Pengenalan

Analisis Deret Waktu (*Time Series Analysis*)

MA 2081 Statistika Dasar

29 November 2012

Utriweni Mukhaiyar

Ilustrasi

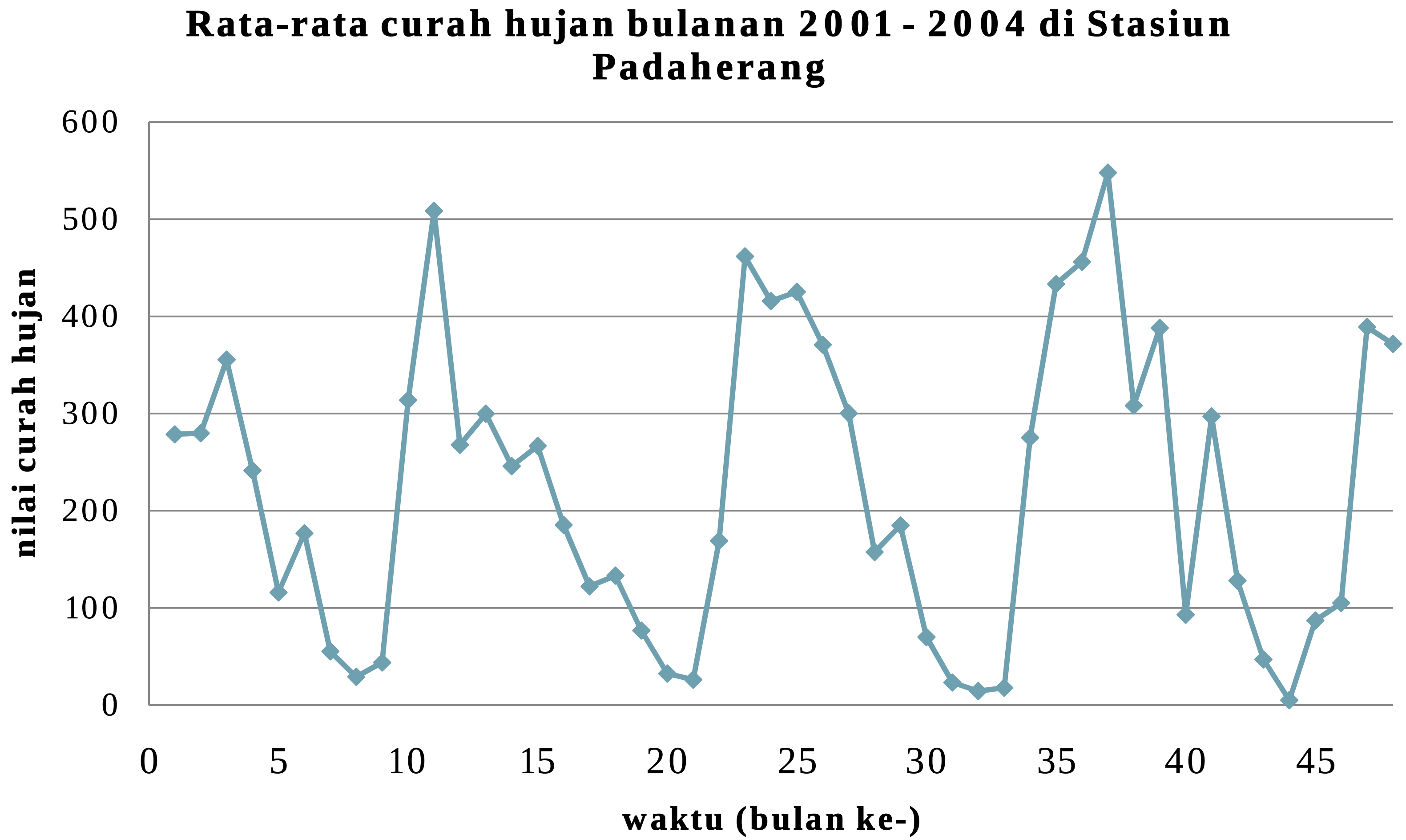
- Berikut adalah data rata-rata curah hujan bulanan yang diamati dari Stasiun Padaherang pada tahun 2001 – 2004.

