Tutorial 3 - STAT 485/685

Trevor Thomson

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

September 28, 2020

Today's Plan

1. Recap of Tutorial 2
   - Weak Stationarity
   - Examples
2. Trends
   - Constant Mean
   - Linear Trend
   - Seasonal Trends
   - Residual Analysis
3. Examples
   - Example 1: Questions 3.5 & 3.11
   - Example 2: Questions 3.6 & 3.12

Recap of Tutorial 2

Weak Stationarity

- **Stochastic Process**: A collection of random variables indexed by some set $\mathcal{I}$:

$$\{Y_t : t \in \mathcal{I}\}$$

- We need to make simplifying (yet reasonable) assumptions about the structure of $\{Y_t : t \in \mathcal{I}\}$
  - Probability laws that govern the behaviour of the stochastic process do not change over time.
- Two forms of stationarity; only interested in *weak stationarity*...
  - If we say $\{Y_t : t \in \mathcal{I}\}$ is *stationary* $\Rightarrow$ $\{Y_t : t \in \mathcal{I}\}$ is *weakly stationary*.

Recap of Tutorial 2

Weak Stationarity

- **Weakly Stationary** (AKA Second-Order Stationary): $\{Y_t : t \in \mathcal{I}\}$ is weakly stationary if it satisfies conditions (A) and (B) below:

  (A) $E(Y_t) = \mu$ - mean function is constant over time
  (B) $\gamma_{t,t-k} = \gamma_{0,k}$ - variance is constant over time, and $Cov(Y_t, Y_{t-k})$ depends only on k.

- To show that $\{Y_t : t \in \mathcal{I}\}$ is stationary, we need to show

  (I) $E(Y_t)$ does not depend on time t.
  (II) $Cov(Y_t, Y_{t-k})$ does not depend on time t.

  If either (I) or (II) does not hold, $\{Y_t : t \in \mathcal{I}\}$ is not stationary.

- We looked at Questions 2.14, 2.19, and 2.21
- In general, the questions constructed time series $\{Y_t : t \in \mathcal{I}\}$ from $\{e_t : t \in \mathcal{I}\}$.
    - **The sequence $\{e_t : t \in \mathcal{I}\}$ is a white noise process**:
      A collection of iid random variables with $E(e_t) = 0$, and $Var(e_t) = \sigma_e^2$.

SFU

Recap of Tutorial 2
Examples

- We looked at Questions 2.14, 2.19, and 2.21

- In general, the questions constructed time series $\{Y_t : t \in \mathcal{I}\}$ from $\{e_t : t \in \mathcal{I}\}$.
  - **The sequence** $\{e_t : t \in \mathcal{I}\}$ **is a white noise process**:
    A collection of iid random variables with $E(e_t) = 0$, and $Var(e_t) = \sigma_e^2$.

- To determine if $\{Y_t : t \in \mathcal{I}\}$ is stationary, we generally followed these four steps:
  1. Write Y_t in terms of $e_t, e_{t-1}, \cdots$.
  2. Compute $E(Y_t)$
     (a) Exploit the fact that $E(e_t) = 0$.
     (b) Check to see if $E(Y_t)$ is a function of time t.
  3. Compute $Cov(Y_t, Y_{t-k})$
     (a) Use properties of covariances (see Appendix of Ch. 2) to get $Cov(Y_t, Y_{t-k})$ in terms of $Cov(e_{t-j_1}, e_{t-k_1})$, $Cov(e_{t-j_2}, e_{t-k_2})$, $\cdots$, $Cov(e_{t-j_m}, e_{t-k_m})$, for some m.
     (b) Exploit the fact that

$$Cov(e_{t-j_\ell}, e_{t-k_\ell}) = \begin{cases} \sigma_e^2 & \text{if } j_\ell = k_\ell \\ 0 & \text{otherwise.} \end{cases} \quad \ell = 1, \cdots, m$$

     (c) Check to see if $Cov(Y_t, Y_{t-k})$ is a function of time t .
  4. Assess if $\{Y_t : t \in \mathcal{I}\}$ is stationary through Steps 2. (b) and 3. (c).

SFU

Example - 2.14 (b): Let $Y_t = \nabla X_t = X_t - X_{t-1}$, where $X_t = \theta_0 + te_t$, for some constant θ_0.

