Econometrics

Ferdian Fadly

2024-02-26

Table of contents

About Me 5
Education 5
Master of Statistics 5
Bachelor of Applied Statistics 5
Teaching Experience 5
STAT7055 - Introductory Statistics For Business and Finance 5
STAT1008 - Quantitative Research Methods 5
MST3106 - Panel Data Analytics 5
EPSEM4B - Econometrics (On-going) 5
Working/Research Experience 6
Statistician 6
Research Fellow 6
Research Fellow 6
Committee of Department of Research 6
Disclaimer : This Module is currently under development. Please provide Feedback and Suggestions. 6

1 Pendahuluan 7
1.1 Pengenalan terhadap Model 7
1.1.1 Definisi Model 7
1.1.2 Tujuan Model: 7
1.2 Trade-off Modelling 7
1.3 Jenis Model 8
1.4 Ekonometrika 8
1.4.1 Definisi Ekonometrika 8
1.4.2 Tujuan Ekonometrika 8
1.5 Tahapan Ekonometrika 8
1.6 Diskusi: Spesifikasi Model (termasuk Identifikasi Masalah) 9
1.6.1 Tips Mencari Isu/Masalah 9
1.6.2 Isu/Permasalahan, Kendaraan, dan Merk Kendaraan 9
1.7 Challenge 1 9

2 Regresi Linear (Ordinary Least Squares) 10
Tahapan Regresi 10

2.1 Spesifikasi Model Umum 10
2.1.1 Spesifikasi Model Regresi OLS 11
2.1.2 Spesifikasi Model Contoh: Hubungan Pendapatan terhadap Pengeluaran Makanan 11
2.2 Estimasi Parameter dengan Regresi 12
2.2.1 Estimasi Selang Kepercayaan 14
2.3 Pengujian Hipotesis, Asumsi, dan Kelayakan Model 14
2.3.1 Overall Test 14
2.3.2 Partial Test 15
2.3.3 Pengecekan Asumsi/Diagnosa Residual-Residual Diagnostic 15
2.3.4 Kelayakan Model 19
2.3.5 Contoh Pengecekan Asumsi yang sesuai/Ideal 19
2.4 Analisis/Interpretasi/Peramalan 23
2.4.1 Peramalan 23
2.4.2 Interpretasi 23
Extras: Model Log-log 24
Extras 2: Model dengan Asumsi Homoskedastisitas yang terlanggar 26
Challenge 2 29

3 Time Series Regression (e.g. ECM) 30
3.1 Stationeritas 30
3.2 Uji Unit Root untuk Stasioneritas 37
3.2.1 AR1, Model Autoregressive Orde Pertama 37
3.2.2 ADF Test USA Funds 38
3.2.3 ADF Test USA Bonds 38
3.3 Differensiasi 40
3.4 Kointegrasi 43
3.5 Error Correction Model (ECM) 46
3.5.1 Two Steps Engle Granger 46
3.5.2 One Step ECM 49
3.5.3 ARDL 50

4 Model Persamaan Simultan 53
4.1 Definisi 53
4.2 Spesifikasi Model 53
4.2.1 Persamaan Struktural 53
4.2.2 Persamaan reduced 53
4.3 Permasalahan Identifikasi 53
4.4 Langkah-langkah dengan Contoh 1 54
4.5 Contoh 2 56

5 Dynamic Panel Data Model 60
5.1 Introduction 60

5.2 Model Specification . . . 60
5.3 R implementation for Arrelano and Bond (1991) . . . 60
5.4 Estimation . . . 61
5.5 Hypotheses Testing . . . 63
5.5.1 Arrelano Bond Serial Correlation . . . 63
5.5.2 Hansen J-test . . . 63
5.5.3 Wald Test . . . 64

6 Summary **65**

References **66**

About Me

Education

Master of Statistics

The Australian National University (ANU) 2021-2022

Bachelor of Applied Statistics

Sekolah Tinggi Ilmu Statistik (STIS) 2007-2011

Teaching Experience

STAT7055 - Introductory Statistics For Business and Finance

The Australian National University (ANU) 2022

STAT1008 - Quantitative Research Methods

The Australian National University (ANU) 2022

MST3106 - Panel Data Analytics

Universitas Riau (UNRI) 2023

EPSEM4B - Econometrics (On-going)

Universitas Muhammadiyah Riau (UMRI) 2024

Working/Research Experience

Statistician

BPS - Statistics Indonesia (2011 - present)

Research Fellow

Lensa Garuda Nusantara (LGN) (2022 - present)

Research Fellow

Mata Garuda Institute (MGI) (2020 - 2022)

Committee of Department of Research

ANU Indonesian Student Association (ANUISA) (2021 - 2022)

Disclaimer : This Module is currently under development. Please provide Feedback and Suggestions.

1 Pendahuluan

1.1 Pengenalan terhadap Model

1.1.1 Definisi Model

- Model: Wujud Gambaran dari Fenomena/Gejala Sebenarnya
- Peneliti menggunakan model untuk merekonstruksi **hubungan (baik telaah teoritis maupun empiris)** antar variabel.

1.1.2 Tujuan Model:

- Analisis (exploratory or Confirmatory)
- Peramalan (Forecasting)
- Evaluasi (Control, Evaluation)

1.2 Trade-off Modelling

Menyeimbangkan antara idealisasi konsep dan penetapan Asumsi

Modelling: Seni Membangun model karena Sistem Dunia Nyata yang begitu kompleks

1.3 Jenis Model

- Model Verbal/Logika
- Model Fisik
- Model Geometri
- Model Matematika
  - Model Deterministik
  - Model Stochastic Contoh: **Ekonometrika**

1.4 Ekonometrika

1.4.1 Definisi Ekonometrika

- Fenomena/Isu/Permasalahan yang **Dipetakan** berdasarkan teori ***Ekonomi*** yang ada
- **Dinyatakan** dengan persamaan/model ***Matematika***
- **Digunakan** Metode, Kriteria dan Prosedur ***Statistika*** untuk menganalisis permasalahan yang ada

1.4.2 Tujuan Ekonometrika

- Analisis Hubungan
- Peramalan
- Kontrol/Analisis Sebab Akibat (Causal)

1.5 Tahapan Ekonometrika

- Spesifikasi Model (termasuk: Identifikasi Masalah)
- Pengumpulan Data
  - Cross Section (Individu: n, periode: 1)
  - Time Series (Individu: 1, periode: t)
  - Pooled (Individu: n, periode: t)
- Estimasi Parameter
- Pengujian Hipotesis, Asumsi, dan Kelayakan Model
- Analisis/Interpretasi/Peramalan

1.6 Diskusi: Spesifikasi Model (termasuk Identifikasi Masalah)

1.6.1 Tips Mencari Isu/Masalah

- Memastikan Ketersediaan Data
  - Contoh sumber-sumber Data Sekunder:
    - Badan Pusat Statistik (bps.go.id)
    - Bank Indonesia (bi.go.id)
    - ASEAN Stats (aseanstats.org)
    - United Nations Statistical Division (unstats.un.org)
    - International Monetary Fund (imf.org)
    - World Bank (data.worldbank.org)
- Mencari Masalah dengan Cara Mengexplorasi Data

1.6.2 Isu/Permasalahan, Kendaraan, dan Merk Kendaraan

1.7 Challenge 1

- Temukan Isu/Masalah yang menurut Bapak/Ibu/Saudara layak untuk dijadikan sebuah topik Penelitian.

2 Regresi Linear (Ordinary Least Squares)

Tahapan Regresi

- Spesifikasi Model (termasuk: Identifikasi Masalah)
- Pengumpulan Data
  - Cross Section (Individu: n, periode: 1)
  - Time Series (Individu: 1, periode: t)
  - Pooled (Individu: n, periode: t)
- Estimasi Parameter
- Pengujian Hipotesis, Asumsi, dan Kelayakan Model
- Analisis/Interpretasi/Peramalan

```r
# Packages Library yang digunakan
library(bookdown)
library(PoEdata)
library(knitr)
library(xtable)
library(printr)
library(stargazer)
library(rmarkdown)
library(lmtest) #for coeftest() and bptest().
library(broom) #for glance() and tidy()
library(PoEdata) #for PoE4 datasets
library(car) #for hccm() robust standard errors
library(sandwich)
library(knitr)
```

2.1 Spesifikasi Model Umum

Model Regresi Linier mengasumsikan bahwa terdapat hubungan linier antara variabel terikat dan variabel bebas.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + \varepsilon_i$$

Dimana:

- y_i = variabel tidak bebas individu -i
- x_{ki} = variabel bebas ke-k individu -i
- ε_i = residual individu -i

2.1.1 Spesifikasi Model Regresi OLS

- Linear: $y = \beta_0 + \beta_1 X$
- Semi-Log/Log-Lin: $log(y) = \beta_0 + \beta_1 X$
- Semi-Log/Lin-Log: $log(y) = \beta_0 + \beta_1 log(X)$
- Double-Log/Log-Log: $log(y) = \beta_0 + \beta_1 log(X)$
- etc

