{\phi}_p r_p).$$

Moving Average Process: Suppose that $\{Y_t : t \in \mathcal{I}\}$ is an $MA(q)$ process.

Recall that

$$\rho_k = Corr(Y_t, Y_{t-k}) = \begin{cases} \dfrac{\sum_{\ell=0}^{q-k} \theta_{k+\ell}\theta_\ell}{\sum_{j=0}^{q} \theta_j^2} & \text{if } k = 0, 1, \cdots, q \\ 0 & \text{if } k > q \end{cases}.$$

$\Rightarrow$ Replace ρ_k with r_k:

$$r_k = \begin{cases} \dfrac{\sum_{\ell=0}^{q-k} \theta_{k+\ell}\theta_\ell}{\sum_{j=0}^{q} \theta_j^2} & \text{if } k = 0, 1, \cdots, q \\ 0 & \text{if } k > q \end{cases}.$$

$\Rightarrow$ Obtain q equations from $r_1, r_2, \cdots, r_q$.

$\Rightarrow$ Solve these equations for $\theta_1, \cdots, \theta_q$

$\Rightarrow$ Obtain the method of moment estimates $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_q$.

$\Rightarrow$ The method of moment estimator for σ_e^2 is

$$\hat{\sigma}_e^2 = \frac{s^2}{1 + \hat{\theta}_1^2 + \hat{\theta}_2^2 + \cdots + \hat{\theta}_q^2}.$$

Recall: For a time series $\{Y_t : t \in \mathcal{I}\}$ with $\mathcal{I} = \{1, \cdots, n\}$, the *linear trend model*:

$$Y_t = \mu_t + X_t,$$

where $\mu_t = \beta_0 + \beta_1 t$.

$$Y_t = \beta_0 + \beta_1 t + X_t.$$

- How do we estimate β_0 and β_1?

  $\Rightarrow$ Find the values of β_0 and β_1 that minimize

$$Q(\beta_0, \beta_1) = \sum_{t=1}^{n} X_t^2 = \sum_{t=1}^{n} (Y_t - \beta_0 - \beta_1 t)^2.$$

  $Q(\beta_0, \beta_1)$ is referred to as the *least squares objective function*.

$\Rightarrow$ Find the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ by

(a) Differentiating $Q(\beta_0, \beta_1)$ with respect to β_0 and β_1

(b) Set both of the resulting equations to zero.

(c) Solve the equations for β_0 and β_1.

Parameter Estimation

Least Squares Estimation

Autoregressive Process: Suppose that $\{Y_t : t \in \mathcal{I}\}$ is an $AR(p)$ process with mean μ:

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

$\Rightarrow$ Find the values of $\phi_1, \phi_2, \cdots, \phi_p$, and μ that minimize

$$\begin{aligned} S_c(\phi_1, \phi_2, \cdots, \phi_p, \mu) &= \sum_{t=p+1}^{n} e_t^2 \\ &= \sum_{t=p+1}^{n} [(Y_t - \mu) - \phi_1(Y_{t-1} - \mu) - \phi_2(Y_{t-2} - \mu) - \cdots - \phi_p(Y_{t-p} - \mu)]^2 \end{aligned}$$

$\Rightarrow$ The estimates $\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p$ and $\hat{\mu}$ are obtained by

(a) Differentiating $S_c(\phi_1, \phi_2, \cdots, \phi_p, \mu)$ with respect to $\phi_1, \phi_2, \cdots, \phi_p$, and μ

(b) Set the $p + 1$ resulting equations to zero.

(c) Solve the equations for $\phi_1, \phi_2, \cdots, \phi_p$, and μ.

$\Rightarrow$ It turns out that

(i) $\hat{\mu} \approx \bar{Y}$.

(ii) The remaining p equations involving $\phi_1, \phi_2, \cdots, \phi_p$ after some algebra...

$\Rightarrow$ very similar to the p sample Yule-Walker equations!

Moving Average Process: Suppose that $\{Y_t : t \in \mathcal{I}\}$ is an $MA(q)$ process:

$$\begin{aligned} Y_t &= e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \\ \Rightarrow e_t &= Y_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \end{aligned} \tag{1}$$

How do we perform least squares with error terms on both sides of the equation?

$\Rightarrow$ Using Equation (1), we can compute $e_1, e_2, \cdots, e_n$ recursively if we have initial conditions for $e_0, e_{-1}, e_{-2}, \cdots, e_{-q}$.

$\Rightarrow$ Initial conditions are $e_0 = e_{-1} = e_{-2} = \cdots = e_{-q} = \underbrace{0}_{E(e_t)}$.

Therefore, we have $e_t = e_t(\theta_1, \theta_2, \cdots, \theta_q)$ for any $t \in \mathcal{I} = \{1, \cdots, n\}$.

