Tutorial 7 - STAT 485/685

Trevor Thomson

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

November 2, 2020

Today's Plan

1. Recap of Tutorial 6
   - Stationarity Through Differencing
   - ARIMA Models
   - Constant Terms in ARIMA Models
   - Other Transformations
2. Model Specification
   - Properties of the Sample Autocorrelation Function
   - The Partial and Extended Autocorrelation Function
3. Examples
   - `larain` Dataset
   - `wages` Dataset

Recap of Tutorial 6

Stationarity Through Differencing

Question: If a time series $\{Y_t : t \in \mathcal{I}\}$ is not stationary, can we find a stationary time series $\{W_t : t \in \mathcal{I}\}$, such that W_t is derived from $\{Y_t : t \in \mathcal{I}\}$?

- **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 $f(.)$.

Question: If a time series $\{Y_t : t \in \mathcal{I}\}$ is not stationary, can we find a stationary time series $\{W_t : t \in \mathcal{I}\}$, such that W_t is derived from $\{Y_t : t \in \mathcal{I}\}$?

- **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 $f(.)$.

For **Approach 1**, we showed that if

- $Y_t = \beta_0 + \beta_1 t + X_t$, where $\{X_t : t \in \mathcal{I}\}$ is a zero-mean stationary series with autocovariance function γ_k, and β_0 and β_1 are non-zero constants,

  $\Rightarrow \{W_t : t \in \mathcal{I}\}$ is stationary, where $W_t = \nabla Y_t$.

- $Y_t = Y_{t-1} + e_t$, where $\{e_t : t \in \mathcal{I}\}$ is white noise,

  $\Rightarrow \{W_t : t \in \mathcal{I}\}$ is stationary, where $W_t = \nabla Y_t$.

Recap of Tutorial 6

ARIMA Models

Definition: $\{Y_t : t \in \mathcal{I}\}$ is an integrated autoregressive moving average model if the dth difference $W_t = \nabla^d Y_t$ is a stationary $ARMA(p, q)$. That is, we can construct $\{W_t : t \in \mathcal{I}\}$, where

$$\begin{aligned} W_t &= [\phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t] + [-\theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}] \\ &= \sum_{j=1}^{p} \phi_j W_{t-j} - \sum_{j=0}^{q} \theta_j e_{t-j}. \end{aligned}$$

If so, we say that $\{Y_t : t \in \mathcal{I}\}$ is an $ARIMA(p, d, q)$ process.

Note: For practical purposes, we only allow for $d \in \{0, 1, 2\}$.

We can then apply the models from Chapter 4 with $\{W_t : t \in \mathcal{I}\}$.

$\Rightarrow$ Use the fact that $W_t = \nabla^d Y_t$ to then apply the model to $\{Y_t : t \in \mathcal{I}\}$.

Special Cases:

- $ARIMA(0, d, q) \Rightarrow IMA(d, q)$
- $ARIMA(p, d, 0) \Rightarrow ARI(p, d)$
- $ARIMA(p, 0, q) \Rightarrow ARMA(p, q)$

Recap of Tutorial 6

Constant Terms in ARIMA Models

If $\{W_t : t \in \mathcal{I}\}$ is an $ARMA(p, q)$ process, let

$$W_t^* = W_t + c$$

$\Rightarrow E(W_t^*) = c$

$\Rightarrow Cov(W_t^*, W_{t-k}^*) = Cov(W_t, W_{t-k})$.

$$W_t = \sum_{j=1}^{p} \phi_j W_{t-j} - \sum_{j=0}^{q} \theta_j e_{t-j}$$

$$W_t^* = \theta_0 + \sum_{j=1}^{p} \phi_j W_{t-j}^* - \sum_{j=0}^{q} \theta_j e_{t-j}$$

We see that

$$\theta_0 = c - \sum_{j=1}^{p} c\phi_j$$

$$c = \frac{\theta_0}{1 - \sum_{j=1}^{p} \phi_j}$$

$\Rightarrow$ if we include an intercept term in an $ARMA(p, q)$ model, we can model stationary processes with non-zero means.

- **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 $f(.)$.

If we want to transform our data, how to choose $f(.)$?

Box-Cox Power Transformations: For a given value of λ and for $Y_t > 0$ for all $t \in \mathcal{I}$, a *power transformation* with parameter λ is defined by

$$g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log x & \text{if } \lambda = 0 \end{cases}.$$

We see that if

- $\lambda = 0 \Rightarrow$ logarithm transformation
- $\lambda = \frac{1}{2} \Rightarrow$ square-root transformation
- $\lambda = -1 \Rightarrow$ inverse transformation
- $\lambda = 1 \Rightarrow$ no transformation

$\Rightarrow$ use an estimate $\hat{\lambda}$ to help us specify $f(x)$

In R: `BoxCox.ar`

- Computes a log-likelihood function for a grid of λ-values based on a normal likelihood function.
- Generates a 95% confidence interval for λ, where the centre is $\hat{\lambda}$.
- Use the 95% confidence interval to guide us in selecting a proper λ.

