Tutorial 9 - STAT 485/685

Trevor Thomson

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

November 16, 2020

Today's Plan

1. Model Specification (Continued)
   - Nonstationarity
   - Other Specification Methods
2. Parameter Estimation
   - Method of Moments
3. Examples
   - `hare` Dataset
   - `oil.price` Dataset

Recall: For a *stationary* time series $\{Y_t : t \in \mathcal{I}\}$, where $\mathcal{I} = \{1, 2, \cdots, n\}$, recall the autocorrelation function (ACF) for lag $k \geq 0$:

$$\rho_k = Corr(Y_t, Y_{t-k}) = \frac{Cov(Y_t, Y_{t-k})}{\sqrt{Var(Y_t)Var(Y_{t-k})}} = \frac{\gamma_k}{\gamma_0}.$$

How do we estimate ρ_k?

$\Rightarrow Cov(Y_t, Y_{t-k}) = E[(Y_t - E(Y_t))(Y_{t-k} - E(Y_{t-k}))]$

$$\Rightarrow \widehat{Cov}(Y_t, Y_{t-k}) = \frac{1}{n-k}\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})$$

$$\underbrace{\approx}_{\text{for large } n} \frac{1}{n}\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})$$

where $\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$ is an estimate for $E(Y_t) = E(Y_{t-k}) = \mu$.

$\Rightarrow Var(Y_t) = Cov(Y_t, Y_t)$, and

$Var(Y_{t-k}) = Cov(Y_{t-k}, Y_{t-k})$

$\Rightarrow Var(Y_t) = Var(Y_{t-k})$ due to *stationarity*

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

The Sample ACF:

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

A useful hypothesis test for each $k > 0$:

$$\begin{aligned} H_0 &: \rho_k = 0 \\ H_a &: \rho_k \neq 0. \end{aligned}$$

To conduct this hypothesis test, we need the distribution of r_k:

- Turns out that if $\{Y_t : t \in \mathcal{I}\}$ is a stationary process:

  $$r_k \xrightarrow{d} \mathcal{N}\left(\rho_k, \frac{c_{kk}}{n}\right), \quad \text{as } n \to \infty,$$

  where

  $$c_{kk} = \sum_{i=-\infty}^{\infty} (\rho_{i+k}^2 + \rho_{i-k}\rho_{i+k} - 4\rho_k\rho_i\rho_{i+k} + 2\rho_k^2\rho_i^2).$$

Model Specification (Continued)
Nonstationarity

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

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

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

- **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 $d = 2$?

Model Specification (Continued)

Nonstationarity

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

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

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

- **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!

Example: Consider $\{Y_t : t \in \mathcal{I}\}$ where $Y_t = Y_{t-1} + e_t$. We see that this process fails the $AR(1)$ stationarity condition.

$d = 1$ Let $W_t = Y_t - Y_{t-1} = e_t$ so that $\{W_t : t \in \mathcal{I}\}$ is white noise (i.e. stationary). Therefore, $\{Y_t : t \in \mathcal{I}\}$ is an $ARIMA(0, 1, 0)$ process.

$d = 2$ If we let $X_t = W_t - W_{t-1} = e_t - e_{t-1}$, then $\{X_t : t \in \mathcal{I}\}$ is an $MA(1)$ process with $\theta_1 = 1$.

Note that the "true" θ_1 parameter is 0.

$d = 3$ If we let $Z_t = X_t - X_{t-1} = e_t - 2e_{t-1} + e_{t-1}$, then $\{Z_t : t \in \mathcal{I}\}$ is an $MA(2)$ process with $\theta_1 = 2$ and $\theta_2 = -1$.

Model Specification (Continued)

Nonstationarity

What if we use r_k to estimate ρ_k if $\{Y_t :\in \mathcal{I}\}$ is not stationary?

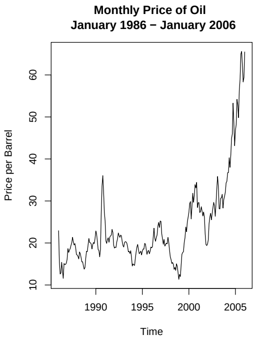
$\Rightarrow$ Interpretation of ρ_k?