$\Rightarrow$ Hence $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_q$ are the minimizers of

$$S_c(\theta_1, \theta_2, \cdots, \theta_q) = \sum [e_t(\theta_1, \theta_2, \cdots, \theta_q)]^2$$

$\Rightarrow$ **Note**: $S_c(\theta_1, \theta_2, \cdots, \theta_q)$ is non-linear in $\theta_1, \theta_2, \cdots, \theta_q$.

$\Rightarrow$ $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_q$ are obtained via numerical methods (e.g. Gauss-Newton, Nelder-Mead).

Parameter Estimation

Maximum Likelihood Estimation

Basic Idea: Consider a random variable $Y \sim f(y;\theta)$, where $\theta \in \Theta$ is a finite-dimensional parameter, and Θ is the parameter space. Here, $f(.)$ is known, but θ is unknown.

- $f(.)$ is a probability density function if Y is continuous.
- $f(.)$ is a probability mass function if Y is discrete.

Suppose we observe $Y_1, Y_2, \cdots, Y_n$ from a sample of size n.

The *likelihood function* of θ is defined as

$$L(\theta) = L(\theta;\boldsymbol{y}) = g(\boldsymbol{y};\theta).$$

, where $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)'$, and $g(.)$ is the joint density function of $Y_1, Y_2, \cdots, Y_n$.

Note: If $Y_1, Y_2, \cdots Y_n$ is a *random* sample $\Rightarrow$ we observe n iid observations of Y

$$L(\theta) = g(\boldsymbol{y};\theta) = \prod_{i=1}^{n} f(y_i;\theta).$$

The *maximum likelihood estimate* (MLE) of θ is

$$\begin{aligned}\hat{\theta} &= \underset{\theta\in\Theta}{\operatorname{argmax}}\{L(\theta)\} \\ &= \underset{\theta\in\Theta}{\operatorname{argmax}}\{\log L(\theta)\}\end{aligned}$$

Interpretation: Most likely value of θ given the observed data.

Properties:

(1.) **Consistency**: $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \to \infty$, where θ_0 denotes the *true* value of θ.

(2.) **Invariance**: Let $h(\theta)$ be a function of θ. Then the MLE of $h(\theta)$ is $h(\hat{\theta})$.

(3.) **Asymptotic Normality**: $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, I^{-1}(\theta))$ as $n \to \infty$.

Here, $I(\theta)$ is referred to as the *Fisher Information matrix*:

$$I(\theta) = Var\left(\frac{\partial \log f(Y;\theta)}{\partial \theta}\right) = -E\left(\frac{\partial^2 \log f(Y;\theta)}{\partial \theta \partial \theta'}\right).$$

Note: A consistent estimate for $Var(\hat{\theta})$ is

$$\widehat{Var}(\hat{\theta}) = -\left(\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}\right)^{-1}\Bigg|_{\theta=\hat{\theta}}.$$

(4.) **Efficient**: Suppose $\tilde{\theta}$ is an unbiased estimator of θ. Then

$$Var(\tilde{\theta}) \geq \frac{I^{-1}(\theta_0)}{n}.$$

$\Rightarrow$ This is referred to as the *Cramer-Rao lower bound*.

Parameter Estimation
Maximum Likelihood Estimation

If we observe a time series $\{Y_t : t \in \mathcal{I}\}$ where $\mathcal{I} = \{1, 2, \cdots, n\} \Rightarrow$ we observe $Y_1, Y_2, \cdots, Y_n$

$$L(\theta) = g(\boldsymbol{y}; \theta) \neq \prod_{i=1}^{n} f(y_i; \theta).$$

This is because $Y_1, Y_2, \cdots, Y_n$ are not independent from each other.

However, we can still write the joint distribution of $Y_1, Y_2, \cdots, Y_n$ as

$$g(\boldsymbol{y}; \theta) = f(y_1) \times \underbrace{f(y_2, \cdots, y_n | y_1; \theta).}_{f(y_2|y_1;\theta) \times \cdots \times f(y_n|y_{n-1}, \cdots, y_2, y_1; \theta)}$$

- **Unconditional MLE**: $L(\theta) = g(y; \theta) = f(y_1; \theta) \times f(y_2, \cdots, y_n | y_1; \theta)$.
- **Conditional MLE**: $L(\theta) = g(y; \theta) \propto f(y_2, \cdots, y_n | y_1; \theta)$.

$\Rightarrow$ For large n, the unconditional MLE and conditional MLE are similar.