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 $k > 0$:

$$H_0 : \rho_k = 0$$
$$H_a : \rho_k \neq 0.$$

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

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

Therefore, for each $k > 0$ an approximate 95% confidence interval for ρ_k is

$$r_k \pm \underbrace{2}_{\approx 1.96} \widehat{SE}(r_k),$$

where $\widehat{SE}(r_k)$ is an estimate of $SE(r_k) = \sqrt{Var(r_k)}$.

- **Example 1**: If $\{Y_t : t \in \mathcal{I}\}$ is white noise,

$$\rho_k = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}.$$

Hence

$$\begin{aligned} c_{kk} &= \rho_0^2 + \underbrace{\rho_{-2k}}_{0} \rho_0 - 4\rho_k \underbrace{\rho_{-k}}_{0} \rho_0 + 2\rho_k^2 \underbrace{\rho_{-k}^2}_{0} \\ &= \rho_0^2 \\ &= 1 \end{aligned}$$

Then $Var(r_k) = \frac{1}{n}$ (which is known!)

$\Rightarrow$ 95% confidence interval for ρ_k is

$$r_k \pm \frac{2}{\sqrt{n}}.$$

Note: The `acf()` function in R plots the 95% confidence interval error bounds $\pm \frac{2}{\sqrt{n}}$ by default!

- **Example 2**: If $\{Y_t : t \in \mathcal{I}\}$ is an $MA(q)$ process

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

$\Rightarrow$ It turns out that

$$c_{kk} = 1 + 2\sum_{j=1}^{q} \rho_j^2 \quad \text{for } k > q$$

$$\Rightarrow Var(r_k) = \frac{1}{n}\left[1 + 2\sum_{j=1}^{q} \rho_j^2\right] \quad \text{for } k > q.$$

Therefore, if we are testing if $\{Y_t : t \in \mathcal{I}\}$ is an $MA(k-1)$ process

$$\widehat{SE}(r_k) = \sqrt{\frac{1}{n}\left[1 + 2\sum_{j=1}^{k-1} r_j^2\right]}$$

$\Rightarrow$ 95% confidence interval for ρ_k is

$$r_k \pm 2\sqrt{\frac{1}{n}\left[1 + 2\sum_{j=1}^{k-1} r_j^2\right]}$$

Note: Specify `ci.type = "ma"` with `acf()` in R to plot the corresponding 95% confidence interval error bounds.

SFU

- **Example 3**: If $\{Y_t : t \in \mathcal{I}\}$ is an $AR(p)$ process, ρ_k is obtained by solving the p Yule-Walker equations.

  Recall: The ACF for the AR(1) process:

$$\rho_k = \phi_1^k, \quad \text{for } k \geq 0.$$

$\Rightarrow$ We see that ρ_k *decays* with k.

  - The issue of the ACF being non-zero for all lags is present in an $AR(p)$ process.

$\Rightarrow$ Since the ACF does not *cut off* after any particular lag, we need another method for model specification.

Partial Correlation: Measure of association between random variables X and Y upon removing the effect of controlling variables $\boldsymbol{Z} = (Z_1, \cdots, Z_m)$', for some m:

$$\rho_{XY.\boldsymbol{Z}} = Corr(\hat{\varepsilon}_X, \hat{\varepsilon}_Y),$$

where

$$\hat{\varepsilon}_X = X - \boldsymbol{Z}'\hat{\boldsymbol{\beta}}$$
$$\hat{\varepsilon}_Y = Y - \boldsymbol{Z}'\hat{\boldsymbol{\alpha}},$$

where $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\alpha}}$ are estimated regression vectors.

For a time series: $\{Y_t : t \in \mathcal{I}\}$...

Partial Autocorrelation Function (PACF): The partial correlation between Y_t and Y_{t-k} upon removing the effect of $(Y_{t-1}, Y_{t-2}, \cdots, Y_{t-(k-1)})'$.

$\Rightarrow$ For a stationary time series $\{Y_t : t \in \mathcal{I}\}$, the PACF at lag k, denoted by ϕ_{kk}, is

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_j}$$

$$\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}, \quad \text{for } j = 1, 2, \cdots, k-1,$$

and $\phi_{11} = \rho_1$.

Model Specification

The Partial and Extended Autocorrelation Function

Problem: ρ_k, ϕ_{kk}, and $\phi_{k,j}$ are unknown parameters.

- Estimate them with $\underbrace{\hat{\rho}_k}_{r_k}$, $\hat{\phi}_{kk}$, and $\hat{\phi}_{k,j}$ with the observed time series $\{Y_t : t \in \mathcal{I}\}$!

$$\hat{\phi}_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_j}$$

$$\hat{\phi}_{k,j} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1,k-j}, \quad \text{for } j = 1, 2, \cdots, k-1,$$

and $\hat{\phi}_{11} = r_1$.

- Turns out that if $\{Y_t : t \in \mathcal{I}\}$ is an $AR(p)$ process:

$$\hat{\phi}_{kk} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{n}\right) \quad \text{as } n \to \infty.$$

Remark: $\phi_{kk} = 0$ for $k > p$. That is, the PACF *cuts off* after lag p if $\{Y_t : t \in \mathcal{I}\}$ is an $AR(p)$ process.

$\Rightarrow$ It turns out that the PACF of an $MA(q)$ process *decays* with k.

- Similar to the ACF with an $AR(p)$ process!

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

$$H_0 : \phi_{kk} = 0$$
$$H_a : \phi_{kk} \neq 0.$$

$\Rightarrow$ an approximate 95% confidence interval for ϕ_{kk} is

$$\hat{\phi}_{kk} \pm \underbrace{2}_{\approx 1.96} \widehat{SE}(\hat{\phi}_{kk})$$
$$\hat{\phi}_{kk} \pm \frac{2}{\sqrt{n}}.$$

Note 1: R will produce estimates $\hat{\phi}_{kk}$ for each $k > 0$, as well as plot $\hat{\phi}_{kk}$ and the corresponding 95% confidence error bounds with `pacf()`; ie $\pm\frac{2}{\sqrt{n}}$.

Note 2: The following table is useful for model identification purposes:

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

Question: Since the ACF and PACF "tails off" for an $ARMA(p,q)$ process, how can we identify p and q?

$\Rightarrow$ A variety of graphical tools have been provided to identify p and q

- The Extended Autocorrelation Function (EACF) is what we will consider using, as it has been shown to have good sampling properties for moderately large sample sizes.

Basic idea: If we knew p (i.e. the autoregressive component), we can "filter it out" of $\{Y_t : t \in \mathcal{I}\}$.

$\Rightarrow$ Obtain the "filtered time series" $\{W_t : t \in \mathcal{I}\}$

$\Rightarrow$ $\{W_t : t \in \mathcal{I}\}$ should "behave" like an $MA(q)$ process.

Specifically, suppose that we know p, and define

$$W_{t,p,j} = Y_t - \tilde{\phi}_1 Y_{t-1} - \tilde{\phi}_2 Y_{t-2} - \cdots - \tilde{\phi}_p Y_{t-p},$$

where $\{\tilde{\phi}_\ell\}_{\ell=1}^{p}$ are estimates of the $AR(p)$ coefficients.

$\Rightarrow$ $\{W_{t,p,j} : t \in \mathcal{I}\}$ should "behave" like an $MA(q)$ process

$\Rightarrow$ Specify $j = q$ by looking at the ACF of $\{W_{t,p,j} : t \in \mathcal{I}\}$.

Problem: We don't know p!

Solution: For each $k \in \{0, 1, 2, \cdots\}$, set $p = k$ and then determine q.

$\Rightarrow$ Have a variety of time series $\{W_{t,k,j} : t \in \mathcal{I}\}$.

Model Specification

The Partial and Extended Autocorrelation Function

We can summarize the information into a table:

- Values of k down the rows $\downarrow$
- Values of j across the columns $\rightarrow$

$\Rightarrow$ The (k, j)th element of the table corresponds to the sample ACF value with the time series $\{W_{t,k,j} : t \in \mathcal{I}\}$.

$\Rightarrow$ Use an "X" if the sample ACF value is statistically significant.

- **Recall**: the distribution of the sample ACF is approximately $\mathcal{N}\left(0, \frac{1}{\sqrt{n-k-j}}\right)$ if the process $\{W_{t,k,j} : t \in \mathcal{I}\}$ is (approximately) an $MA(j)$ process.

  $\Rightarrow$ Construct a 95% confidence interval for the *true* ACF value at lag k.

  $\Rightarrow$ An $ARMA(p, q)$ process should *theoretically* give a triangle of zeros, with the upper left-hand corner corresponding to the orders of the process:

Exhibit 6.4 Theoretical Extended ACF (EACF) for an ARMA(1,1) Model

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	0*	0	0	0	0	0	0	0	0	0	0	0	0
2	x	x	0	0	0	0	0	0	0	0	0	0	0	0
3	x	x	x	0	0	0	0	0	0	0	0	0	0	0
4	x	x	x	x	0	0	0	0	0	0	0	0	0	0
5	x	x	x	x	x	0	0	0	0	0	0	0	0	0
6	x	x	x	x	x	x	0	0	0	0	0	0	0	0
7	x	x	x	x	x	x	x	0	0	0	0	0	0	0

Examples

Due to randomness, we typically don't observe a "nice" table shown on the previous slide:

- The sample EACF usually looks a bit different than its corresponding theoretical EACF.

However, it should help *guide* us to select potential p and q values!

In R, the `eacf()` function computes the sample EACF.

Let's look at the `larain` and `wages` datasets!

- See the R file `Tutorial7.R`