2.1.2 Spesifikasi Model Contoh: Hubungan Pendapatan terhadap Pengeluaran Makanan

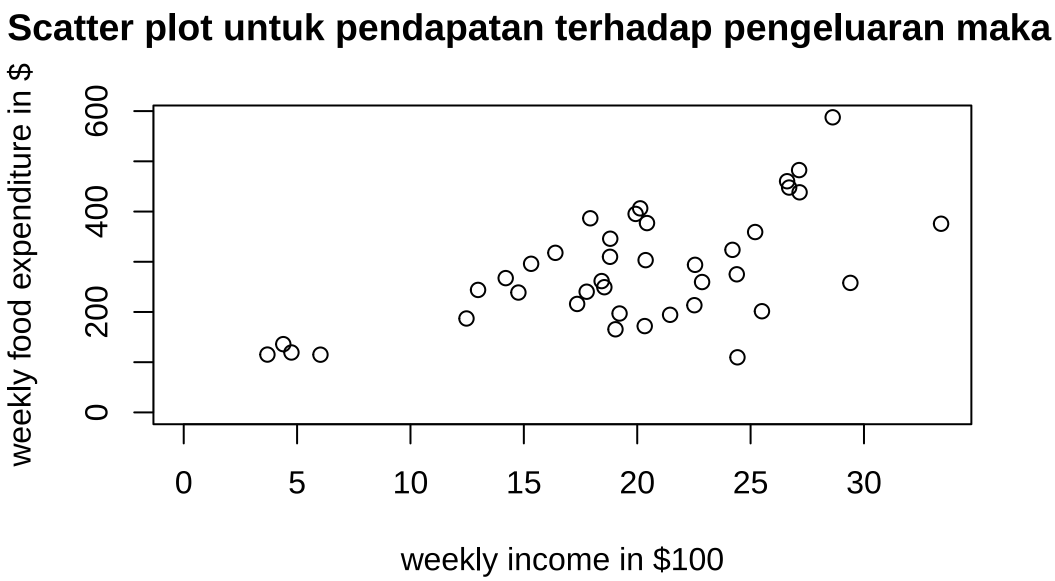
```
library(PoEdata)
data(food)
head(food)
```

food_exp	income
115.22	3.69
135.98	4.39
119.34	4.75
114.96	6.03
187.05	12.47
243.92	12.98

Data berjumlah 40 observasi. Variabel yang diamati ada 2. Data selengkapnya dapat diakses pada link yang disediakan.

```
data("food", package="PoEdata")
plot(food$income, food$food_exp,
     ylim=c(0, max(food$food_exp)),
     xlim=c(0, max(food$income)),
```

```
     xlab="weekly income in $100",
     ylab="weekly food expenditure in $",
     type = "p",
     main="Scatter plot untuk pendapatan terhadap pengeluaran makanan")
```



$$
\text{Food Expenditure}_i = \beta_0 + \beta_1 \text{Income}_i + \varepsilon_i
$$
$$
\varepsilon_i \sim \text{Normal}(0, \sigma^2)
$$

2.2 Estimasi Parameter dengan Regresi

```
mod1 <- lm(food_exp ~ income, data = food)
b0 <- coef(mod1)[[1]]
b1 <- coef(mod1)[[2]]
smod1 <- summary(mod1)
smod1
```

```
Call:
```

```
lm(formula = food_exp ~ income, data = food)

Residuals:
     Min       1Q   Median       3Q      Max 
-223.025  -50.816   -6.324   67.879  212.044 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   83.416     43.410   1.922   0.0622 .  
income        10.210      2.093   4.877 1.95e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 89.52 on 38 degrees of freedom
Multiple R-squared:  0.385, Adjusted R-squared:  0.3688 
F-statistic: 23.79 on 1 and 38 DF,  p-value: 1.946e-05
```

```r
plot(food$income, food$food_exp,
     xlab="weekly income in $100",
     ylab="weekly food expenditure in $",
     ylim=c(0, max(food$food_exp)),
     xlim=c(0, max(food$income)),
     type = "p")
abline(b0,b1)
```

2.2.1 Estimasi Selang Kepercayaan

```
ci <- confint(mod1)
print(ci)
```

```
                2.5 %    97.5 %
(Intercept) -4.463279 171.29528
income       5.972052  14.44723
```

2.3 Pengujian Hipotesis, Asumsi, dan Kelayakan Model

2.3.1 Overall Test

- Gunakan p value dari F-Statistik

H_0 : Tidak ada variabel yang signifikan dalam model

H_1 : Minimal ada 1 variabel yang signifikan dalam model

$\alpha = 0.05$

Keputusan: Tolak H_0, karena $p - value < \alpha$

Kesimpulan: Dengan tingkat keyakinan 95%, cukup bukti untuk menyatakan bahwa minimal ada 1 variabel yang signifikan dalam model.

2.3.2 Partial Test

- Gunakan p value dari t-Statistik

$H_0 : \beta_1 = 0$ variabel tidak signifikan dalam model

$H_1 : \beta_1 \neq 0$ variabel signifikan dalam model

$\alpha = 0.05$

Statistik Uji:

$$t = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)}$$

Keputusan: Tolak H_0, karena $p - value < \alpha$

Kesimpulan: Dengan tingkat keyakinan 95%, cukup bukti untuk menyatakan bahwa variabel signifikan dalam model.

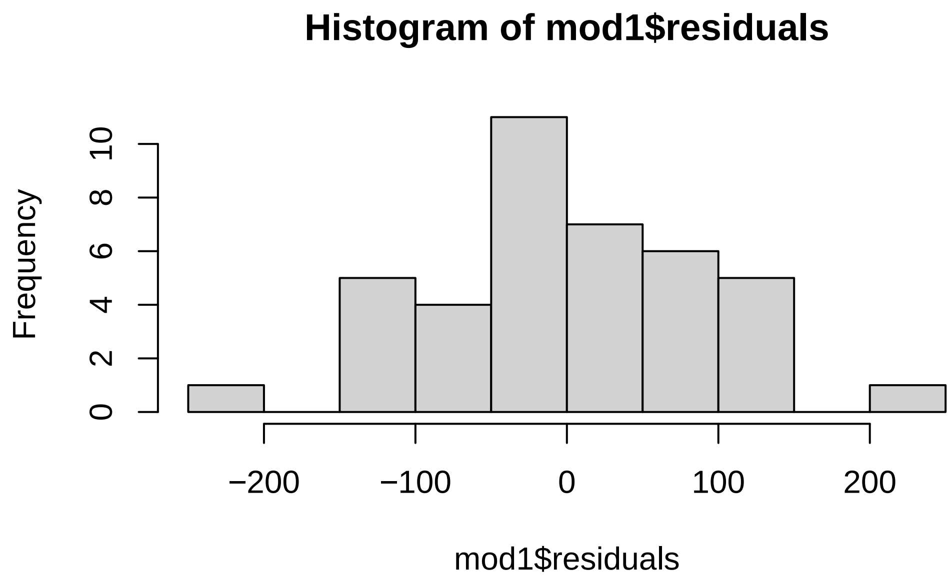
2.3.3 Pengecekan Asumsi/Diagnosa Residual-Residual Diagnostic