Sumber : Modul 3 Praktikum Mekanika Medium Kontinu “ Medan Gravitasi”

Tahun	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Agust	Sep	Okt	Nop	Des
2001	278.59	279.78	355.29	241.34	115.9	176.9	55.32	29.08	43.82	313.68	508.49	267.82
2002	299.78	245.88	266.64	185.27	122.22	133.1	76.78	32.4	26.09	169.05	461.62	415.73
2003	425.21	370.8	300.23	157.43	184.96	69.93	23.28	14.39	17.86	275.23	433.23	456.02
2004	547.8	308.2	388	93	297	128	47	5	87	105	389	371.6

- Apabila nilai curah hujan saat ini dianggap dipengaruhi oleh rata-rata curah hujan kemarin dst, maka data rata-rata curah hujan di atas dapat dikategorikan sebagai suatu deret waktu (*time series*).

Plot Data
berdasarkan waktu



Proses Stokastik

- Proses stokastik adalah barisan peubah acak $\{Y_t, t \in T\}$
- Setiap proses stokastik memuat ruang keadaan S dan indeks parameter T
  S : semua nilai yang mungkin dari Y_t
  S dan T dapat bernilai diskrit atau kontinu
- Contoh proses stokastik:
  a. Cuaca harian kota Bandung
  b. Banyaknya trombosit/hari pasien demam berdarah sejak ia terinfeksi
  c. Laju pertumbuhan populasi orang utan (% per tahun)
  d. Waktu antara mekarnya bunga bangkai yang ke-n dengan bunga bangkai yang ke n+1
- Misal y_t nilai dari Y_t maka barisan nilai $\{y_t, t \in T\}$ disebut realisasi dari $\{Y_t, t \in T\}$

Time Series

- Jika T : waktu, maka $\{Y_t\ ,\ t \in T\}$ disebut time series
- Realisasinya disebut data TS
- Studi berkaitan dengan TS disebut analisis TS
- Permasalahan dalam analisis TS :
  "Bagaimana menentukan model Y_t sehingga model tersebut dapat digunakan untuk forecasting (prakiraan di waktu mendatang)*?? "*