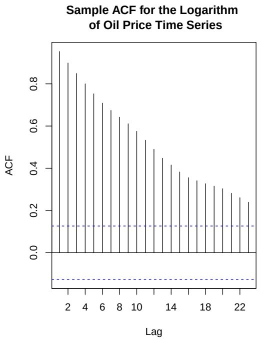
$\Rightarrow$ What are the asymptotic properties of r_k?

$\Rightarrow$ For $k > 0$, r_k fails to die out rapidly as the lags increase.

$\Rightarrow$ Similar challenges occur with the sample PACF.

$\Rightarrow$ **Example**: Consider the `oil.price` dataset.





Model Specification (Continued)

Other Specification Methods

We have so far used the sample EACF and following table to specify p and q:

	$AR(p)$	$MA(q)$	$ARMA(p, q)$
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

$\Rightarrow$ Sometimes the choice of p and q is not straightforward.

$\Rightarrow$ ... But what if we propose a variety of p or q values?

> $\Rightarrow$ Fit several models for different p and q values.
>
> $\Rightarrow$ Use *model selection* tools to let the data select the optimal p and q values.

Model Specification (Continued)

Other Specification Methods

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!

Model Specification (Continued)

Other Specification Methods

Consider the following time series $\{Y_t : t \in \mathcal{I}\}$, where

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

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

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

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

$\Rightarrow$ Similar idea as the *seasonal means* model we considered in Chapter 3.

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

$\Rightarrow$ How do we identify a subset $ARMA(p, q)$ model?

1. Suppose that we have a rough idea of what p and q are.
   - e.g. sample ACF / PACF plots.
2. Fit model(s) for a specified p and q
3. Use AIC / AIC_c / BIC to conduct model selection.

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

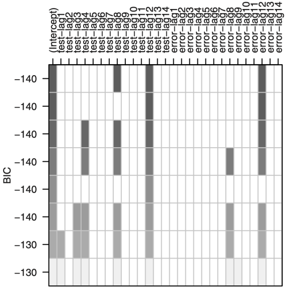
The textbook (see Section 6.5 page 132)...

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 $k = 1, \cdots, r$.

Example:

- If we let $r = 1$ and $g(x) = x \Rightarrow$ we want to estimate $g(\mu_1) = \mu_1$.
- If we let $r = 2$ and $g(x, y) = y - x^2 \Rightarrow$ we want to estimate $g(\mu_1, \mu_2) = \underbrace{\mu_2 - \mu_1^2}_{Var(X)}$

Since $g(.)$ is a known function, we need to estimate $\mu_1, \cdots, \mu_r$.

Note: If we let $X_1, \cdots, X_n$ denote n realizations of X from a random sample

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{p} \mu_k \quad \text{as } n \to \infty.$$

by the *law of large numbers*.

$\Rightarrow$ 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

$$g(\hat{\mu}_1, \cdots, \hat{\mu}_k).$$

Parameter Estimation

Method of Moments

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 $\theta_1, \cdots, \theta_q$.

Parameter Estimation

Method of Moments

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

Recall the Yule-Walker equations

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

$\Rightarrow$ Replace ρ_k with r_k:

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

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

$\Rightarrow$ Obtain the method of moment estimates $\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p$. (Continued on next slide)

Parameter Estimation

Method of Moments

If we want to estimate σ_e^2, we have that

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t \\ \therefore \underbrace{E(Y_t^2)}_{\gamma_0} &= \phi_1 \underbrace{E(Y_t Y_{t-1})}_{\gamma_1} + \phi_2 \underbrace{E(Y_t Y_{t-2})}_{\gamma_2} + \cdots + \phi_p \underbrace{E(Y_t Y_{t-p})}_{\gamma_p} + \underbrace{E(Y_t e_t)}_{\sigma_e^2} \\ \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \cdots + \phi_p \gamma_p + \sigma_e^2 \\ \therefore 1 &= \phi_1 \rho_1 + \phi_2 \rho_2 + \cdots + \phi_p \rho_p + \frac{\sigma_e^2}{\gamma_0} \\ \sigma_e^2 &= \gamma_0 (1 - \phi_1 \rho_1 - \phi_2 \rho_2 - \cdots - \phi_p \rho_p). \end{aligned}$$

$\Rightarrow$ if we estimate γ_0 with the sample variance

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

then the method of moment estimator for σ_e^2 is

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

Example: Suppose that $\{Y_t : t \in \mathcal{I}\}$ is an $AR(2)$ process. Then

$$
\begin{aligned}
r_1 &= \phi_1 + \phi_2 r_1 \qquad (1)\\
r_2 &= \phi_1 r_1 + \phi_2 \qquad (2)
\end{aligned}
$$

$\Rightarrow$ solve Equations (1) and (2) for ϕ_1 and ϕ_2:

$$
\begin{aligned}
\Rightarrow \phi_1 &= r_1(1-\phi_2) \quad \text{from (1)}\\
\Rightarrow r_2 &= \underbrace{r_1(1-\phi_2)}_{\phi_1} r_1 + \phi_2 \quad \text{from (2)}\\
&= r_1^2(1-\phi_2) + \phi_2\\
\Rightarrow \hat{\phi}_2 &= \frac{r_2 - r_1^2}{1-r_1^2}\\
\Rightarrow \hat{\phi}_1 &= r_1(1-\hat{\phi}_2) = \frac{r_1(1-r_2)}{1-r_1^2}.
\end{aligned}
$$

$\Rightarrow$ Estimate σ_e^2 with

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

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 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_q$. (Continued on next slide)

Parameter Estimation

Method of Moments

If we want to estimate σ_e^2, we have that

$$\gamma_k = Cov(Y_t, Y_{t-k}) = \begin{cases} \sigma_e^2 \sum_{\ell=0}^{q-k} \theta_{k+\ell}\theta_\ell & \text{if } k = 0, 1, \cdots, q \\ 0 & \text{if } k > q \end{cases}$$

so that in particular,

$$\gamma_0 = \sigma_e^2(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2).$$

$\Rightarrow$ if we estimate γ_0 with the sample variance

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

then the method of moment estimator for σ_e^2 is

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

Parameter Estimation

Method of Moments

Example: Suppose that $\{Y_t : t \in \mathcal{I}\}$ is an $MA(1)$ process. Then

$$r_1 = \frac{\theta_0 \theta_1}{\theta_0^2 + \theta_1^2} = -\frac{\theta_1}{1 + \theta_1^2} \tag{3}$$

$\Rightarrow$ solve Equation (3) for θ_1:

$$\Rightarrow r_1 \theta_1^2 + \theta_1 + r_1 = 0$$
$$\Rightarrow \hat{\theta}_1 = \frac{-1 \pm \sqrt{1 - 4r_1^2}}{2r_1}$$

Note that we for $\hat{\theta}_1$ to be a real number: we need

$$1 - 4r_1^2 \geq 0$$
$$|r_1| \leq 0.5$$

Note that $\hat{\theta}_1$ provides *two* estimates (due to "$\pm$"). Since we need $|\theta_1| < 1$, we take $\hat{\theta}_1$ as

$$\hat{\theta}_1 = \frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}$$

$\Rightarrow$ Estimate σ_e^2 with

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

Examples

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

- Use the `ar` function to estimate parameters from an $AR(p)$ process
  - Fits models of the form

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

$\Rightarrow$ we will fit an $AR(2)$ model for $W_t = \sqrt{Y}_t$ with the `hare` dataset:

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

$\Rightarrow$ $\hat{\mu} = 5.82$, $\hat{\phi}_1 = 1.12$, $\hat{\phi}_2 = -0.52$, and $\hat{\sigma}_e^2 = 1.97$.

- Use our defined `estimate.ma1.mom` function to estimate θ_1 from an $MA(1)$ process

$$Y_t - \mu = e_t - \theta_1 e_{t-1}$$

$\Rightarrow$ we will fit an $MA(1)$ model for $W_t = \log Y_t - \log Y_{t-1}$ with the `oil.price` dataset:

$$W_t - \mu = e_t - \theta_1 e_{t-1}$$

$\Rightarrow$ $\hat{\mu} = 0.004$, $\hat{\theta}_1 = -0.222$, and $\hat{\sigma}_e^2 = 0.007$.

$\Rightarrow$ See `Week11_Tutorial.R`