1. $Y_t = \theta_0 + te_t - [\theta_0 + (t-1)e_{t-1}] = te_t - (t-1)e_{t-1}$
2. $E(Y_t) = E(te_t - (t-1)e_{t-1}) = tE(e_t) - (t-1)E(e_{t-1}) = 0.$
3. $$Cov(Y_t, Y_{t-k}) = t(t-k)\underbrace{Cov(e_t, e_{t-k})}_{Cov(e_{t-j_1}, e_{t-k_1})} - t(t-k-1)\underbrace{Cov(e_t, e_{t-(k+1)})}_{Cov(e_{t-j_2}, e_{t-k_2})}$$
$$-(t-1)(t-k)\underbrace{Cov(e_{t-1}, e_{t-k})}_{Cov(e_{t-j_3}, e_{t-k_3})} + (t-1)(t-k-1)\underbrace{Cov(e_{t-1}, e_{t-k-1})}_{Cov(e_{t-j_4}, e_{t-k_4})}$$

$j_1 = 0$, $k_1 = k$.
Need $j_1 = k_1$
$\Rightarrow k = 0$.

$j_2 = 0$, $k_2 = k + 1$.
Need $j_2 = k_2$
$\Rightarrow k = -1$.

$j_3 = 1$, $k_3 = k$.
Need $j_3 = k_3$
$\Rightarrow k = 1$.

$j_4 = 1$, $k_4 = k + 1$.
Need $j_4 = k_4$
$\Rightarrow k = 0$.

$$Cov(Y_t, Y_{t-k}) = \begin{cases} -t^2\sigma_e^2 & \text{if } k = -1 \\ [t^2 + (t-1)^2]\sigma_e^2 & \text{if } k = 0 \\ -(t-1)^2\sigma_e^2 & \text{if } k = 1 \\ 0 & \text{otherwise} \end{cases}$$

4. Since $Cov(Y_t, Y_{t-k})$ is a function of t, $\{Y_t\}$ is not stationary.

Trends

- For a time series $\{Y_t : t \in \mathcal{I}\}$, we can express each term as

$$Y_t = \mu_t + X_t,$$

where $E(Y_t) = \mu_t \Rightarrow E(X_t) = 0$.

- **Goal**: Model and estimate μ_t.
    - μ_t *may* or *may not* depend on time t.

In what follows, we specify $\mathcal{I}$ to be $\mathcal{I} = \{1, \cdots, n\}$.

Trends

- For a time series $\{Y_t : t \in \mathcal{I}\}$, we can express each term as

$$Y_t = \mu_t + X_t,$$

  where $E(Y_t) = \mu_t \Rightarrow E(X_t) = 0$.

- **Goal**: Model and estimate μ_t.
  - μ_t *may* or *may not* depend on time t.

  In what follows, we specify $\mathcal{I}$ to be $\mathcal{I} = \{1, \cdots, n\}$.

There are a variety of forms we can specify for μ_t. The ones we consider in Ch. 3 are:

(a) $\mu_t = \mu$ - constant mean

(b) $\mu_t = \beta_0 + \beta_1 t$ - linear trend

(c) $\mu_t = \mu_{t+k}$ for some k - seasonal mean

E.g.

- $k = 1$ - same as the constant mean
- $k = 12$ - monthly
- $k = 4$ - seasonal
- $k = 24$ - hourly

(d) $\mu_t = \beta_0 + \beta \cos(2\pi f t + \Phi) = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$ - cosine trend

SFU

Constant Mean Model:

$$Y_t = \mu + X_t,$$

where $E(X_t) = 0$, and $\mu_t = \mu$ is a constant.

- Most natural estimator for μ is the *sample mean*:

$$\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t.$$

- We will fit the constant mean model with the `larain` dataset in R.
  - See `Tutorial3.R`

Constant Mean Model:

$$Y_t = \mu + X_t,$$

where $E(X_t) = 0$, and $\mu_t = \mu$ is a constant.

- Most natural estimator for μ is the *sample mean*:

$$\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t.$$

- We will fit the constant mean model with the `larain` dataset in R.
    - See `Tutorial3.R`
- **Note**: Let $\mu = \beta_0$, so that

$$Y_t = \beta_0 + X_t.$$

Since $E(X_t) = 0$, this looks like a regression model with only an intercept term!

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*.

Exercise: What is wrong with the following objective function?

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

- 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.

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} (Y_t - \bar{Y})(t - \bar{t})}{\sum_{t=1}^{n} (t - \bar{t})^2},$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{t},$$

where $\bar{t} = \frac{1}{n} \sum_{t=1}^{n} t = \frac{n+1}{2}$.