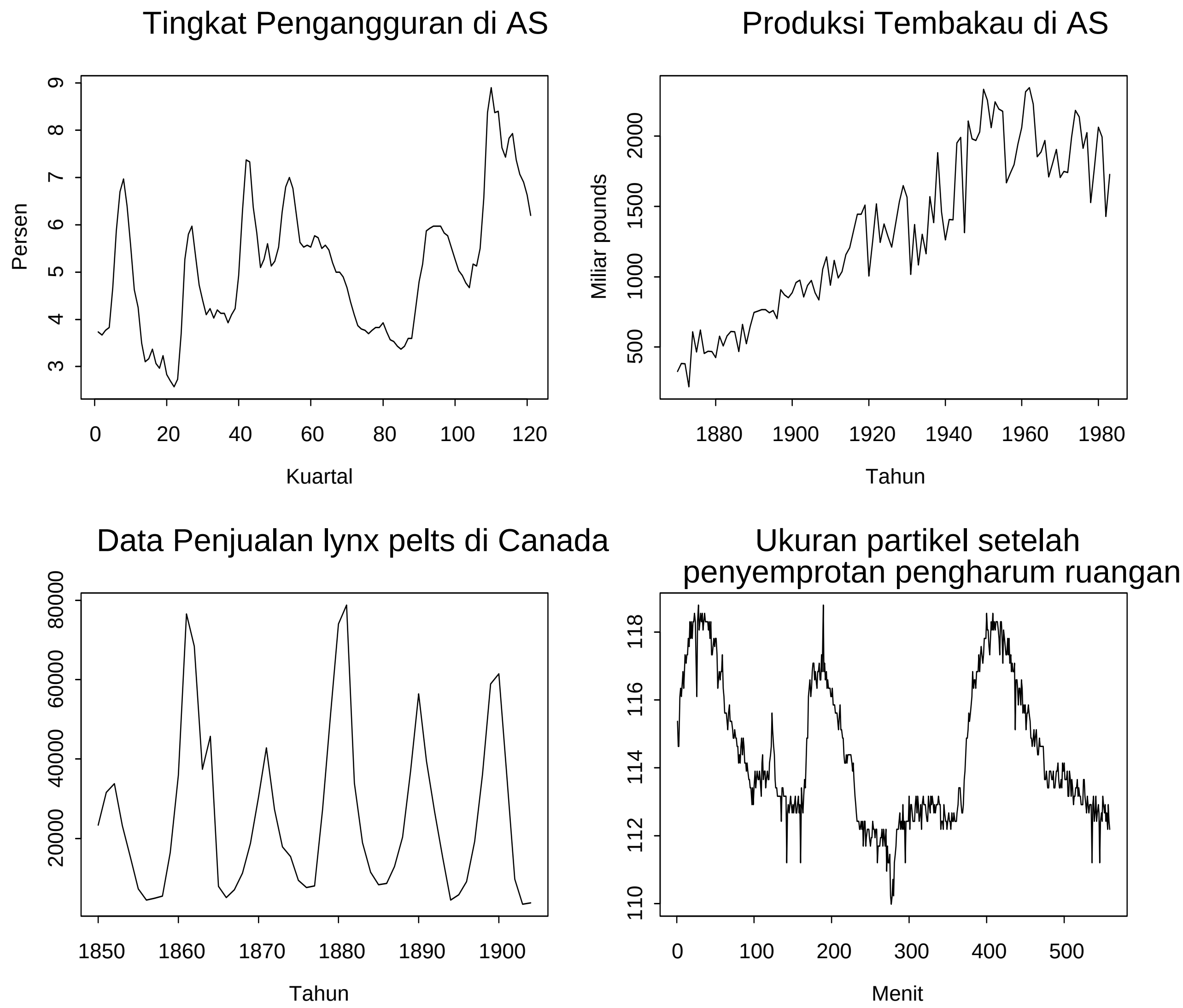
- Secara umum, model TS dapat ditulis

$$Y_t = f\,(.) + e_t \qquad (1)$$

  Asumsi galat: $e_t \sim N\,(0,\sigma^2)$ dan tidak berkorelasi

- Jika f linier dalam parameter-parameternya maka persamaan (1) disebut model linier TS
- Koleksi semua model linier TS dinamakan model ARIMA(p,d,q) (Box-Jenkins, 1976)

Contoh *Time Series*



Manfaat dan Tujuan TS

- Memodelkan data TS sehingga dapat dilihat perilaku data lebih lanjut
- Melakukan prediksi ke depan atau prakiraan jangka pendek *(short-time forecasting)*

Beberapa Konsep Dasar dalam TS
Kestasioneran

- TS $\{Y_t, t \in T\}$ stasioner jika untuk setiap t,
  1. $E[Y_t] = \mu$ (konstan)
  2. $\text{kov}(Y_t, Y_{t-k}) = \gamma_k$ (tidak tergantung t)

- Secara visual, data TS $\{Y_t, t \in T\}$ stasioner jika data TS berfluktuasi di sekitar rataannya dengan variansi konstan

Beberapa Konsep Dasar dalam TS

ACF, fungsi autokorelasi

- ACF (fungsi autokorelasi) : fungsi antara lag k dan ρ_k dengan, $\rho_k = corr\,(Y_t, Y_{t-k})$.

  ACF sampel:

$$r_k = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

  $r_k = 0$ (secara signifikan) jika

$$-1{,}96\frac{1}{\sqrt{n}} < r_k < 1{,}96\frac{1}{\sqrt{n}}$$

Beberapa Konsep Dasar dalam TS

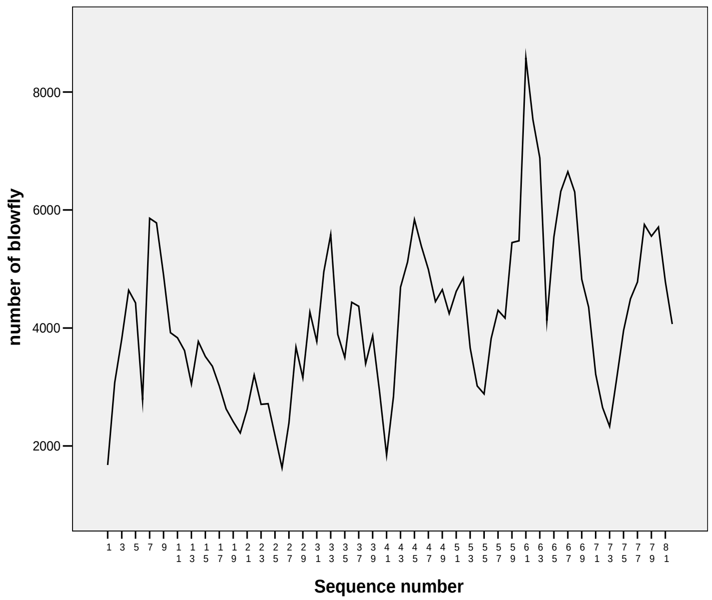
PACF, fungsi parsial autokorelasi

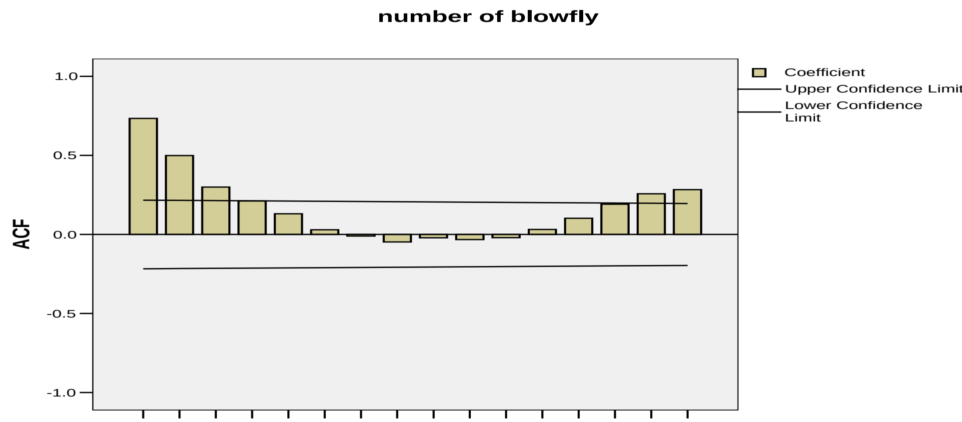
- PACF (fs. autokorelasi parsial) : fungsi antara lag k dengan ϕ_{kk} di mana $\phi_{kk} = corr\,(Y_t, Y_{t-k})$ setelah pengaruh $Y_1, Y_2, \ldots, Y_{k-1}$ ditiadakan.

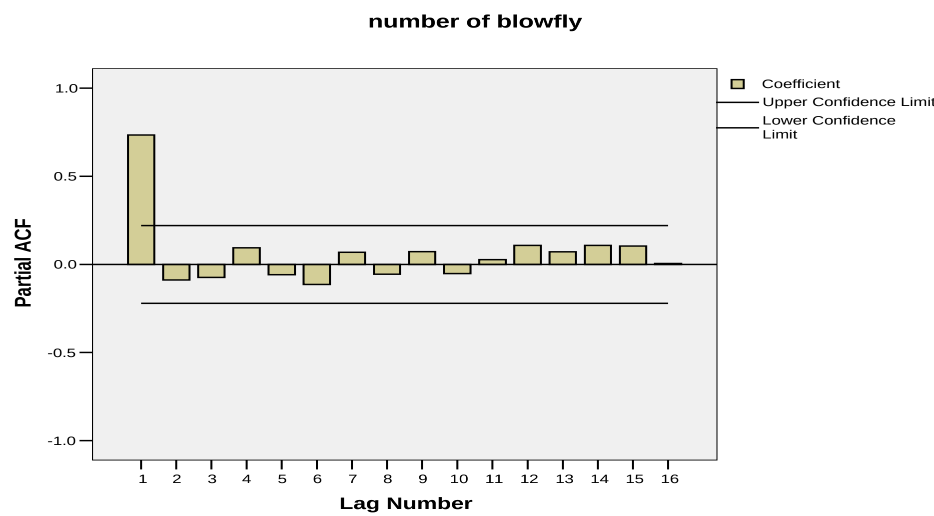
- PACF dapat didefinisikan juga sebagai koefiesien suku terakhir dari regresi Y_t dengan $Y_1, Y_2, \ldots, Y_k$.

  Artinya, jika $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_k Y_{t-k}$ maka PACF sampel untuk lag k = taksiran dari ϕ_k.
  atau $\hat{\phi}_{kk} = \hat{\phi}_k$
  $\hat{\phi}_{kk} = 0$ (secara signifikan) jika $-1{,}96\frac{1}{\sqrt{n}} < \hat{\phi}_{kk} < 1{,}96\frac{1}{\sqrt{n}}$

Contoh ACF dan PACF dengan SPSS







Dari menu SPSS, pilih
Graphs
Time Series
Autocorrelations...
pilih variabel yang akan dihitung ACF dan PACF-nya
OK

Model-model *Time Series*

Untuk TS Stasioner

1. Autoregresi (AR) : *"regresi terhadap TS yg lalu & galat sekarang"*

   AR(1): $Y_t = \mu + \phi_1 Y_{t-1} + e_t$, di mana $1 < \phi_1 < 1$

   AR(2): $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$,

   di mana $1 < \phi_2 < 1$, $\phi_2 + \phi_1 < 1$, $\phi_2 - \phi_1 < 1$

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

2. Moving Average (MA) : *"regresi terhadap galat yang lalu dan galat sekarang"*

   MA(1): $Y_t = \mu + e_t - \theta_1 e_{t-1}$, di mana $1 < \theta_1 < 1$

   MA(2): $Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$

   di mana $1 < \phi_2 < 1$, $\phi_2 + \phi_1 < 1$, $\phi_2 - \phi_1 < 1$

   MA(q): $Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}$

Model-model *Time Series*

Untuk TS Stasioner

3. Autoregresi-Moving Average (ARMA)

"regresi terhadap TS yang lalu dan semua galat"

ARMA(1,1): $Y_t = \mu + \phi_1 Y_{t\text{-}1} + e_t - \theta_1 e_{t\,\text{-}1}$

ARMA(p,q):

$Z_t = \mu + (\phi_1 Y_{t\text{-}1} + \ldots + \phi_p Y_{t\text{-}p}) + (e_t - \theta_1 e_{t\,\text{-}1} - \ldots - \theta_q e_{t\,\text{-}q})$

Catatan: AR(p) = ARMA(p,0), MA(q) = ARMA(0,q)

Model-model *Time Series*

Untuk TS tidak Stasioner

- Misal TS $\{Y_t\}$ tidak stasioner.
  Buat TS baru yg stasioner, sebut $\{Z_t\}$ dengan cara diferensi, yaitu $Z_t = Y_t - Y_{t-1}$, untuk setiap t.
- Maka
  "*ARMA(p,q) untuk $\{Z_t\}$ disebut ARIMA (p,1,q) untuk $\{Z_t\}$*"

- Jika diferensi dilakukan d kali, ditulis ARIMA(p,d,q)
- *Catatan:* ARMA(p,q) = ARIMA ($p,0,q$)

Metode Box Jenkins

Tahap awal:

Pemeriksaan kestasioneran:

- Plot TS
- Jika stasioner, lanjutkan ke "*tiga tahap iteratif*". Jika tidak lakukan transformasi atau diferensi

Tiga tahap iteratif :

1. Identifikasi
2. Penaksiran parameter
3. Uji diagnostik (pemeriksaan asumsi sisa)

Jika pada uji diagnostik, ada asumsi yang dilanggar ulangi lagi 3 tahap iteratif

Identifikasi

Model	ACF	PACF
AR(p)	Menurun secara eksponensial atau membentuk gelombang sinus teredam	*Cut off* setelah lag p
MA(q)	*Cut off* setelah lag q	Menurun secara eksponensial atau membentuk gelombang sinus teredam

- Mengidentifikasi orde (p,q) model ARMA melalui kriteria Akaike (AIC)

$$\text{AIC} \approx n \log \quad + 2m\,, \qquad m = \#\text{ parameter}$$

- Hitung nilai AIC untuk setiap (p,q). Orde yang dipilih adalah (p,q) dengan nilai AIC yang paling kecil

Penaksiran Parameter

- Metode: - Kuadrat terkecil (untuk model AR)
  - Maksimum likelihood
  - Melard (digunakan SPSS)

- Contoh penaksiran parameter melalui SPSS
  Dari menu, pilih *Analyze* → *Forecasting* → *Create Models ...*
  Pilih nama TS sebagai *Dependent variable*
  Masukkan orde model ARIMA

Uji Diagnosis

Ingat asumsi galat: $e_t \sim N(0,\sigma^2)$ dan tidak berkorelasi

Pengujian asumsi:

Cara 1:

- Plot sisaan

  berfluktuasi di sekitar 0 $\Rightarrow$ E[e_t] = 0

  nilai sisaan di sekitar $\pm$ 1,96$\hat{\sigma}^2$ $\Rightarrow$ Var(e_t) = σ^2

- plot ACF serta plot PACF-nya

  r_k dan $\hat{\phi}_{kk}$ signifikan 0 $\Rightarrow$ sisaan "tidak berkorelasi"

Cara 2: Uji Ljung-Box

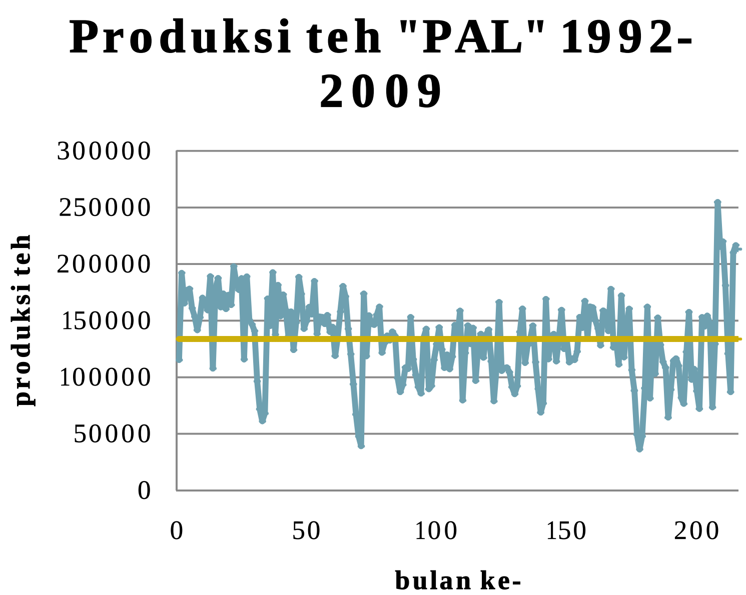
- Uji "H_0: korelasi antar sisaan = 0" dengan statistik Ljung-Box

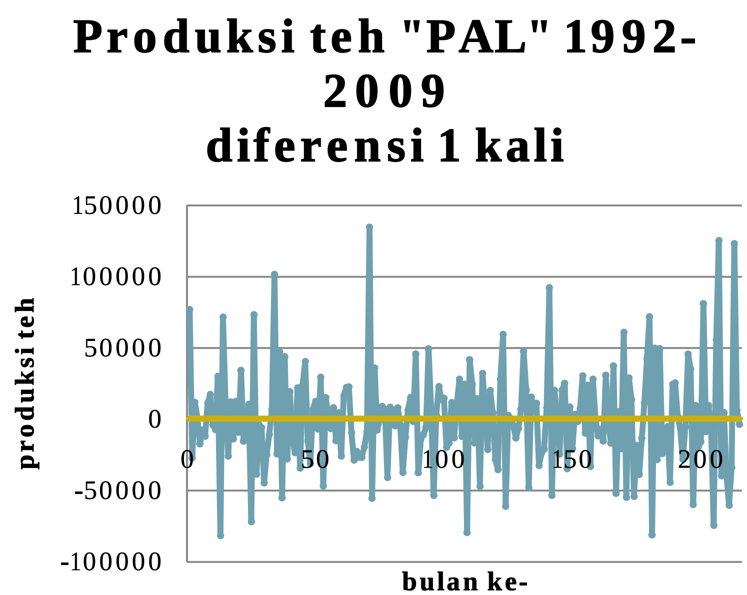
$$Q^* = n(n+2)\sum_{k=1}^{h} \frac{r_k^2}{n-k}$$

- Jika $Q^* > \chi^2_{\alpha,\nu}$ dengan $\nu = h - m$ dan m = # parameter, maka H_0 ditolak

Contoh

- Hasil produksi bulanan perkebunan teh di lokasi PAL tahun 1992-2009 (T = 216)



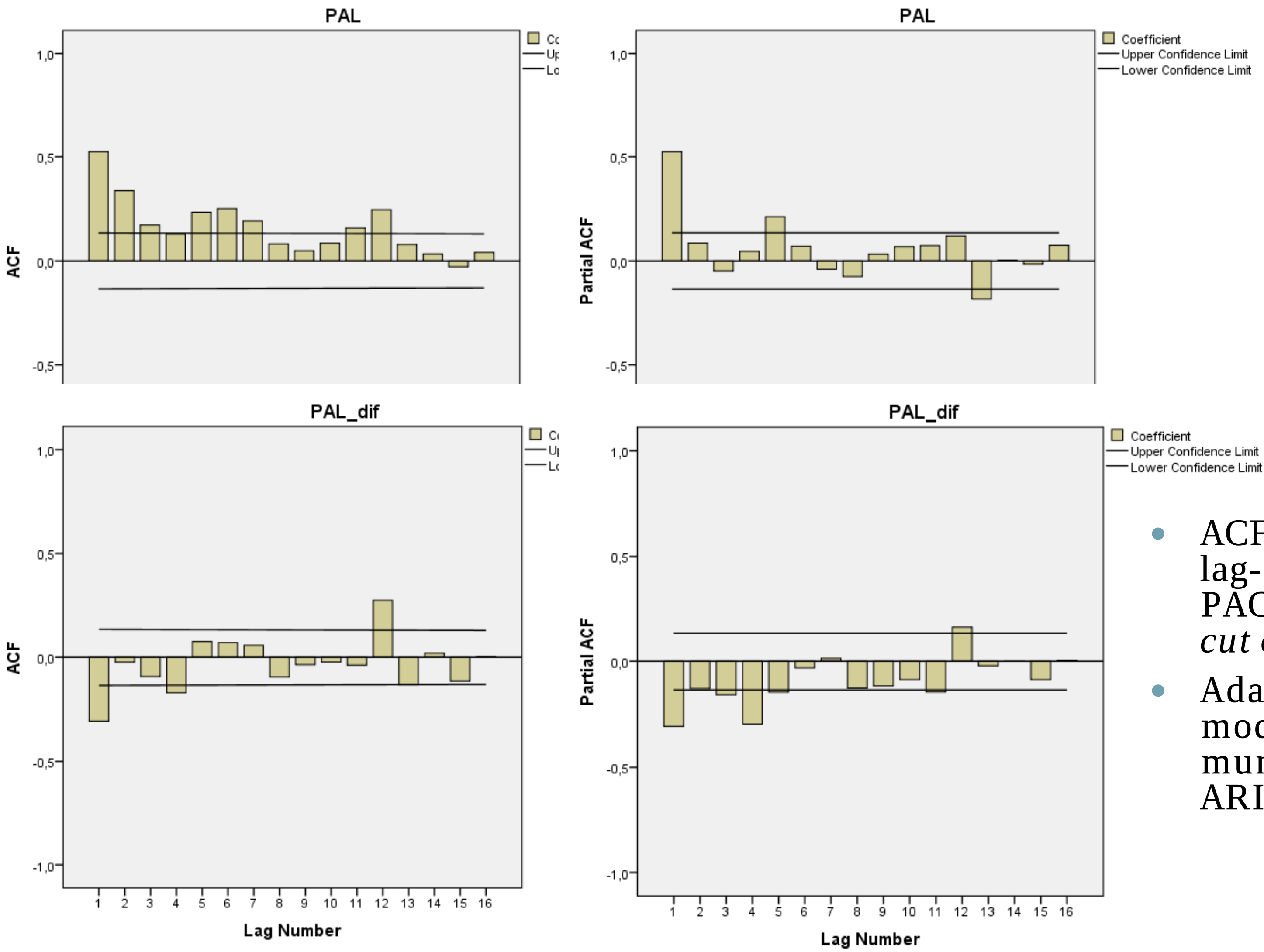


Contoh

Sari Numerik Data

Data perkebunan teh PAL	
Mean	133793.6
Standard Error	2488.531
Median	136781
Mode	#N/A
Standard Deviation	36573.79
Sample Variance	1.34E+09
Kurtosis	0.222436
Skewness	-0.07241
Range	218458
Minimum	36305
Maximum	254763
Sum	28899412
Count	216

Data perkebunan teh PAL (diff 1 kali)	
Mean	455.7023
Standard Error	2407.674
Median	-1515
Mode	-15033
Standard Deviation	35303.43
Sample Variance	1.25E+09
Kurtosis	1.855309
Skewness	0.701741
Range	216395
Minimum	-81536
Maximum	134859
Sum	97976
Count	215



F menurun
erti gelombang
us teredam
langkan PACF *cut*
setelah lag-1.
odel yang
ingkin adalah
(1)

- ACF *cut off* setelah lag-1 sedangkan PACF juga seperti *cut off* setelah lag-1.
- Ada beberapa model yang mungkin, seperti ARIMA(1,1,1)

AR (1)

ARIMA Model Parameters

					Estimate	SE	t	Sig.
PAL-Model_1	PAL	No Transformation	Constant		134113,420	4518,793	29,679	,000
			AR	Lag 1	,535	,059	9,127	,000

Model Statistics

Model	Number of Predictors	Model Fit statistics			Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	MAPE	Normalized BIC	Statistics	DF	Sig.	
PAL-Model_1	0	,281	21,160	20,738	35,952	17	,005	0

Diperoleh AR(1) : $Y_t = 134113,420 + 0,535Y_{t-1} + e_t$

ARIMA (1,1,1)

ARIMA Model Parameters

					Estimate	SE	t	Sig.
PAL-Model_1	PAL	No Transformation	Constant		-19,205	279,639	-,069	,945
			AR	Lag 1	,434	,079	5,477	,000
			Difference		1			
			MA	Lag 1	,934	,039	23,750	,000

Model Statistics

Model	Number of Predictors	Model Fit statistics			Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	MAPE	Normalized BIC	Statistics	DF	Sig.	
PAL-Model_1	0	,224	21,063	20,774	30,238	16	,017	0

Diperoleh ARIMA(1,1,1) : $Z_t = -19,205 + 0,434Z_{t-1} - 0,934e_{t-1} + e_t$

Contoh

Kesimpulan

- Berdasarkan hasil Ljung-Box, dimana pada model AR(1) H_0 ditolak (sisaan berkorelasi) untuk semua $1\% \leq \alpha \leq 10\%$, sedangkan ARIMA(1,1,1) tidak ditolak untuk $\alpha < 1,7\%$.
- Oleh karena itu model ARIMA(1,1,1) bisa dianggap lebih cocok (dengan sisaan yang tidak berkorelasi) sehingga dapat digunakan untuk melakukan *short-time forecast* dengan menggunakan persamaan :

$$Z_t = -19,205 + 0,434 Z_{t-1} - 0,934 e_{t-1} + e_t$$

$$\widehat{Y}_{t+1} - Y_t = -19,205 + 0,434(Y_t - Y_{t-1}) - 0,934 e_{t-1}$$

$$\widehat{Y}_{t+1} = -19,205 + 1,434 Y_t - 0,434 Y_{t-1} - 0,934 e_{t-1}$$

Referensi

- Box, G. E. P. dan Jenkins, G. M. (1976): *Time Series Analysis: Forecasting & Control*, Holden-Day Inc., San Fransisco
- Cryer, J. D. dan Chan, K. S. (2008): *Time Series Analysis with Applications in R*, Springer, New York.