Parameter Estimation

Maximum Likelihood Estimation

Example: Suppose that $\{Y_t : t \in \mathcal{I}\}$ is an $AR(1)$ process with mean μ:

$$\begin{aligned} Y_t - \mu &= \phi_1(Y_{t-1} - \mu) + e_t \\ \Rightarrow e_t &= (Y_t - \mu) - \phi_1(Y_{t-1} - \mu) \end{aligned}$$

Suppose that $e_t \sim \mathcal{N}(0, \sigma_e^2)$

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

$\Rightarrow$ Letting $\theta = (\mu, \phi_1, \sigma_e^2)'$, the conditional MLE of θ is

$$\begin{aligned} L(\theta) &\propto f(y_2|y_1;\theta) \times f(y_3|y_2, y_1;\theta) \times \cdots \times f(y_n|y_{n-1}, \cdots, y_2, y_1;\theta) \\ &= f(y_2|y_1;\theta) \times f(y_3|y_2;\theta) \times f(y_n|y_{n-1};\theta) \quad \text{(since } \{Y_t : t \in \mathcal{I}\} \text{ is an } AR(1) \text{ process)} \\ &= (2\pi\sigma_e^2)^{-(n-1)/2} \exp\left\{ -\frac{1}{2\sigma_e^2} \sum_{t=2}^{n} [(y_t - \mu) - \phi_1(y_{t-1} - \mu)]^2 \right\}, \end{aligned}$$

since $Y_t|\{Y_{t-1} = y_{t-1}\} \sim \mathcal{N}(\mu + \phi_1[y_{t-1} - \mu], \sigma_e^2)$.

$\Rightarrow$ The MLE of θ is

$$\begin{aligned} \hat{\theta} &= \underset{\theta \in \Theta}{\operatorname{argmax}}\{L(\theta)\} \\ &= \underset{\theta \in \Theta}{\operatorname{argmax}}\{\log L(\theta)\}. \end{aligned}$$

Parameter Estimation

Bootstrapping ARIMA Models

Question 1: How do we estimate the variance (or standard error) for the least squares estimator, or method of moment estimator?

- Recall the simulation study we considered in Week 5's tutorial

  $\Rightarrow$ The `lm` function in R *under-estimates* the standard error.

Question 2: If $h(\theta)$ is a function of $\theta \Rightarrow$ the MLE of $h(\theta)$ is $h(\hat{\theta})$. What is the variance of $h(\hat{\theta})$?

- Delta method?

$\Rightarrow$ The *bootstrap* is a method to estimate the variance of the estimate.

Two versions: (i) non-parametric bootstrap, and (ii) parametric bootstrap. Let's focus on (ii).

Let $e_t \sim \mathcal{N}(0, \sigma_e^2)$, γ denote a vector consisting of all $ARMA(p,q)$ parameters, and $\hat{\gamma}$ denote the estimate of γ.

The parametric bootstrap entails following these three steps for each bootstrap $b = 1, 2, \cdots, B$:

(1) Obtain $\hat{e}_1^*, \hat{e}_2^*, \cdots \hat{e}_n^*$ denote *sampled* observations from the $\mathcal{N}(0, \hat{\sigma}_e^2)$ distribution, where $\hat{\sigma}_e^2$ is an estimate of σ_e^2.

(2) Under the $ARMA(p,q)$ model, use $\hat{\gamma}$ and $\{\hat{e}_t^*\}_{t=1}^n$ to generate new observations: $\hat{Y}_1^*, \hat{Y}_2^*, \cdots, \hat{Y}_n^*$.

(3) Use $\{\hat{Y}_t^*\}_{t=1}^n$ to estimate $\hat{\gamma}_b^*$.

$\Rightarrow$ The variance of $\hat{\gamma}$ is estimated by the (sample) variance of $\{\hat{\gamma}_b^*\}_{b=1}^B$.

Note: B is specified to be "large enough" (e.g. $B = 1000$).

Examples

Let's take a look at the `hare` and `oil.price` datasets!

⇒ We will fit model parameters with

- Method of Moments
- Least Squares Estimation
- Maximum Likelihood Estimation

⇒ See `Week12_Tutorial.R`

hare			
	Estimate (S.E.)		
Parameter	Method of Moment	Least Squares	Maximum Likelihood
ϕ_1	1.1177 (?)	1.3631 (0.1332)	1.3514 (0.1286)
ϕ_2	-0.5187 (?)	-0.7792 (0.1362)	-0.7763 (0.1242)
μ	5.8190 (?)	5.7086 (0.4783)	5.7134 (0.4753)
σ_e^2	1.9694 (?)	1.2249 (?)	1.2226 (?)

oil.price			
	Estimate (S.E.)		
Parameter	Method of Moment	Least Squares	Maximum Likelihood
θ_1	-0.2221 (?)	-0.2731 (0.0681)	-0.2956 (0.0693)
σ_e^2	0.0068 (?)	0.0067 (?)	0.0067 (?)

SFU