Therefore, we estimate μ_t with

$$\hat{\mu}_t = \hat{\beta}_0 + \hat{\beta}_1 t.$$

- We will fit the linear trend model with the `wages` dataset in R.
  - See `Tutorial3.R`

- We will fit the linear trend model with the `wages` dataset in R.
  - See `Tutorial3.R`

- **Remark 1**: Consider the linear trend model with $\beta_1 = 0$:

$$Y_t = \beta_0 + \underbrace{0}_{\beta_1} \times t + X_t = \beta_0 + X_t.$$

Since we know $\beta_1 = 0$, the "best estimate" for it $\hat{\beta}_1 = 0$. Then the estimator for $\hat{\beta}_0$ is

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{t} = \bar{Y}.$$

- **Remark 2**: We can consider including further powers of t in μ_t:
  - $\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2$: *quadratic trend model*
  - $\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$: *cubic trend model*
  - $\ldots$

The same procedure to estimate μ_t applies.

Seasonal Mean Model:

$$Y_t = \mu_t + X_t,$$

where $\mu_t = \mu_{t+k}$. For example, $k = 12$:

$$\mu_t = \begin{cases} \beta_1 & \text{for } t = 1, 13, 25, \cdots \\ \beta_2 & \text{for } t = 2, 14, 26, \cdots \\ \vdots & \vdots \\ \beta_{12} & \text{for } t = 12, 24, 36, \cdots \end{cases}$$

We estimate β_j $(j = 1, \cdots, k)$ by minimizing the least squares objective function $Q(\beta_1, \beta_2, \cdots, \beta_{12})$.

Note that we can write the model as

$$\begin{aligned} Y_t &= \beta_1 I(t \in \{1, 13, 25, \cdots\}) + \beta_2 I(t \in \{2, 14, 26, \cdots\}) + \cdots + \beta_{12} I(t \in \{12, 24, 36, \cdots\}) + X_t \\ &= \beta_1 Z_{1t} + \beta_2 Z_{2t} + \cdots \beta_{12} Z_{12,t} + X_t. \end{aligned}$$

Note: There is no intercept term. What happens if we include one?

$\Rightarrow$ The model becomes

$$Y_t = \beta_0 + \beta_1 Z_{1t} + \beta_2 Z_{2t} + \cdots \beta_{12} Z_{12,t}.$$

Then $E(Y_t) = \mu_t = \beta_0 + \beta_j$, for some $j \in \{1, 2, \cdots, 12\}$ (depending on t).

The issue is that we have 13 parameters ($\beta_0, \beta_1, \beta_2, \cdots, \beta_{12}$), but only 12 (linearly independent) equations.

In other words, we can estimate the sum $\beta_0 + \beta_j$ for each j, but we cannot estimate β_0 and β_j on their own.

Note: There is no intercept term. What happens if we include one?

$\Rightarrow$ The model becomes

$$Y_t = \beta_0 + \beta_1 Z_{1t} + \beta_2 Z_{2t} + \cdots \beta_{12} Z_{12,t}.$$

Then $E(Y_t) = \mu_t = \beta_0 + \beta_j$, for some $j \in \{1, 2, \cdots, 12\}$ (depending on t).

The issue is that we have 13 parameters $(\beta_0, \beta_1, \beta_2, \cdots, \beta_{12})$, but only 12 (linearly independent) equations.

In other words, we can estimate the sum $\beta_0 + \beta_j$ for each j, but we cannot estimate β_0 and β_j on their own.

If we want an intercept term, we need to drop one of the β_j terms. By dropping β_1 (the first term),

$$\begin{aligned} Y_t &= \alpha_0 + \alpha_2 I(t \in \{2, 14, 26, \cdots\}) + \cdots + \alpha_{12} I(t \in \{12, 24, 36, \cdots\}) + X_t \\ &= \alpha_0 + \alpha_2 Z_{2t} + \cdots \alpha_{12} Z_{12,t} + X_t, \end{aligned}$$

with $\alpha_1 = 0$. The two models we specified are hence related as

$$\beta_j = \alpha_0 + \alpha_j, \qquad j = 1, \cdots, 12.$$

- The estimator for β_j is

$$\hat{\beta}_j = \frac{1}{N} \sum_{i=0}^{N-1} Y_{j+12i},$$

where N is the number of (complete) years in the data. Note that since $\alpha_0 = \beta_1$,

$$\hat{\alpha}_0 = \frac{1}{N} \sum_{i=0}^{N-1} Y_{1+12i},$$

and since $\beta_j = \alpha_0 + \alpha_j$,

$$\hat{\alpha}_j = \hat{\beta}_j - \hat{\alpha}_0.$$

- We will fit the seasonal trends model with the `tempdub` dataset in R.
  - See `Tutorial3.R`

Cosine Trend Model:

$$Y_t = \mu_t + X_t,$$

where $\mu_t = \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$.

$$Y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + X_t.$$

- Here, f is the frequency of the curve. In the case of observations on a monthly time scale, $f = 1/12$.
- Here, β_0, β_1, β_2 are unknown parameters.
- This looks like a regression model with $\cos(2\pi f t)$, and $\sin(2\pi f t)$ as covariates.
  - Estimate β_0, β_1, and β_2 by minimizing the least squares objective function $Q(\beta_0, \beta_1, \beta_2)$

- We will fit the cosine trend model with the `tempdub` dataset in R.
  - See `Tutorial3.R`

SFU

- We will fit the cosine trend model with the `tempdub` dataset in R.
  - See `Tutorial3.R`

- **Remark**: We have two seasonal trends models. Which one should we use?
  - **Seasonal Mean:** The parameters have a nice interpretation, but there are 12 parameters we need to estimate (which can be a lot!).
  - **Cosine Trend:** The interpretation of the parameters is tricky, but there are only 3 parameters we need to estimate.

- With $Y_t = \mu_t + X_t$,
  - Specify a model for μ_t.
  - Estimate the parameters
    - Obtain $\hat{\mu}_t$
  - **Now what?**

- Estimate the "error term" X_t with

$$\hat{X}_t = Y_t - \hat{\mu}_t.$$

- **Basic idea**: If we specify a "good" model for $\mu_t \Rightarrow X_t \approx 0$.
  - If we properly accounted for the autocorrelation structure present within $\{Y_t : t \in \mathcal{I}\}$, then $\{X_t : t \in \mathcal{I}\}$ should "behave" like white noise.
  - Since we don't know $\{X_t : t \in \mathcal{I}\} \Rightarrow$ we use $\{\hat{X}_t : t \in \mathcal{I}\}$ instead.

- Recall, $\{X_t : t \in \mathcal{I}\}$ is a white noise process if
  - X_t are iid random variables.
  - $E(X_t) = 0$.
  - $Var(X_t) = \sigma_e^2$.

  There are a few general checks to see if $\{X_t : t \in \mathcal{I}\}$ "behaves" like white noise, and other general goodness-of-fit procedures we can conduct:

- **Estimate** σ_e

$$\hat{\sigma}_e = s = \sqrt{\frac{1}{n-p}\sum_{t=1}^{n}(Y_t - \hat{\mu}_t)^2},$$

where p is the number of parameters.

- **Compute R^2 - Coefficient of Determination**

$$R^2 = 1 - \frac{\sum_{t=1}^{n} (Y_t - \hat{\mu}_t)^2}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}.$$

Here, R^2 is the proportion of the variation in Y_t that the variables in $\hat{\mu}_t$ can explain.

  - Generally, $R^2 \approx 1$ means that that the $\hat{\mu}_t$ fits the data well.

- **Compute The Adjusted R^2 - Adjusted Coefficient of Determination**

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p^*-1},$$

where p^* is the number of covariates in the model (excluding the intercept).

  - R^2 increases if we keep adding variables to the model.
  - $\bar{R}^2$ at least accounts for the number of variables included in the model.

- **Plot The Residuals Over Time**

  If $\{X_t : t \in \mathcal{I}\}$ "behaves" like white noise, we should see a random scatter around 0.

- **Histogram of Residuals**

  Empirically plot the distribution of $\{X_t : t \in \mathcal{I}\}$, to see if it follows the normal distribution.

- **Normal Q-Q Plot**

  Another empirical check to see if $\{X_t : t \in \mathcal{I}\}$ is normally distributed.

- **Perform The Runs Test**:

$$\begin{aligned} H_0 &: \text{Elements of } \{X_t : t \in \mathcal{I}\} \text{ are mutually independent} \\ H_a &: \text{Not } H_0. \end{aligned}$$

If we fail to reject H_0, X_t are independent random variables.

- **Compute the Sample Autocorelation Function (ACF)**

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

This is a very useful quantity, **we will see this again later in the course!**

We can perform the following hypothesis test for each k:

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

If we fail to reject H_0 for each k, X_t are independent random variables.

- **Compute the Sample Autocorelation Function (ACF)**

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

  This is a very useful quantity, **we will see this again later in the course!**

  We can perform the following hypothesis test for each k:

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

  If we fail to reject H_0 for each k, X_t are independent random variables.

- We conduct residual diagnostics with the `tempdub` dataset with the cosine trend model in R.
  - See `Tutorial3.R`

Examples

- See the R file `Tutorial3.R` where solutions are provided for Questions 3.5, 3.6, 3.11, and 3.12.