2.3.3.1 Linearitas

2.3.3.2 Independensi

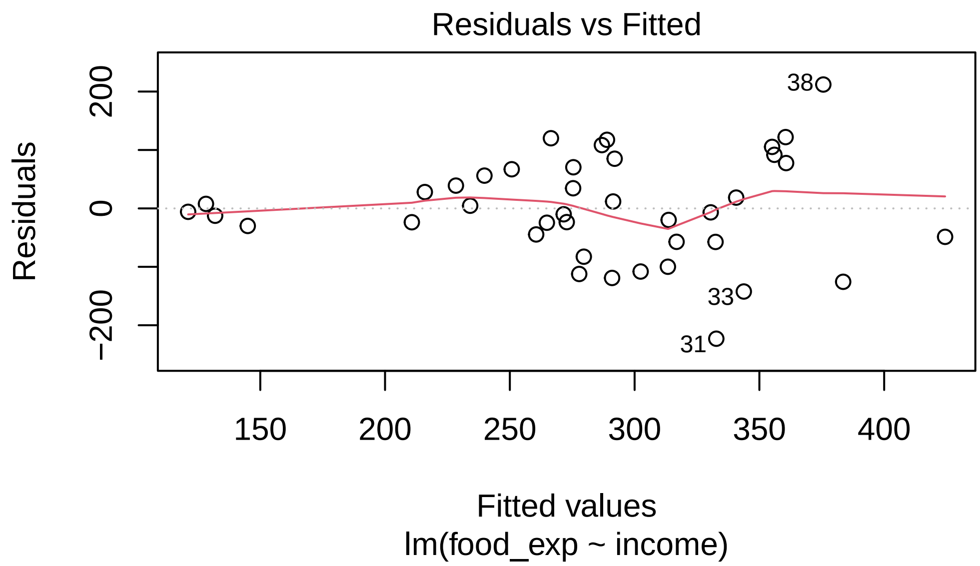
2.3.3.3 Normalitas

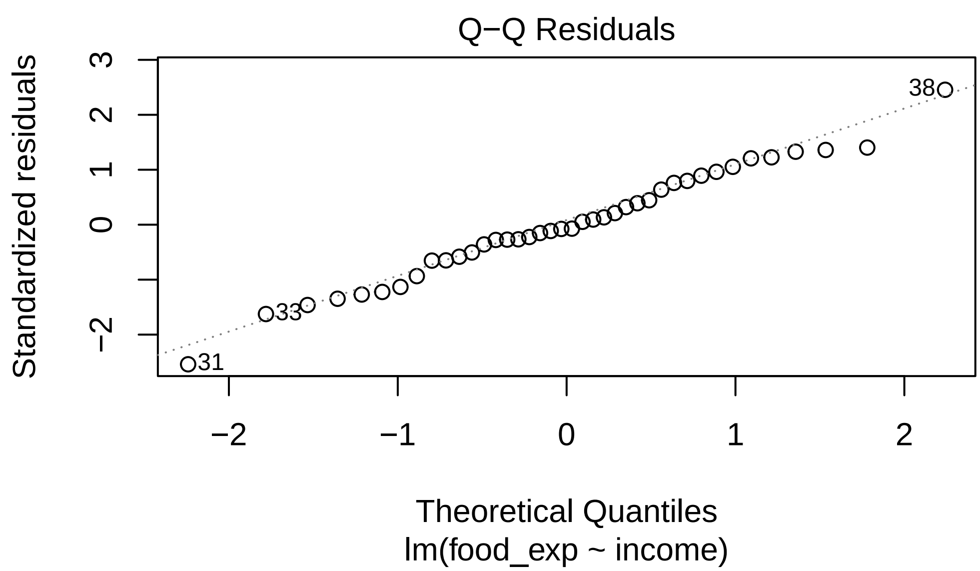
```
hist(mod1$residuals)
```

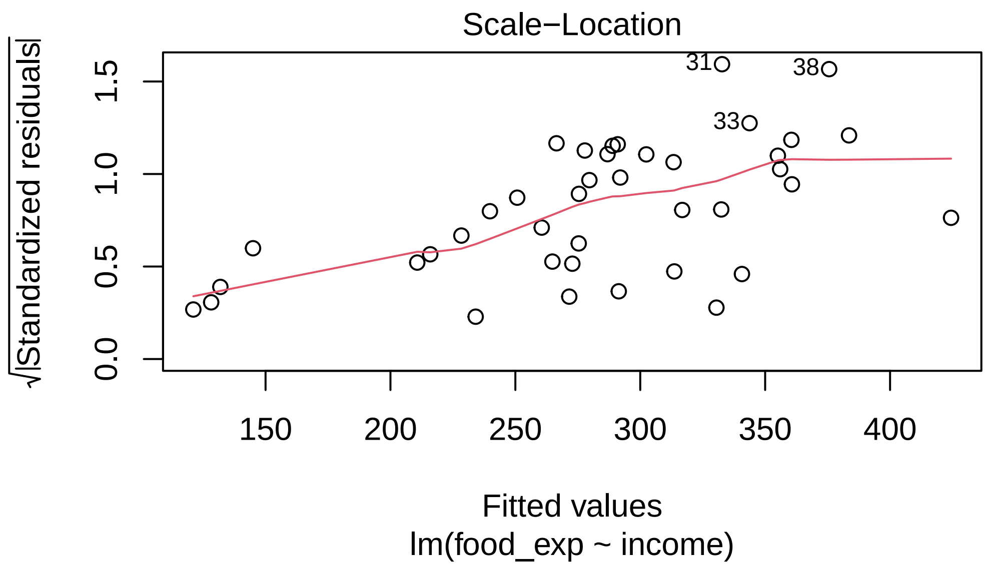


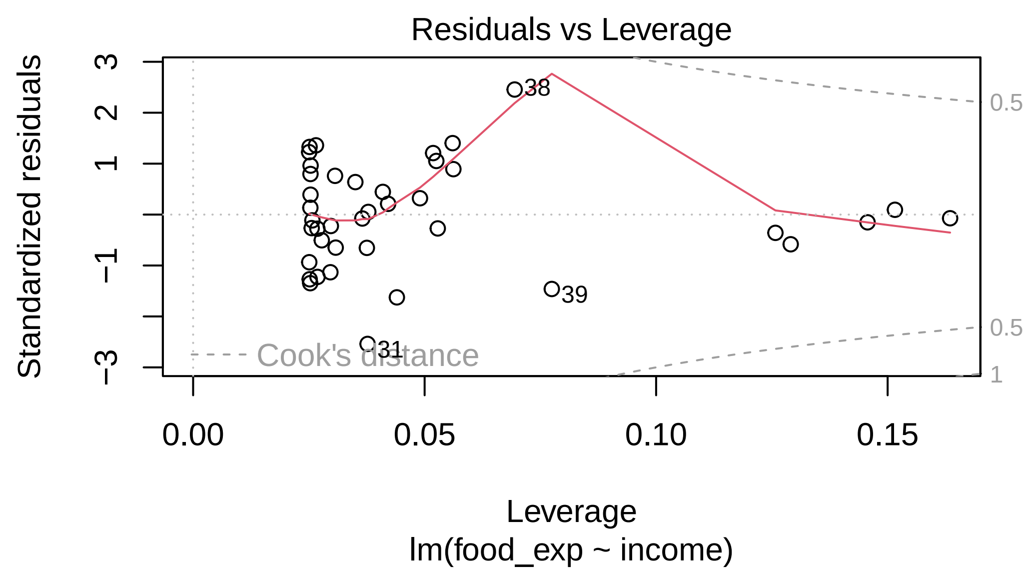
2.3.3.4 Homoskedastisitas

```
plot(mod1)
```


Residuals vs Fitted
Residuals
Fitted values
lm(food_exp ~ income)
38
33
31


Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(food_exp ~ income)
38
33
31


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
150
200
250
300
350
400
31
38
33
Fitted values
lm(food_exp ~ income)


Residuals vs Leverage
Standardized residuals
−3
−1
1
2
3
0.00
0.05
0.10
0.15
38
39
31
Cook's distance
0.5
0.5
1
Leverage
lm(food_exp ~ income)

2.3.4 Kelayakan Model

- Substansi
- Statistik, contohnya: (R^2)

2.3.4.1 Koefisien Determinasi (R^2)

$$SST = SSR + SSE$$

```
anova1=anova(mod1)
anova1
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
income	1	190627.0	190626.984	23.78884	1.95e-05
Residuals	38	304505.2	8013.294	NA	NA

```
r2=anova1$"Sum Sq"[1]/sum(anova1$"Sum Sq")
```

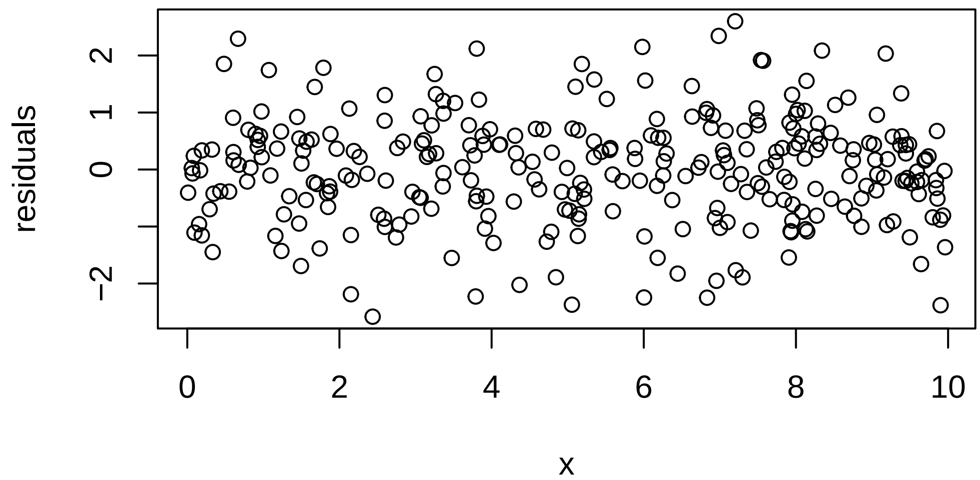
$SST = 4.951322 \times 10^5$

$R^2 = 0.385$

Artinya: 38.5 % proporsi variasi y yang dapat dijelaskan oleh model/variabel x.

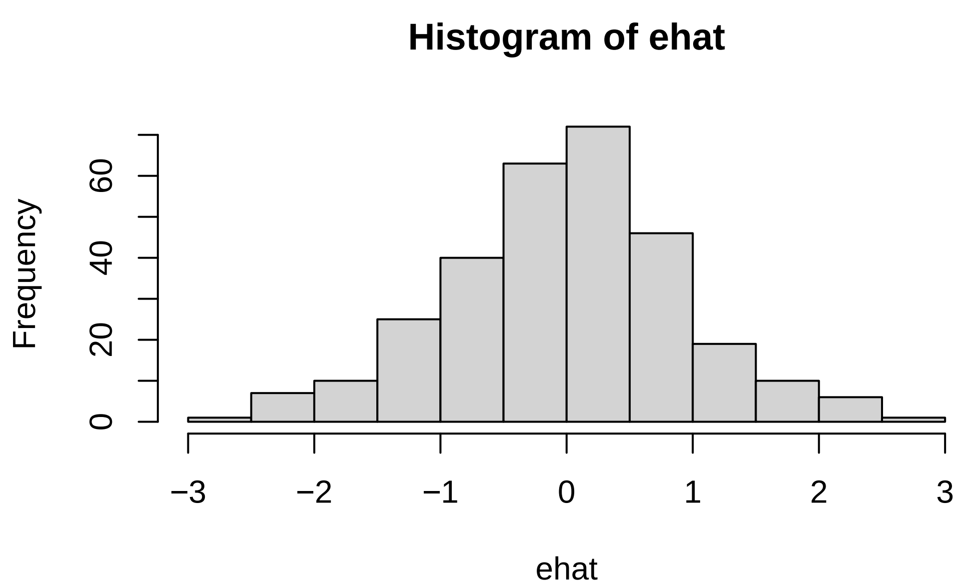
2.3.5 Contoh Pengecekan Asumsi yang sesuai/Ideal

```
set.seed(12345)   #sets the seed for the random number generator
x <- runif(300, 0, 10)
e <- rnorm(300, 0, 1)
y <- 1+x+e
mod3 <- lm(y~x)
ehat <- resid(mod3)
plot(x,ehat, xlab="x", ylab="residuals")
```



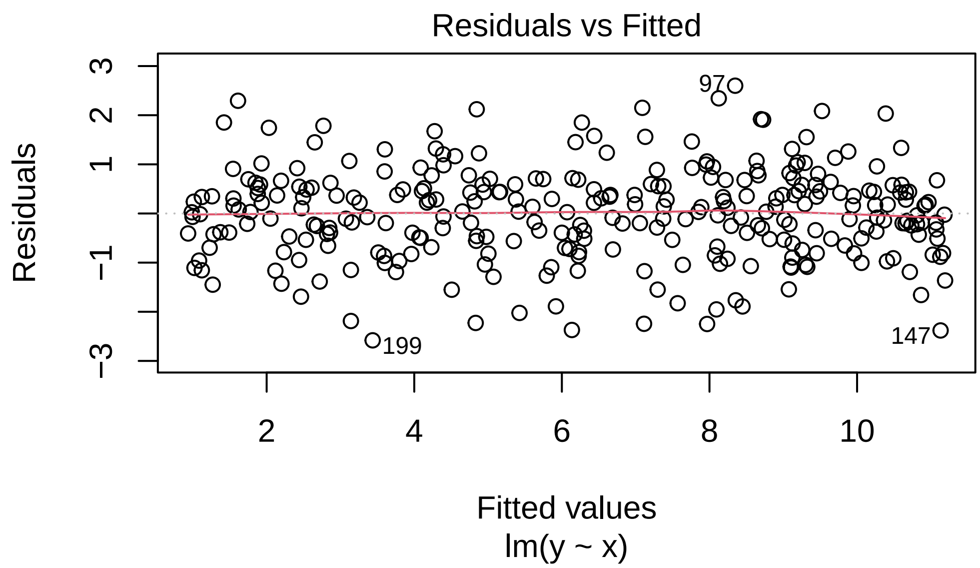
residuals
−2
0
1
2
0
2
4
6
8
10
x


```r
hist(ehat)
```

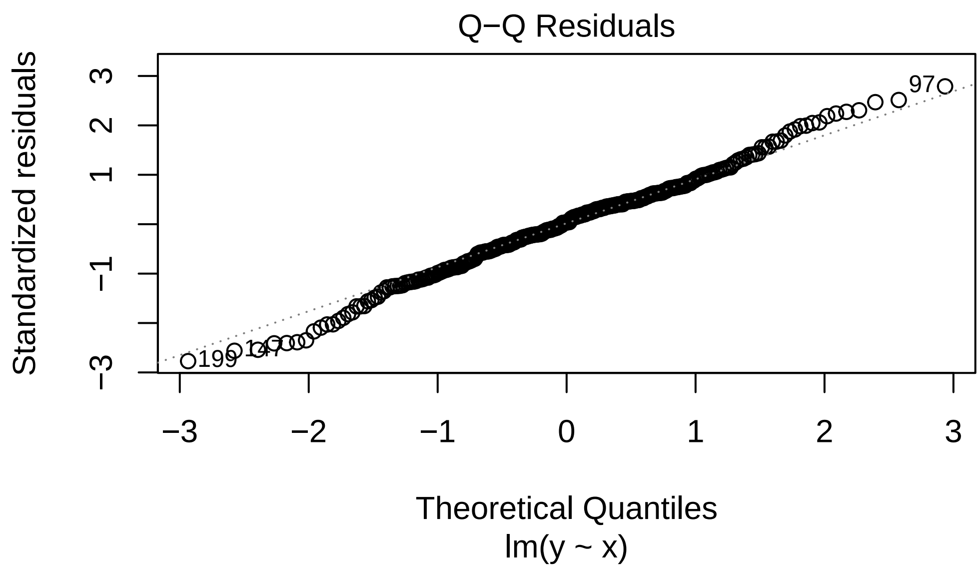


Histogram of ehat
Frequency
0
20
40
60
−3
−2
−1
0
1
2
3
ehat

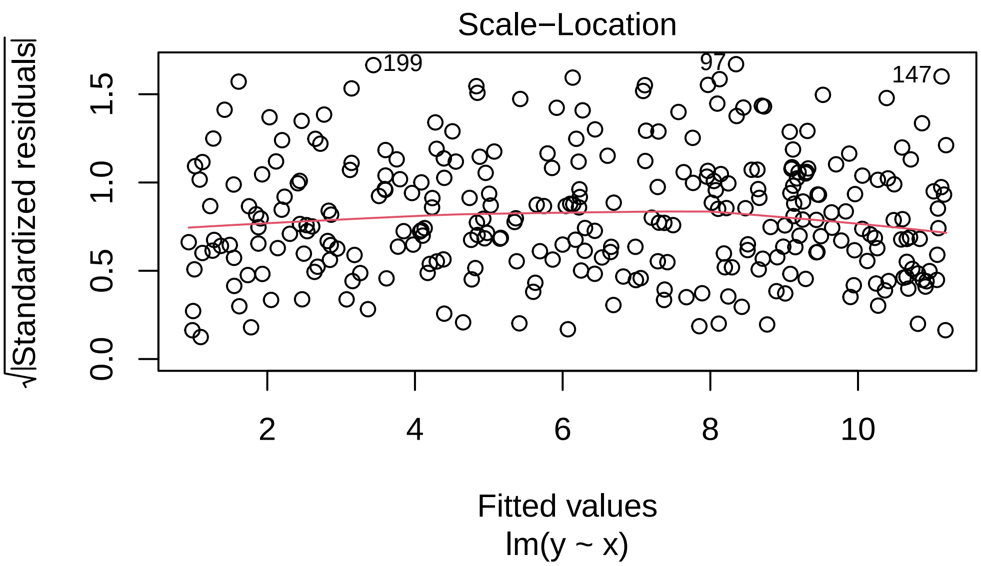
```r
plot(mod3)
```

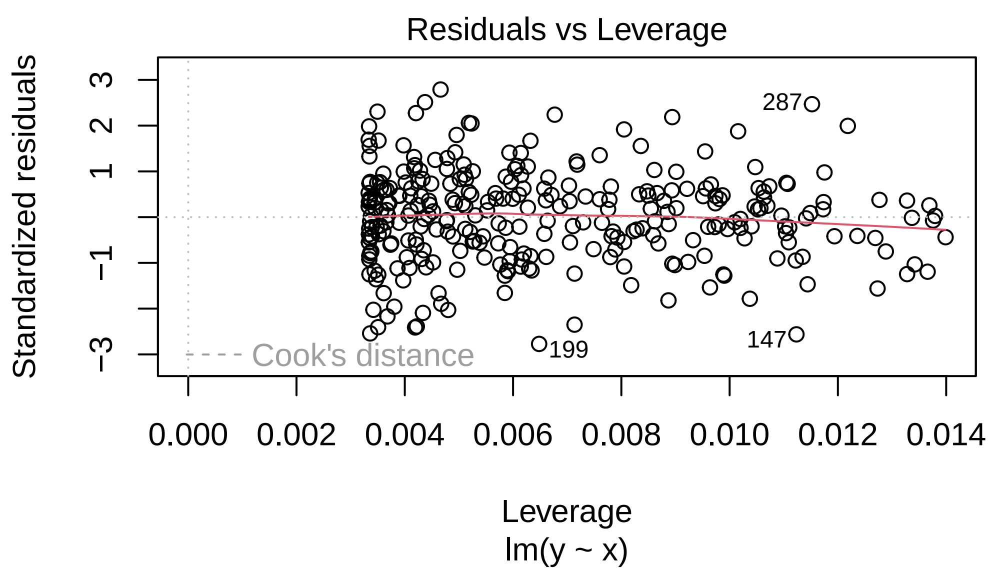


Residuals vs Fitted
Residuals
Fitted values
lm(y ~ x)
97
199
147



Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(y ~ x)
97
199
147


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
199
97
147
2
4
6
8
10
Fitted values
lm(y ~ x)


Residuals vs Leverage
Standardized residuals
−3
−1
1
2
3
287
199
147
Cook's distance
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
Leverage
lm(y ~ x)

2.4 Analisis/Interpretasi/Peramalan

2.4.1 Peramalan

```
newx <- data.frame(income = c(20, 25, 27))
yhat <- predict(mod1, newx)
names(yhat) <- c("income=$2000", "$2500", "$2700")
yhat  # prints the result
```

```
income=$2000        $2500        $2700
    287.6089     338.6571     359.0764
```

2.4.2 Interpretasi

```
mod1
```

```
Call:
lm(formula = food_exp ~ income, data = food)

Coefficients:
(Intercept)       income
      83.42        10.21
```

- Hati-hati menginterpretasikan Estimasi β_0
- Interpretasi dari Estimasi β_1

Setiap Terjadi Kenaikan Income sebesar USD 100 (Kenaikannya sebesar 1 satuan, tetapi satuan x dalam hal ini income adalah USD 100), terjadi kenaikan pengeluaran makanan sebesar USD 10,21 (satuan y dalam hal ini adalah USD).

2.4.2.1 Pembuktian Interpretasi

2.4.2.1.1 Model Linear-Linear

$$y = \beta_0 + \beta X$$

Misalkan nilai awal $y = y_0$, dan nilai awal $X = x_0$. dan nilai akhir $y = y_1$, dan nilai akhir $X = x_1$.

Jelaskan hubungan x dan y jika $x_1 = x_0 + 1$

Persamaan awal

$$y_0 = \beta_0 + \beta x_0$$

Pembuktian

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 \\ &= \beta_0 + \beta_1(x_0 + 1) \\ &= \beta_0 + \beta_1 x_0 + \beta_1 \\ &= y_0 + \beta_1 \\ \beta_1 &= y_1 - y_0 \\ &= \Delta y \end{aligned}$$

Sehingga $\beta_1 = \Delta y$ **Interpretasi** β_1

Ketika X meningkat sebesar 1 satuan x, maka terjadi perubahan pada y sebesar β_1 satuan y.

2.4.2.1.2 Model Linear-Log

2.4.2.1.3 Model Log-Linear

2.4.2.1.4 Model Log-Log

Extras: Model Log-log

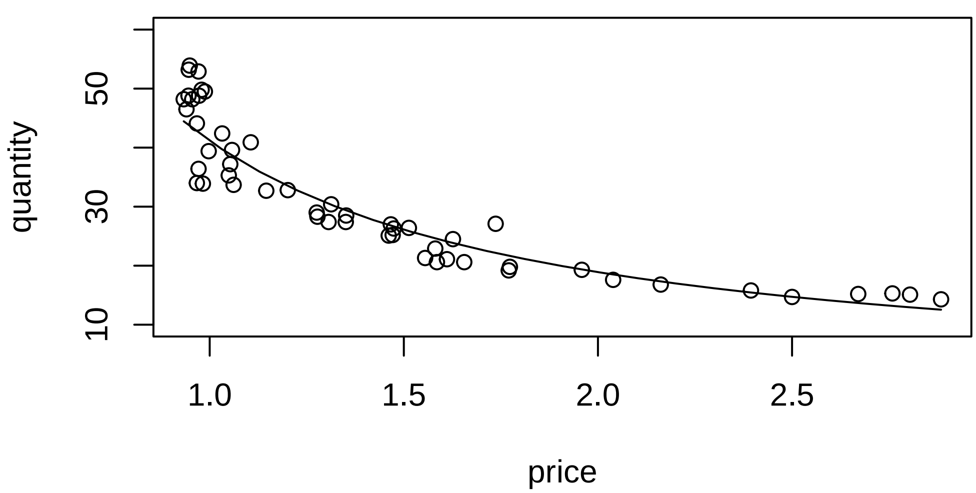
```
# Calculating log-log demand for chicken
data("newbroiler", package="PoEdata")
mod6 <- lm(log(q)~log(p), data=newbroiler)
b1 <- coef(mod6)[[1]]
b2 <- coef(mod6)[[2]]
smod6 <- summary(mod6)
```

```r
tbl <- data.frame(xtable(smod6))
kable(tbl, caption="Model Hubungan Harga terhadap Permintaan Ayam")
```

Table 2.3: Model Hubungan Harga terhadap Permintaan Ayam

	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	3.716944	0.0223594	166.23619	0
log(p)	-1.121358	0.0487564	-22.99918	0

```r
# Drawing the fitted values of the log-log equation
ngrid <- 20 # number of drawing points
xmin <- min(newbroiler$p)
xmax <- max(newbroiler$p)
step <- (xmax-xmin)/ngrid # grid dimension
xp <- seq(xmin, xmax, step)
predicty=exp(b1+b2*log(xp))
plot(newbroiler$p, newbroiler$q, ylim=c(10,60),
     xlab="price", ylab="quantity")
lines(predicty~xp, lty=1, col="black")
```



Extras 2: Model dengan Asumsi Homoskedastisitas yang terlanggar

Karena adanya heteroskedastisitas membuat kesalahan standar kuadrat terkecil menjadi keliru, maka diperlukan metode lain untuk menghitung Regresi.

```
library(car)
foodeq <- lm(food_exp~income,data=food)
kable(tidy(foodeq),caption="Regular standard errors in the 'food' equation")
```

Table 2.4: Regular standard errors in the 'food' equation

term	estimate	std.error	statistic	p.value
(Intercept)	83.41600	43.410163	1.921578	0.0621824
income	10.20964	2.093263	4.877381	0.0000195

```
cov1 <- hccm(foodeq, type="hc1") #needs package 'car'
food.HC1 <- coeftest(foodeq, vcov.=cov1)
kable(tidy(food.HC1),caption="Robust (HC1) standard errors in the 'food' equation")
```

Table 2.5: Robust (HC1) standard errors in the 'food' equation

term	estimate	std.error	statistic	p.value
(Intercept)	83.41600	27.463748	3.037313	0.0042989
income	10.20964	1.809077	5.643566	0.0000018

```
w <- 1/food$income
food.wls <- lm(food_exp~income, weights=w, data=food)
vcvfoodeq <- coeftest(foodeq, vcov.=cov1)
kable(tidy(foodeq),
  caption="OLS estimates for the 'food' equation")
```

Table 2.6: OLS estimates for the 'food' equation

term	estimate	std.error	statistic	p.value
(Intercept)	83.41600	43.410163	1.921578	0.0621824
income	10.20964	2.093263	4.877381	0.0000195

```
kable(tidy(food.wls),
  caption="WLS estimates for the 'food' equation" )
```

Table 2.7: WLS estimates for the 'food' equation

term	estimate	std.error	statistic	p.value
(Intercept)	78.68408	23.788722	3.307621	0.0020641
income	10.45101	1.385891	7.541002	0.0000000

```
kable(tidy(vcvfoodeq),caption=
"OLS estimates for the 'food' equation with robust standard errors" )
```

Table 2.8: OLS estimates for the 'food' equation with robust standard errors

term	estimate	std.error	statistic	p.value
(Intercept)	83.41600	27.463748	3.037313	0.0042989
income	10.20964	1.809077	5.643566	0.0000018

```
data("food", package="PoEdata")
food.ols <- lm(food_exp~income, data=food)
ehatsq <- resid(food.ols)^2
sighatsq.ols  <- lm(log(ehatsq)~log(income), data=food)
vari <- exp(fitted(sighatsq.ols))
food.fgls <- lm(food_exp~income, weights=1/vari, data=food)
```

```
stargazer(food.ols, food.HC1, food.wls, food.fgls, type="text",
          column.labels=c("OLS","HC1","WLS","FGLS"),
#          single.row = TRUE,
#          report = "vc*",
      header = FALSE,
  dep.var.labels.include = FALSE,
  model.numbers = FALSE,
  dep.var.caption="Dependent variable: 'food expenditure'",
  model.names=FALSE
#          df=FALSE,
#          digits=2
)
```

```
=======================================================================
                              Dependent variable: 'food expenditure'
                              -----------------------------------------
                                 OLS        HC1        WLS       FGLS
-----------------------------------------------------------------------
income                        10.210***  10.210***  10.451***  10.633***
                               (2.093)    (1.809)    (1.386)    (0.972)

Constant                       83.416*   83.416***  78.684***  76.054***
                              (43.410)   (27.464)   (23.789)    (9.713)

-----------------------------------------------------------------------
Observations                     40                    40         40
R2                              0.385                 0.599      0.759
Adjusted R2                     0.369                 0.589      0.753
Residual Std. Error (df = 38)   89.517                18.750     1.547
F Statistic (df = 1; 38)      23.789***             56.867*** 119.799***
=======================================================================
Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

```r
stargazer(food.ols, food.HC1, food.wls, food.fgls,
  header=FALSE,
  title="Comparing various 'food' models",
  type=.stargazertype, # "html" or "latex" (in index.Rmd)
  keep.stat="n",  # what statistics to print
  omit.table.layout="n",
  star.cutoffs=NA,
  digits=3,
#  single.row=TRUE,
  intercept.bottom=FALSE, #moves the intercept coef to top
  column.labels=c("OLS","HC1","WLS","FGLS"),
  dep.var.labels.include = FALSE,
  model.numbers = FALSE,
  dep.var.caption="Dependent variable: 'food expenditure'",
  model.names=FALSE,
  star.char=NULL) #supresses the stars
```

Challenge 2

- Lakukan Estimasi Regresi Linear OLS untuk data Pengeluaran Makanan dan Income
- Coba untuk melakukan Estimasi Regresi dari Isu/Permasalahan yang ditemukan dari Challenge 1

3 Time Series Regression (e.g. ECM)

3.1 Stationeritas

```
library(dynlm) #for the `dynlm()` function
library(orcutt) # for the `cochrane.orcutt()` function
library(nlWaldTest) # for the `nlWaldtest()` function
library(zoo) # for time series functions (not much used here)
library(pdfetch) # for retrieving data (just mentioned here)
library(lmtest) #for `coeftest()` and `bptest()`.
library(broom) #for `glance(`) and `tidy()`
library(PoEdata) #for PoE4 datasets
library(car) #for `hccm()` robust standard errors
library(sandwich)
library(knitr) #for kable()
library(forecast)
library(dplyr)
```

Time Series Data adalah data beberapa variabel pada suatu unit pengamatan (seperti individu, negara, atau perusahaan) ketika pengamatan mencakup beberapa periode. Korelasi antara pengamatan selanjutnya, pentingnya tatanan dalam data dan dinamika (nilai data masa lalu mempengaruhi nilai masa kini dan masa depan) merupakan fitur time series data yang tidak terjadi dalam data cross-sectional.

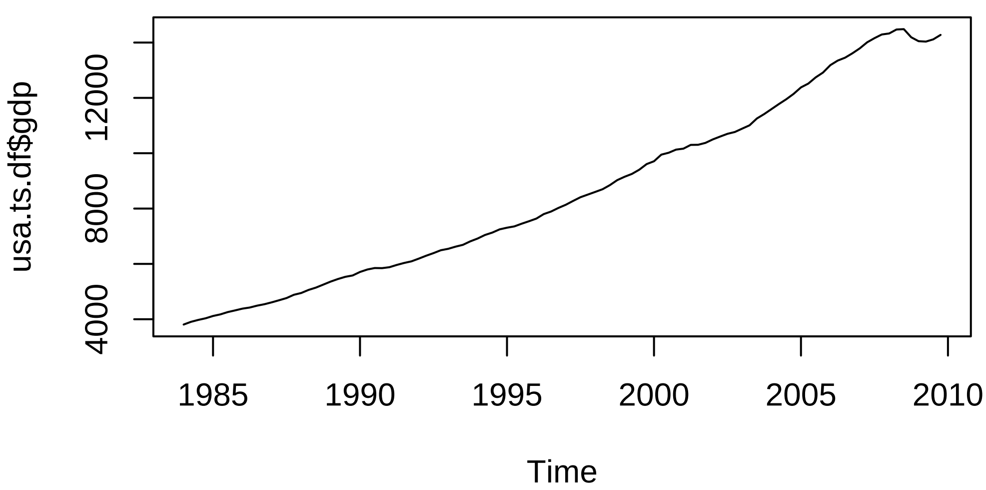
Model Time Series mengasumsikan, selain asumsi regresi linier biasa, bahwa **series-series data tersebut stasioner**, yaitu distribusi error, serta korelasi antar error dalam beberapa periode adalah konstan sepanjang waktu. Distribusi yang konstan mensyaratkan, khususnya, bahwa variabel tersebut tidak menampilkan tren dalam mean atau variansnya; korelasi konstan menyiratkan tidak adanya pengelompokan pengamatan dalam periode tertentu.

Contoh Model Time Series:

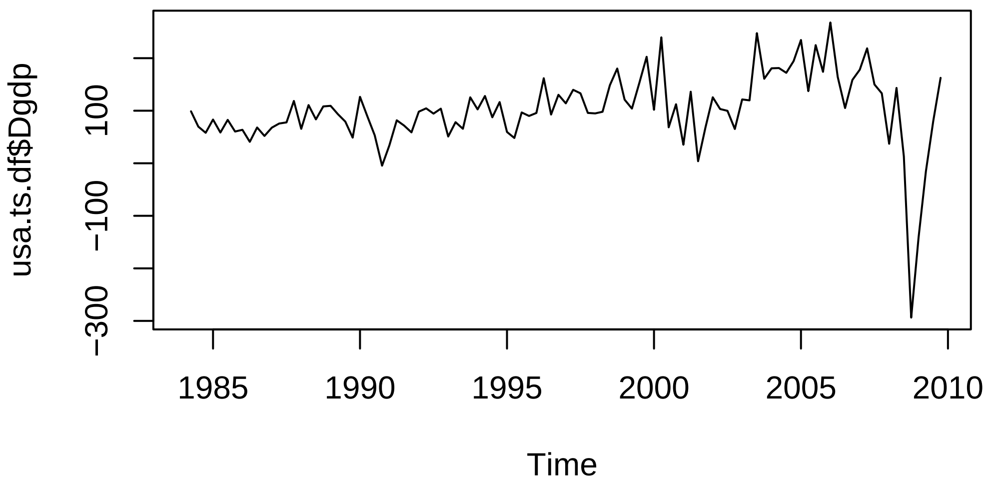
- Stationer, e.g: Regresi OLS, AutoRegressive Distributed Lag Model (ARDL), etc
- Tidak Stasioner, e.g: Error Correction Models (ECM)

Deret waktu dikatakan nonstasioner jika distribusinya, khususnya mean, varians, atau kovarians berdasarkan waktu berubah seiring waktu. Deret waktu nonstasioner tidak dapat digunakan dalam model regresi karena dapat menimbulkan **regresi palsu** , yaitu hubungan yang salah karena, misalnya, tren umum pada variabel yang tidak terkait. Dua atau lebih rangkaian nonstasioner masih dapat menjadi bagian dari model regresi jika keduanya terkointegrasi, yaitu keduanya berada dalam hubungan yang stasioner.

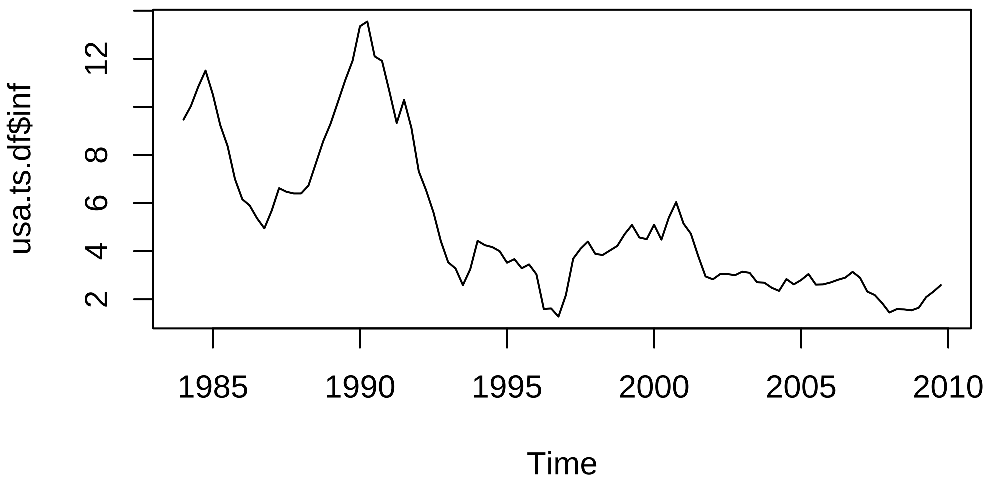
```
data("usa", package="PoEdata")
usa.ts <- ts(usa, start=c(1984,1), end=c(2009,4),
               frequency=4)
Dgdp <- diff(usa.ts[,1])
Dinf <- diff(usa.ts[,"inf"])
Df <- diff(usa.ts[,"f"])
Db <- diff(usa.ts[,"b"])
usa.ts.df <- ts.union(gdp=usa.ts[,1], # package tseries
                      inf=usa.ts[,2],
                      f=usa.ts[,3],
                      b=usa.ts[,4],
                      Dgdp,Dinf,Df,Db,
                      dframe=TRUE)

plot(usa.ts.df$gdp)
```



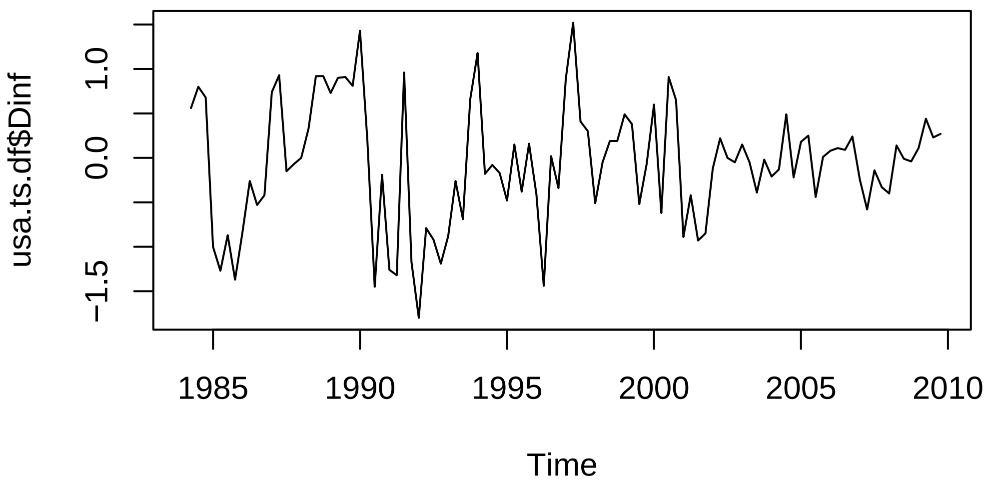
```r
plot(usa.ts.df$Dgdp)
```



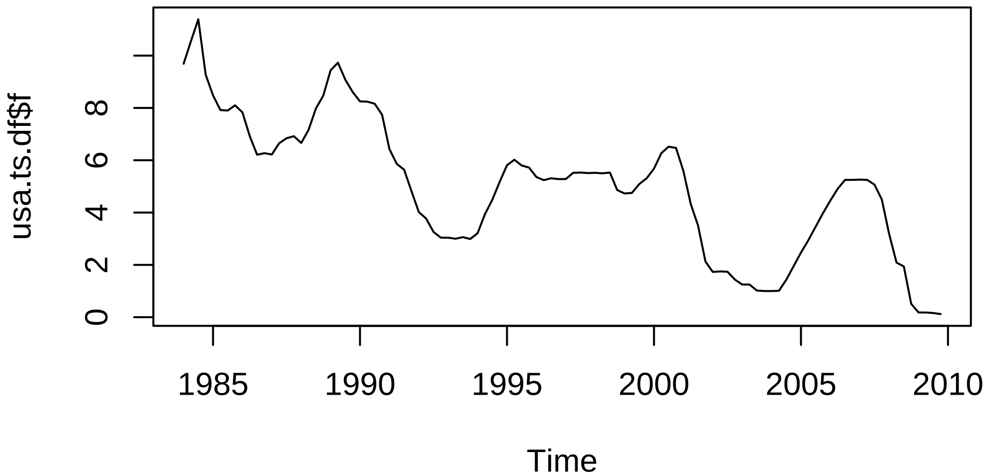

```r
plot(usa.ts.df$inf)
```



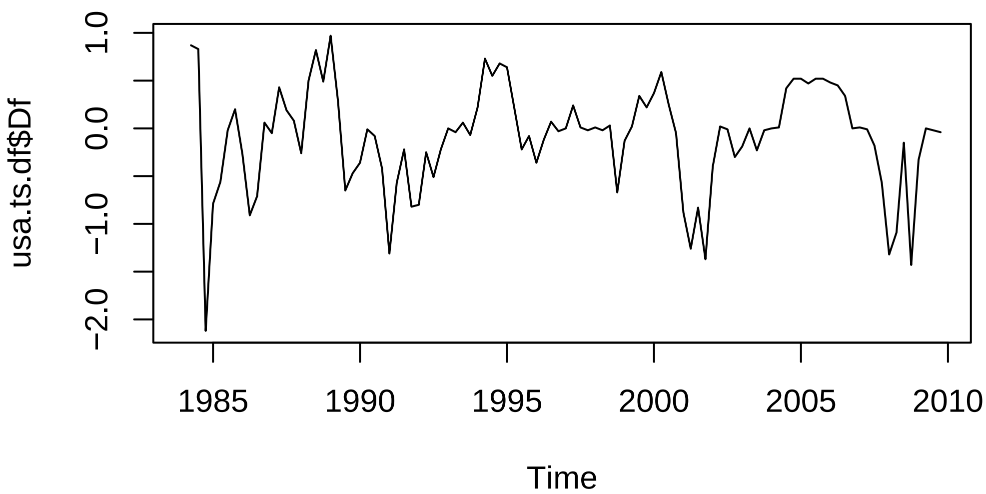

```
plot(usa.ts.df$Dinf)
```



```
plot(usa.ts.df$f)
```

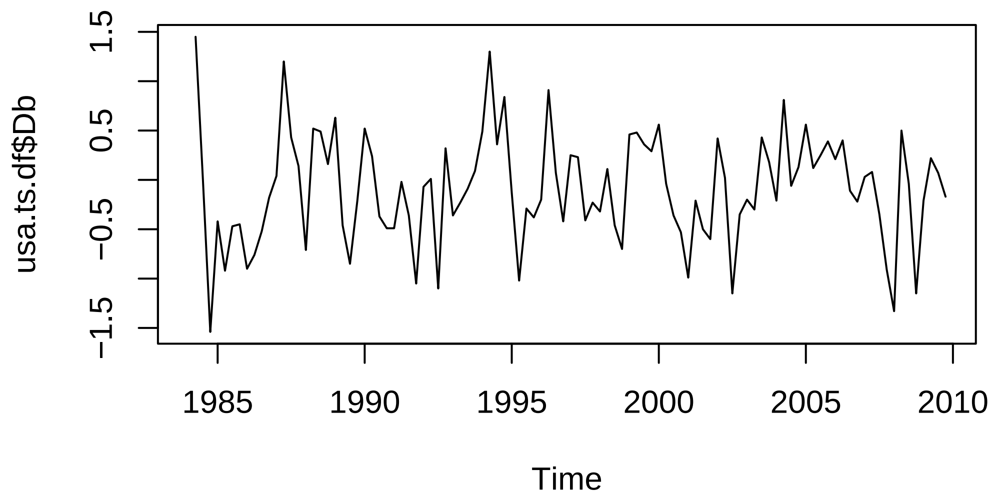



```
plot(usa.ts.df$Df)
```




```r
plot(usa.ts.df$b)
```

```
plot(usa.ts.df$Db)
```



Contoh Dataset:

```
kable(head(usa.ts.df),
caption="Time series data frame constructed with 'ts.union'")
```

Table 3.1: Time series data frame constructed with 'ts.union'

gdp	inf	f	b	Dgdp	Dinf	Df	Db
3807.4	9.47	9.69	11.19	NA	NA	NA	NA
3906.3	10.03	10.56	12.64	98.9	0.56	0.87	1.45
3976.0	10.83	11.39	12.64	69.7	0.80	0.83	0.00
4034.0	11.51	9.27	11.10	58.0	0.68	-2.12	-1.54
4117.2	10.51	8.48	10.68	83.2	-1.00	-0.79	-0.42
4175.7	9.24	7.92	9.76	58.5	-1.27	-0.56	-0.92

3.2 Uji Unit Root untuk Stasioneritas

3.2.1 AR1, Model Autoregressive Orde Pertama

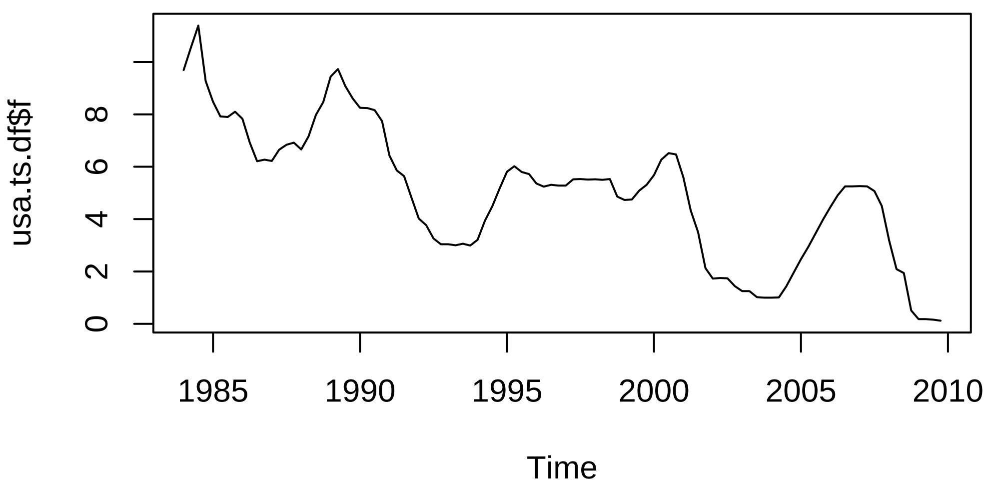
Spesifikasi Model

$$y_t = \rho y_{t-1} + v_t, |\rho| < 1$$

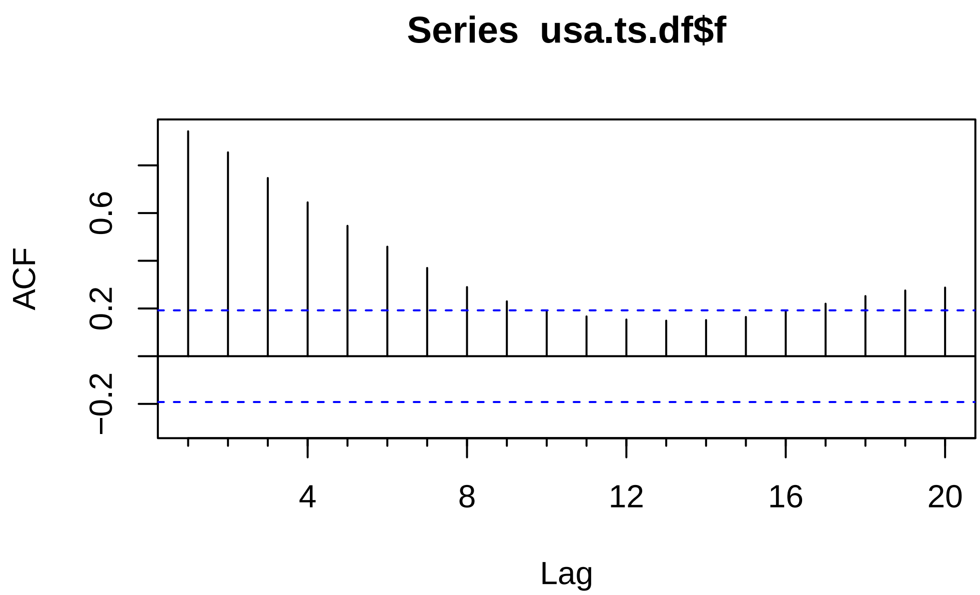
Uji Dickey-Fuller untuk stasioneritas didasarkan pada proses AR(1) sebagaimana didefinisikan dalam Persamaan di atas.

$$H_0 : \rho = 1, H_1 : \rho < 1 \text{ (Variabel Stasioner)}$$

```
plot(usa.ts.df$f)
```




```
Acf(usa.ts.df$f)
```



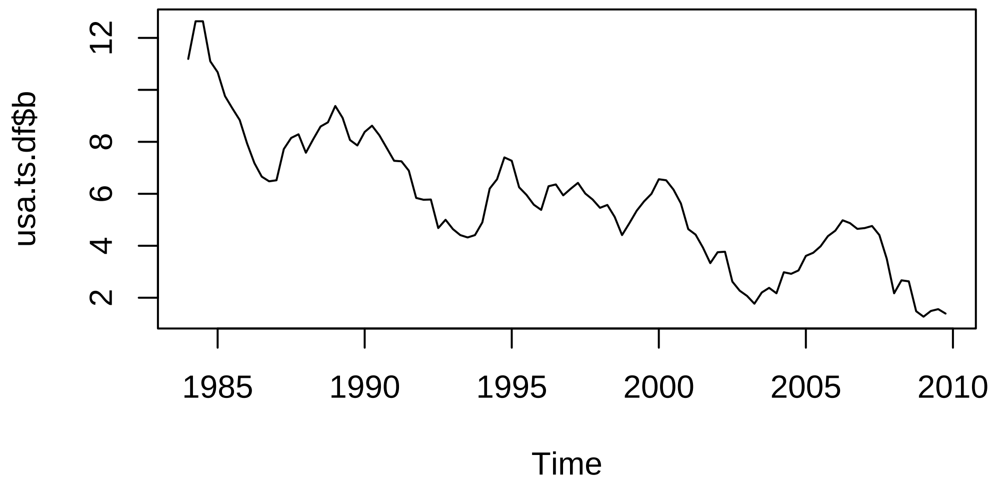
3.2.2 ADF Test USA Funds

```
tseries::adf.test(usa.ts.df$f, k = 10)
```

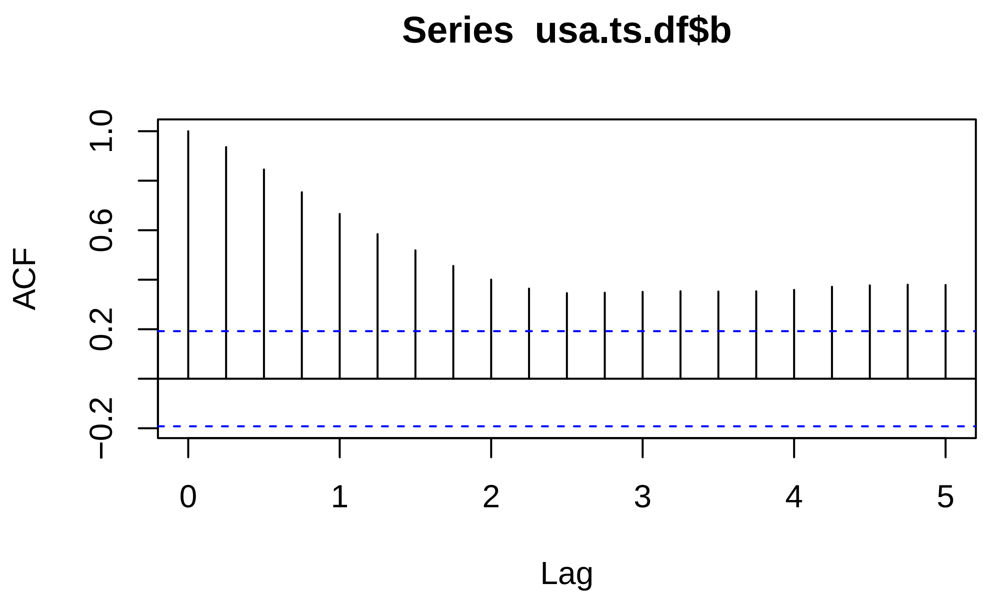
```
	Augmented Dickey-Fuller Test

data:  usa.ts.df$f
Dickey-Fuller = -3.3726, Lag order = 10, p-value = 0.06283
alternative hypothesis: stationary
```

3.2.3 ADF Test USA Bonds

```
plot(usa.ts.df$b)
```




```r
acf(usa.ts.df$b)
```



```r
tseries::adf.test(usa.ts.df$b, k=10)
```

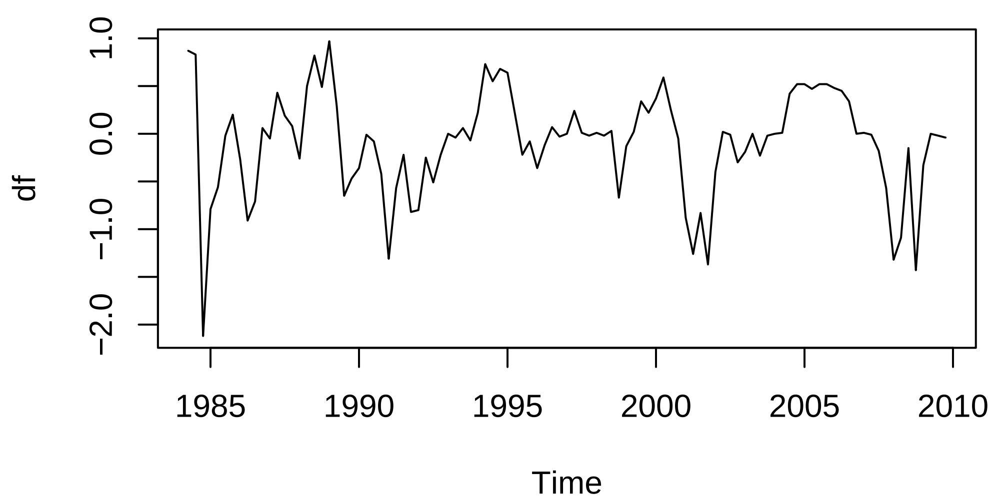
```
    Augmented Dickey-Fuller Test

data:  usa.ts.df$b
Dickey-Fuller = -2.9838, