Xyba Project

Analisis Runtun Waktu

Pembahasan UTS Tahun 2017

1. This document is version: 0.7.7
   Version should be at least 0.9 if you want to share this document to other people
2. You may not share this document if version is less than 1.0 unless you have my permission to do so
3. This document is created by Xyba, Student of Mathematics University of Indonesia Batch 2016
4. Should there be any mistakes or feedbacks you'd like to give, please contact me
5. Last Updated: 19/10/2019

Thank you for your cooperation >v<

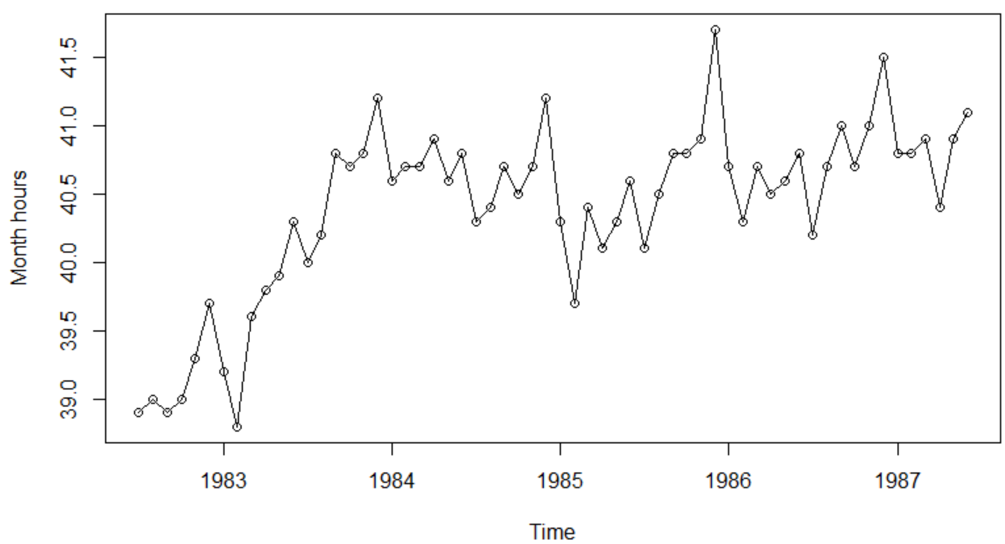
Soal

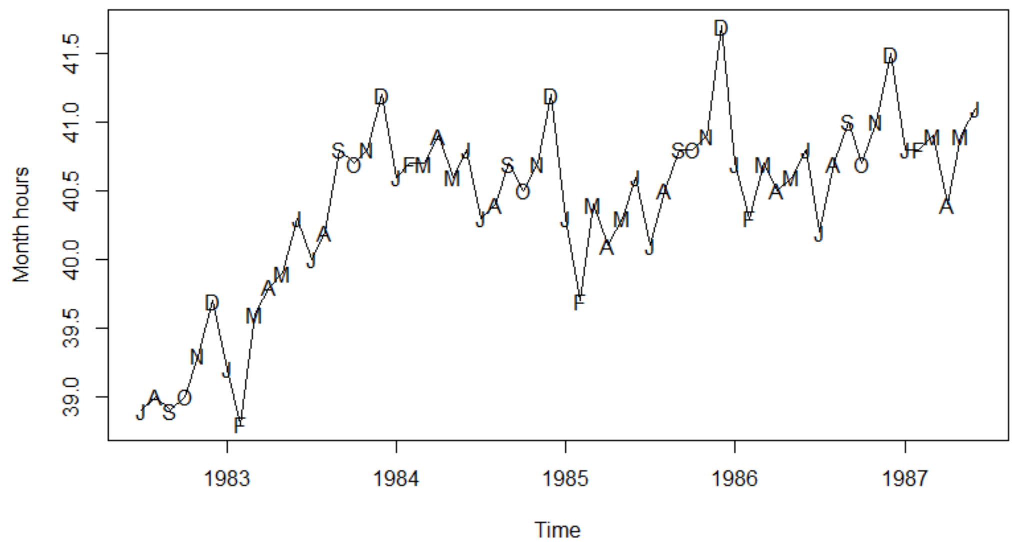
1. Misalkan $\{X_t\}$ adalah runtun waktu yang stasioner dan didefinisikan:
$$Y_t = \begin{cases} X_t & \text{, untuk } t \text{ ganjil} \\ X_t + 6 & \text{, untuk } t \text{ genap} \end{cases}$$
   a. Tentukan $\text{Cov}(Y_t, Y_{t-k})$ untuk semua lag k. Apakah bebas dari t?
   b. Apakah $\{Y_t\}$ stasioner?

2. Diketahui $\{X_t\}$ adalah runtun waktu stasioner dengan fungsi autokovariansi $\gamma_X(k)$. Didefinisikan runtun baru $\{Y_t\}$ yang sebagai berikut:
$$Y_t = \nabla^s X_t$$
   a. Nyatakan fungsi autokovariasi dari $\{Y_t\}$ sebagai fungsi dari $\gamma_X(k)$.
   b. Jika diberikan $X_t = \alpha X_{t-1} + e_t, |\alpha| < 1$, dimana $\{e_t\}$ adalah *white noise*, tentukan fungsi autokovariansi dari $\{Y_t\}$.
   c. Gunakan $s = 1$ yaitu $Y_t = \nabla X_t$, tentukan untuk nilai α berapa sehingga variansi Y_t akan lebih kecil dari variansi X_t.

3. Perhatikan dua model runtun waktu berikut:
   A: $Y_t = 0.9Y_{t-1} + 0.09Y_{t-2} + e_t$
   B: $Y_t = Y_{t-1} + e_t - 0.1e_{t-1}$
   a. Identifikasi masing-masing model sebagai model ARIMA yaitu berapa p, d, dan q dan nilai dari masing-masing parameter ϕ dan θ?
   b. Dalam hal apa kedua model itu mirip? Jelaskan.
   c. Dalam hal apa kedua model itu berbeda? Jelaskan.

4. Berikut ini adalah plot mengenai rata-rata jumlah jam kerja per bulan dari buruh suatu pabrik dari Juli 1982 – Juni 1987.





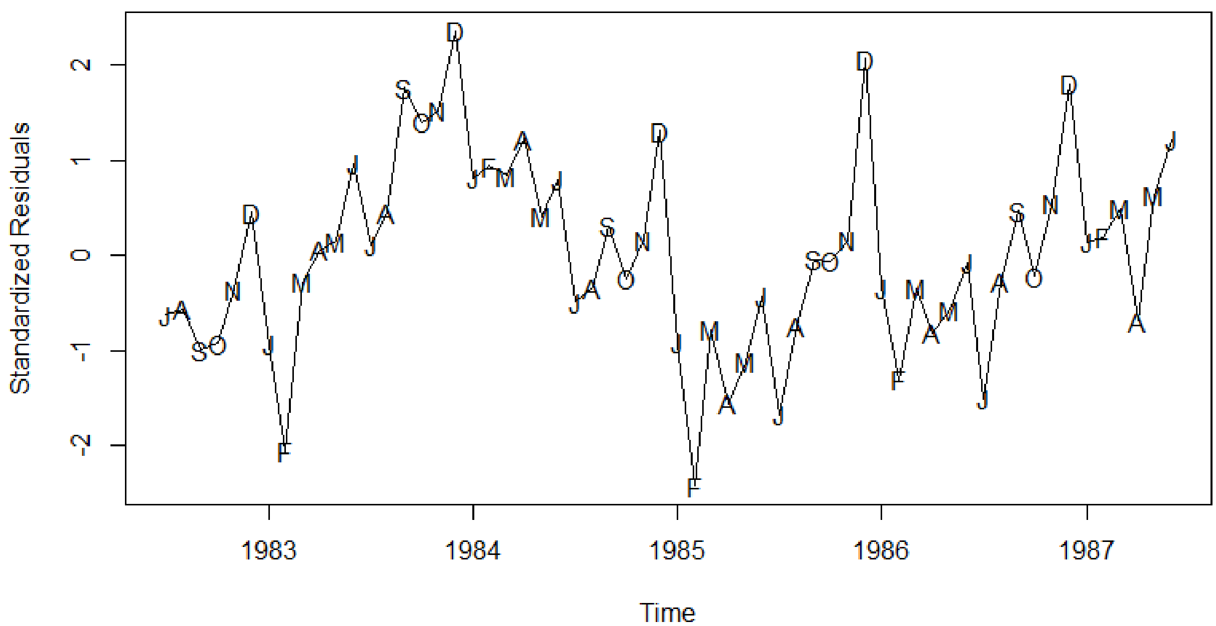
Apabila dilakukan pemodelan least square untuk *trend* kuadratik terhadap waktu, diperoleh output berikut.

```
Call:
lm(formula = hours ~ time(hours) + I(time(hours)^2))

Residuals:
     Min       1Q   Median       3Q      Max 
-1.00603 -0.25431 -0.02267  0.22884  0.98358 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -5.122e+05  1.155e+05  -4.433 4.28e-05 ***
time(hours)       5.159e+02  1.164e+02   4.431 4.31e-05 ***
I(time(hours)^2) -1.299e-01  2.933e-02  -4.428 4.35e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.423 on 57 degrees of freedom
Multiple R-squared:  0.5921,    Adjusted R-squared:  0.5778 
F-statistic: 41.37 on 2 and 57 DF,  p-value: 7.97e-12
```

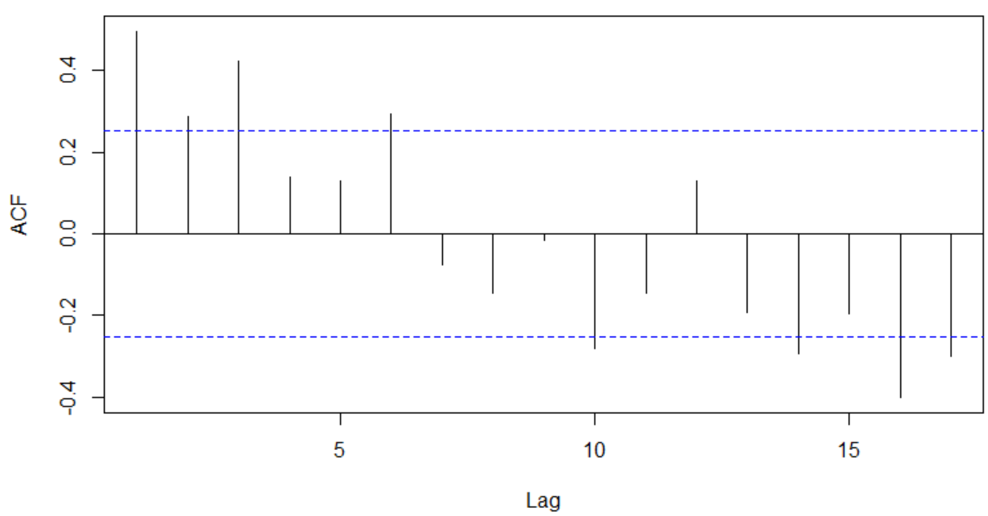
a. Dengan melihat output, tuliskan model regresinya.
b. Lakukan analisis. Apakah model pada (a) signifikan?
c. Berkaitan dengan plot data dan ‘time series’ yang sedang dipelajari, apa kelemahan model (a)? Dengan perkataan lain, pola apa yang tidak dapat terlihat jika menggunakan model pada (a)? (Hint. pertimbangkan pola musiman)

d. Dengan memerhatikan plot ke-2, tuliskan model yang menurut Anda lebih baik dari model (a) sehingga kelemahannya dapat diatasi. Berikan alasan.

e. *Run Test* dalam *Time Series* bertujuan untuk menguji apakah ada autokorelasi pada data dengan melihat kerandoman residualnya. Hipotesisnya:
H_0: *there is no autocorrelation in the errors*; dan
H_1: *there is autocorrelation*.
Berikut ini adalah hasil *run test* menggunakan R software. Apa kesimpulan dari *run test* ini?




```
$pvalue
[1] 0.00012

$observed.runs
[1] 16

$expected.runs
[1] 30.96667

$n1
[1] 31

$n2
[1] 29

$k
[1] 0
```



Jawaban

1. Misalkan $\{X_t\}$ adalah runtun waktu yang stasioner dan didefinisikan:
$$Y_t = \begin{cases} X_t & \text{, untuk } t \text{ ganjil} \\ X_t + 6 & \text{, untuk } t \text{ genap} \end{cases}$$
   a. Tentukan $\text{Cov}(Y_t, Y_{t-k})$ untuk semua lag k. Apakah bebas dari t?
   b. Apakah $\{Y_t\}$ stasioner?

Jawab:

a. WLOG misal t ganjil. Akan ditentukan $\text{Cov}(Y_t, Y_{t-k})$, $\forall k \in \mathbb{Z}$.
Pertama, perlu ditentukan terlebih dahulu nilai dari $\text{Var}(Y_t)$. Perhatikan bahwa Y_t dapat ditulis ulang sebagai $Y_t = aX_t + b$ dengan $a = 1$ dan:
$$b = \begin{cases} 0 & \text{, untuk } t \text{ ganjil} \\ 6 & \text{, untuk } t \text{ genap} \end{cases}$$
Karena $\{X_t\}$ merupakan runtun waktu stasioner dan a, b merupakan konstanta maka diperoleh $\text{Cov}(Y_t, Y_t) = \text{Var}(Y_t) = \text{Var}(X_t)$ pada lag $k = 0$.
   - Misal k ganjil, yakni $k = 2m + 1, m \in \mathbb{Z}$. Maka karena $\{X_t\}$ merupakan runtun waktu stasioner, diperoleh pada lag k ganjil:
$$\text{Cov}(Y_t, Y_{t-k}) = \text{Cov}(X_t, X_t + 6) = \text{Cov}(X_t, X_t) = \text{Var}(X_t)$$
   - Misal k genap selain nol, yakni $k = 2m, m \in \mathbb{Z}\backslash\{0\}$. Maka diperoleh pada lag k genap:
$$\text{Cov}(Y_t, Y_{t-k}) = \text{Cov}(X_t, X_t) = \text{Var}(X_t)$$

Sehingga diperoleh $\text{Cov}(Y_t, Y_{t-k}) = \text{Var}(X_t)$ untuk semua lag k.
Misal p, q adalah konstanta. Karena $\{X_t\}$ merupakan runtun waktu stasioner:
$$\text{Cov}(X_t + p, X_{t-k} + q) = \text{Cov}(X_t, X_{t-k})$$
yang bebas dari t untuk setiap k. Sehingga $\text{Cov}(Y_t, Y_{t-k}) = \text{Cov}(X_t, X_t)$ bebas dari t.

b. Perhatikan:
$$\mu_t = E(Y_t) = \begin{cases} E(X_t) & \text{, untuk } t \text{ ganjil} \\ E(X_t) + 6 & \text{, untuk } t \text{ genap} \end{cases}$$
Karena μ_t bergantung pada t, diperoleh Y_t tidak stasioner.

∴ Sehingga:

a. $\text{Cov}(Y_t, Y_{t-k}) = \text{Var}(X_t)$ bebas dari t untuk semua lag k.

b. Runtun waktu Y_t tidak stasioner.

Catatan: Soal ini merupakan variasi dari soal buku 2.6

2. Diketahui $\{X_t\}$ adalah runtun waktu stasioner dengan fungsi autokovariansi $\gamma_X(k)$. Didefinisikan runtun baru $\{Y_t\}$ yang sebagai berikut:

$$Y_t = \nabla^s X_t$$

   a. Nyatakan fungsi autokovariasi dari $\{Y_t\}$ sebagai fungsi dari $\gamma_X(k)$.
   b. Jika diberikan $X_t = \alpha X_{t-1} + e_t, |\alpha| < 1$, dimana $\{e_t\}$ adalah *white noise*, tentukan fungsi autokovariansi dari $\{Y_t\}$.
   c. Gunakan $s = 1$ yaitu $Y_t = \nabla X_t$, tentukan untuk nilai α berapa sehingga variansi Y_t akan lebih kecil dari variansi X_t.

Jawab:

a. Akan dinyatakan fungsi autokovariasi dari $\{Y_t\}$ sebagai fungsi dari $\gamma_X(k)$. Perhatikan:

Untuk $s = 1$, maka $Y_t = \nabla X_t = X_t - X_{t-1}$.

Untuk $s = 2$, maka:

$$Y_t = \nabla^2 X_t = \nabla[X_t - X_{t-1}] = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}$$

Untuk $s = 3$, maka:

$$\begin{aligned} Y_t = \nabla^3 X_t = \nabla[X_t - 2X_{t-1} + X_{t-2}] &= (X_t - X_{t-1}) - 2(X_{t-1} - X_{t-2}) + (X_{t-2} - X_{t-3}) \\ &= X_t - 3X_{t-1} + 3X_{t-2} - X_{t-3} \end{aligned}$$

dan seterusnya.

Dapat diobservasi bahwa Y_t mengikuti pola sebagai berikut:

$$Y_t = \nabla^s X_t = \sum_{i=0}^{s} (-1)^i \binom{s}{i} X_{t-i}$$

Misal $c_i = (-1)^i \binom{s}{i}$ maka $Y_t = c_0 X_t + c_1 X_{t-1} + \cdots + c_s X_{t-s}$. Diperoleh:

$$\begin{aligned} \gamma_Y(k) &= \text{Cov}(Y_t, Y_{t-k}) \\ &= \text{Cov}(\nabla^s X_t, \nabla^s X_{t-k}) \\ &= \text{Cov}(c_0 X_t + c_1 X_{t-1} + \cdots + c_s X_{t-s}, c_0 X_{t-k} + c_1 X_{t-k-1} + \cdots + c_s X_{t-k-s}) \\ &= \sum_{p=0}^{s} \sum_{q=0}^{s} c_p c_q \text{Cov}\big(X_{t-q}, X_{t-k-p}\big) \\ &= \sum_{p=0}^{s} \sum_{q=0}^{s} (-1)^p \binom{s}{p} (-1)^q \binom{s}{q} \text{Cov}\big(X_{t-q}, X_{t-q+(q-k-p)}\big) \\ &= \sum_{p=0}^{s} \sum_{q=0}^{s} (-1)^{p+q} \binom{s}{p} \binom{s}{q} \gamma_X(q-k-p) \\ &= \sum_{p=0}^{s} \sum_{q=0}^{s} (-1)^{p+q} \binom{s}{p} \binom{s}{q} \gamma_X(k+p-q) \end{aligned}$$

b. Diberikan X_t adalah model AR(1), yakni $X_t = \alpha X_{t-1} + e_t$ dengan $|\alpha| < 1$ dimana $\{e_t\}$ adalah *white noise*. Akan ditentukan fungsi autokovariansi dari $\{Y_t\}, Y_t = \nabla^s X_t$.
Karena X_t adalah model AR(1), diperoleh:

$$\gamma_X(k) = \alpha^k \frac{\operatorname{Var}(e_t)}{1-\alpha^2} = \alpha^k \frac{\sigma_e^2}{1-\alpha^2}$$

Dari bagian a dengan substitusi tersebut, diperoleh:

$$\begin{aligned}\gamma_Y(k) &= \sum_{p=0}^{s}\sum_{q=0}^{s}(-1)^{p+q}\binom{s}{p}\binom{s}{q}\gamma_X(k+p-q) \\ &= \sum_{p=0}^{s}\sum_{q=0}^{s}(-1)^{p+q}\binom{s}{p}\binom{s}{q}\alpha^{k+p-q}\frac{\sigma_e^2}{1-\alpha^2}\end{aligned}$$

c. Asumsikan X_t masih model yang sama dengan bagian b. Asumsi ini diperlukan karena berhubungan dengan fungsi autokovariansi $\gamma_X(k)$ yang terkait.
Dengan $s = 1$ maka $Y_t = \nabla X_t$. Akan ditentukan $\alpha \ni \operatorname{Var}(Y_t) < \operatorname{Var}(X_t)$.
Dari bagian b dengan $s = 1$, diperoleh autokovariansi untuk $\{Y_t\}$:

$$\begin{aligned}\gamma_Y(k) = \operatorname{Cov}(Y_t, Y_{t-k}) &= \sum_{p=0}^{1}\sum_{q=0}^{1}(-1)^{p+q}\binom{1}{p}\binom{1}{q}\alpha^{k+p-q}\frac{\sigma_e^2}{1-\alpha^2} \\ &= \sum_{p=0}^{1}\sum_{q=0}^{1}(-1)^{p+q}\alpha^{k+p-q}\frac{\sigma_e^2}{1-\alpha^2}\end{aligned}$$

Sedangkan untuk $\{X_t\}$:

$$\gamma_X(k) = \operatorname{Cov}(X_t, X_{t-k}) = \alpha^k \frac{\sigma_e^2}{1-\alpha^2}$$

Tetapkan $k = 0$, maka diperoleh:

$$\operatorname{Cov}(Y_t, Y_t) = \operatorname{Var}(Y_t) = \sum_{p=0}^{1}\sum_{q=0}^{1}(-1)^{p+q}\alpha^{p-q}\frac{\sigma_e^2}{1-\alpha^2}, \quad \operatorname{Cov}(X_t, X_t) = \operatorname{Var}(X_t) = \frac{\sigma_e^2}{1-\alpha^2}$$

Sehingga:

$$\begin{aligned}\operatorname{Var}(Y_t) &= \sum_{p=0}^{1}\sum_{q=0}^{1}(-1)^{p+q}\alpha^{p-q}\operatorname{Var}(X_t) \\ &= \sum_{p=0}^{1}\left[(-1)^p\alpha^p\operatorname{Var}(X_t) + (-1)^{p+1}\alpha^{p-1}\operatorname{Var}(X_t)\right] \\ &= \left[\operatorname{Var}(X_t) - \frac{1}{\alpha}\operatorname{Var}(X_t)\right] + \left[-\alpha\operatorname{Var}(X_t) + \operatorname{Var}(X_t)\right] \\ &= -\frac{(\alpha-1)^2}{\alpha}\operatorname{Var}(X_t)\end{aligned}$$

Perhatikan:

$$\mathrm{Var}(Y_t) < \mathrm{Var}(X_t) \Leftrightarrow -\frac{(\alpha-1)^2}{\alpha}\mathrm{Var}(X_t) < \mathrm{Var}(X_t) \Leftrightarrow -\frac{(\alpha-1)^2}{\alpha} < 1$$
$$\Leftrightarrow \alpha^2 - 2\alpha + 1 > -\alpha \Leftrightarrow \alpha^2 - \alpha + 1 > 0$$

Misal $f(\alpha) = \alpha^2 - \alpha + 1, \alpha \in \mathbb{R}$. Fungsi kuadrat f merupakan fungsi konveks ke atas karena koefisien α^2 positif. Sehingga minimum global dari f dapat diperoleh dengan menyelesaikan $f'(\alpha) = 0$ untuk α.

$$f'(\alpha) = 0 \Leftrightarrow 2\alpha - 1 = 0 \Leftrightarrow \alpha = \frac{1}{2}$$

Substitusikan ke f, diperoleh:

$$f\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^2 - \frac{1}{2} + 1 = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

Sehingga diperoleh:

$$\alpha^2 - \alpha + 1 \geq \frac{3}{4}, \forall\alpha \in \mathbb{R}$$

Artinya untuk sembarang $|\alpha| < 1$, diperoleh:

$$\alpha^2 - \alpha + 1 \geq \frac{3}{4} > 0$$

yang berarti $\mathrm{Var}(Y_t) < \mathrm{Var}(X_t), \forall\alpha, |\alpha| < 1$.

∴ Sehingga:

a. Untuk runtun waktu $\{Y_t\}$ dengan $Y_t = \nabla^s X_t$,

$$\gamma_Y(k) = \sum_{p=0}^{s}\sum_{q=0}^{s}(-1)^{p+q}\binom{s}{p}\binom{s}{q}\gamma_X(k+p-q)$$

b. Jika $X_t = \alpha X_{t-1} + e_t, |\alpha| < 1$, maka:

$$\gamma_Y(k) = \sum_{p=0}^{s}\sum_{q=0}^{s}(-1)^{p+q}\binom{s}{p}\binom{s}{q}\alpha^{k+p-q}\frac{\sigma_e^2}{1-\alpha^2}$$

c. Jika $s = 1$, maka $\mathrm{Var}(Y_t) < \mathrm{Var}(X_t)$ berlaku untuk setiap α dengan $|\alpha| < 1$.

3. Perhatikan dua model runtun waktu berikut:
   A: $Y_t = 0.9Y_{t-1} + 0.09Y_{t-2} + e_t$
   B: $Y_t = Y_{t-1} + e_t - 0.1e_{t-1}$
   a. Identifikasi masing-masing model sebagai model ARIMA yaitu berapa p, d, dan q dan nilai dari masing-masing parameter ϕ dan θ?
   b. Dalam hal apa kedua model itu mirip? Jelaskan.
   c. Dalam hal apa kedua model itu berbeda? Jelaskan.

Jawab:

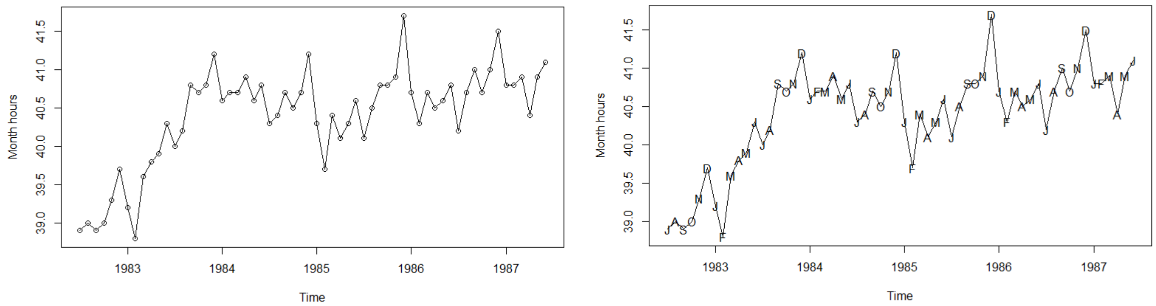
a. Akan diidentifikasi model A dan B sebagai model ARIMA(p, d, q) serta nilai dari masing-masing parameter ϕ dan θ.
   - Untuk model A yakni $Y_t = 0.9Y_{t-1} + 0.09Y_{t-2} + e_t$, perhatikan bahwa proses ini memiliki 2 suku AR dan tidak memiliki suku MA. Diperoleh $\phi_1 = 0.9$ dan $\phi_2 = 0.09$. Perhatikan:
   $$\phi_1 + \phi_2 = 0.9 + 0.09 = 0.99 < 1$$
   $$\phi_2 - \phi_1 = 0.09 - 0.9 = -0.81 < 1$$
   $$|\phi_2| = |0.09| = 0.09 < 1$$
   Artinya proses ini stasioner. Model A adalah AR(2) atau ARIMA(2,0,0). Parameter-paramater model yaitu $\phi_1 = 0.9$ dan $\phi_2 = 0.09$.
   - Untuk model B yakni $Y_t = Y_{t-1} + e_t - 0.1e_{t-1}$, perhatikan bahwa proses ini memiliki 1 suku AR dan 1 suku MA. Diperoleh $\phi_1 = 1$ dan $\theta_1 = -0.1$. Karena $\phi_1 \geq 1$ maka suku AR ini membuat model tidak stasioner. Tulis ulang:
   $$Y_t = Y_{t-1} + e_t - 0.1e_{t-1} \Leftrightarrow Y_t - Y_{t-1} = e_t - 0.1e_{t-1} \Leftrightarrow W_t = e_t - 0.1e_{t-1}$$
   dengan $W_t = \nabla Y_t = Y_t - Y_{t-1}$. Sehingga diperoleh model:
   $$W_t = e_t - 0.1e_{t-1}$$
   Proses ini tidak memiliki suku AR dan memiliki 1 suku MA. Diperoleh $\theta_1 = -0.1$. Dengan 1 kali differencing, diperoleh proses stasioner. Model B adalah IMA(1,1) atau ARIMA(0,1,1). Parameter model yaitu $\theta_1 = -0.1$.

b. Model A dan B mirip dalam hal model menginkorporasikan suku autoregresif Y_{t-1} dan suku *white noise* e_t.

c. Model A dan B berbeda karena model A hanya memiliki suku AR sedangkan model B hanya tidak memiliki suku AR. Perbedaan yang lain adalah model B hanya stasioner jika dilakukan differencing sedangkan model A sudah stasioner.

Catatan: Soal ini merupakan soal buku 5.7

4. Berikut ini adalah plot mengenai rata-rata jumlah jam kerja per bulan dari buruh suatu pabrik dari Juli 1982 – Juni 1987.

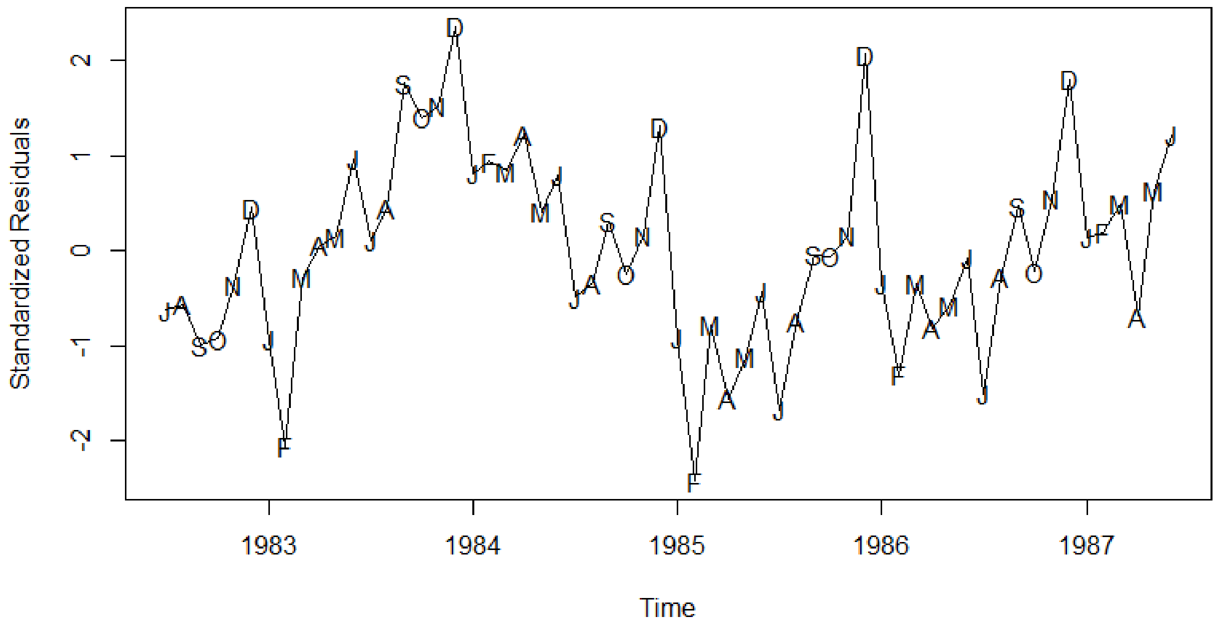


Apabila dilakukan pemodelan least square untuk *trend* kuadratik terhadap waktu, diperoleh output berikut.

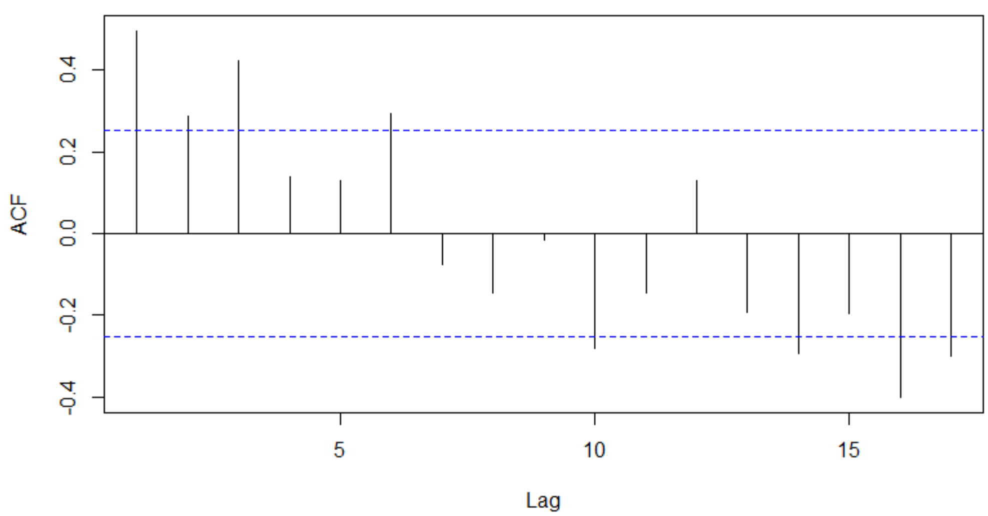
```
Call:
lm(formula = hours ~ time(hours) + I(time(hours)^2))

Residuals:
     Min       1Q   Median       3Q      Max 
-1.00603 -0.25431 -0.02267  0.22884  0.98358 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -5.122e+05  1.155e+05  -4.433 4.28e-05 ***
time(hours)       5.159e+02  1.164e+02   4.431 4.31e-05 ***
I(time(hours)^2) -1.299e-01  2.933e-02  -4.428 4.35e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.423 on 57 degrees of freedom
Multiple R-squared:  0.5921,	Adjusted R-squared:  0.5778 
F-statistic: 41.37 on 2 and 57 DF,  p-value: 7.97e-12
```

a. Dengan melihat output, tuliskan model regresinya.
b. Lakukan analisis. Apakah model pada (a) signifikan?
c. Berkaitan dengan plot data dan ‘time series’ yang sedang dipelajari, apa kelemahan model (a)? Dengan perkataan lain, pola apa yang tidak dapat terlihat jika menggunakan model pada (a)? (Hint. pertimbangkan pola musiman)
d. Dengan memerhatikan plot ke-2, tuliskan model yang menurut Anda lebih baik dari model (a) sehingga kelemahannya dapat diatasi. Berikan alasan.
e. *Run Test* dalam *Time Series* bertujuan untuk menguji apakah ada autokorelasi pada data dengan melihat kerandoman residualnya. Hipotesisnya:

H_0: *there is no autocorrelation in the errors*; dan
H_1: *there is autocorrelation*.
Berikut ini adalah hasil *run test* menggunakan R software. Apa kesimpulan dari *run test* ini?




```
$pvalue
[1] 0.00012

$observed.runs
[1] 16

$expected.runs
[1] 30.96667

$n1
[1] 31

$n2
[1] 29

$k
[1] 0
```



Jawab:

a. Berdasarkan output dari pemodelan least square pada software R, diperoleh model regresi (setelah pembulatan) sebagai berikut:

$$\hat{y} = \beta_0 + \beta_1 t + \beta_2 t^2 = -5.122(10^5) + 5.159(10^2)t - 1.299(10^{-1})t^2$$
$$= -512200 + 515.9t - 0.1299t^2$$

dimana t menyatakan variabel *hours*.

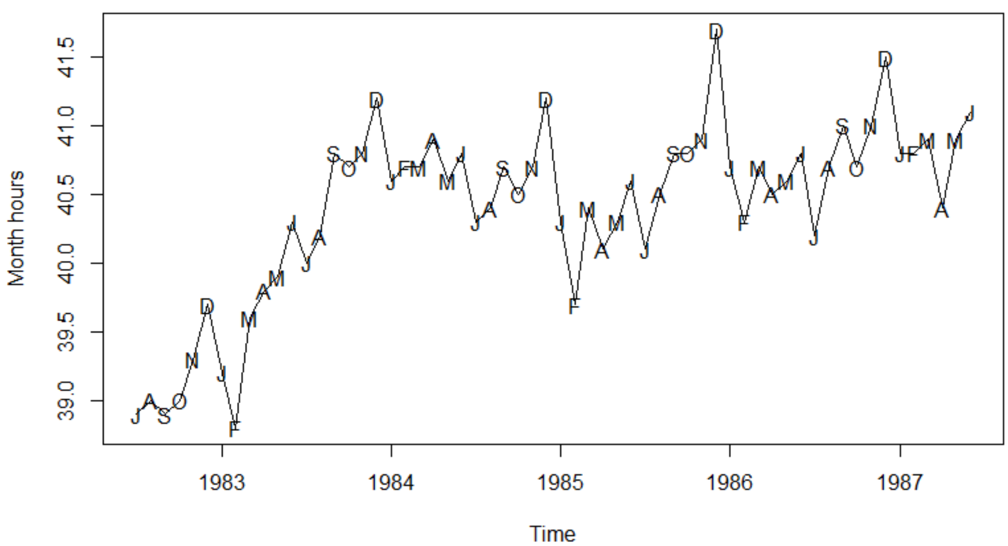
b. Diperoleh model hasil fitting dari bagian a sebagai berikut:

$$\hat{y} = -512200 + 515.9t - 0.1299t^2$$

Berdasarkan output dari pemodelan least square yang sama, diperoleh bahwa model miliki adj R-squared 0.5778 yang menandakan bahwa model kurang cocok dengan data. Walau demikian, semua variabel yang digunakan berguna dengan kepercayaan 99% karena $\Pr(>|t|) < 0.01$ untuk setiap variabel. Artinya model sedikit menangkap pola data namun tidak menggeneralisasi dengan baik. Diperlukan model yang memerhatikan pola data dengan lebih baik.

c. Kelemahan model pada bagian a seperti yang sudah dijelaskan pada bagian b adalah kekurangan dalam menangkap pola yang terkandung dalam data. Ini dikarenakan data merupakan data berbentuk runtun waktu, dimana setiap titik data berkorespondensi dengan suatu waktu, yakni tahun dan bulan pada kasus ini. Umumnya data dengan bentuk ini memiliki faktor autokorelasi yang tidak dipertimbangkan pada model regresi. Jika kita perhatikan pada grafik, data *hours* ini memiliki suatu pola musiman seperti rata-rata menurun dari bulan Januari ke Februari (terlihat jelas pada 1983, 1985, dan 1986). Pola musiman ini muncul dalam beberapa kasus yang dapat ditentukan secara numerik tidak hanya dengan observasi menggunakan analisis runtun waktu. Sehingga inilah yang tidak terlihat oleh model pada bagian a, karena model bagian a merupakan model regresi yang tidak memertimbangkan adanya autokorelasi.

d. Untuk menentukan model yang baik untuk data *hours* berdasarkan grafik yang tersedia, perlu dipertimbangkan 3 hal yakni *trend, season,* dan *cycle*. Berikut grafik data *hours*.



- Sebuah *trend* ada jika tampak kenaikan atau penurunan jangka panjang dalam data. Terlihat bahwa data mengalami kenaikan walau tidak signifikan. Anggap bahwa ada kemungkinan bahwa *trend* naik ada pada data dengan periode observasi ini.

- Sebuah *season* ada jika data terpengaruh oleh faktor musiman seperti bulan dalam suatu tahun. Seperti yang sudah didiskusikan pada bagian c, data ini memiliki faktor *season*.
- Sebuah *cycle* ada jika tertunjukkan kenaikan dan penurunan yang tidak memiliki frekuensi tetap. Terlihat bahwa antara suatu tahun dan tahun yang lain, dari suatu bulan ke bulan yang lain tidak terdapat frekuensi tetap yang memengaruhi kenaikan maupun penurunan. Sehingga tidak ada *cycle* pada data ini.

Sehingga model yang baik untuk data ini adalah model yang memertimbangkan faktor *season* dan faktor *trend* jika merupakan faktor signifikan. Faktor *trend* dapat diuji misal dengan menggunakan tes non-parametrik Spearman. Dengan menggunakan R diperoleh koefisien Spearman (setelah pembulatan) pada data hours yang sudah distandarisasi residualnya sebesar 0.0322. Artinya faktor *trend* tidak signifikan.

Sehingga model yang baik untuk data ini adalah model yang memertimbangkan faktor *season*. Model yang dapat digunakan diantaranya model Seasonal Naïve dan model Seasonal ARIMA. Sehingga dengan alasan-alasan tersebut, kelemahan pada model bagian a dapat teratasi karena faktor musiman dalam data sudah diinkorporasikan dalam data.

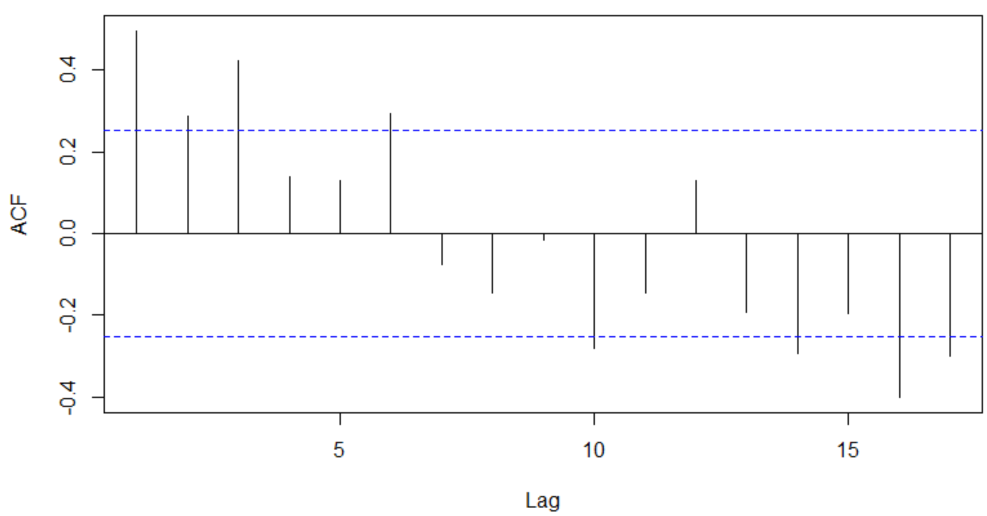
e. Diberikan hasil *run test* menggunakan R untuk menguji hipotesis adanya autokorelasi berdasarkan kerandoman residual,

   H_0: *there is no autocorrelation in the errors*; dan

   H_1: *there is autocorrelation*.

   sebagai berikut:

```
$pvalue
[1] 0.00012

$observed.runs
[1] 16

$expected.runs
[1] 30.96667

$n1
[1] 31

$n2
[1] 29

$k
[1] 0
```



Dari *run test* tersebut, perhatikan bahwa observed.runs kurang dari expected.runs. Ini mengindikasikan bahwa nilai-nilai yang bertetangga memiliki dependensi secara positif dan cenderung untuk dekat atau serupa dari waktu ke waktu. Sehingga H_0 ditolak dan dapat disimpulkan terdapat autokorelasi pada data *hours*.

∴ Sehingga:

a. Model regresi berdasarkan *trend* kuadratik terhadap waktu diberikan oleh
$$\hat{y} = -512200 + 515.9t - 0.1299t^2$$
dimana t menyatakan variabel *hours*.

b. Model pada bagian a tidak signifikan karena kurang menangkap pola pada data.

c. Kelemahan model pada bagian a adalah tidak memerhatikan adanya faktor autokorelasi karena data berbentuk runtun waktu. Pola yang tidak terlihat adalah pola musiman.

d. Model yang lebih baik daripada model pada bagian a adalah model yang memerhatikan faktor musim, seperti model Seasonal Naïve atau Seasonal ARIMA. Dengan ini, kelemahan model sebelumnya teratasi.

e. Berdasarkan hasil *run test*, disimpulkan bahwa terdapat autokorelasi pada data *hours*.

REFERENSI

[1] Cryer, J.D. dan Chan K-S. (2008). *Time Series Analysis: With Applications in R*. ed 2. Springer Text in Statistics.

[2] Hyndman, R.J. dan Athanasopoulos G. (2018). *Forecasting: principles and practice*. ed 2. Diakses pada 19 Oktober 2019. ⟨ otexts.com/fpp2/ ⟩

[3] Larsson, J. (2017). *Solutions to Time Series Analysis: with Applications in R*. Diakses pada 18 Oktober 2019. ⟨ jolars.github.io/TSAsolutions/index.html ⟩

[4] Mudit. (2016). *Time series Analysis Chapter 3: Trends*. Diakses pada 18 Oktober 2019. ⟨ rpubs.com/mudy/161970 ⟩

[5] R Package Documentation. (2019). *runs: Runs test*. Diakses pada 19 Oktober 2019. ⟨ rdrr.io/cran/TSA/man/runs.html ⟩

AFTERWORD

- Soal asli yang disediakan oleh Departemen Radian HMD Matematika terlalu buram sehingga grafik plot disintesis sendiri oleh penulis menggunakan R.
- Apabila request untuk kode R cukup banyak, penulis mungkin memberikan link kode terkait pada versi berikutnya.
- ‘buku’ yang dimaksud pada catatan nomor 1 dan 3 mengacu pada referensi [1].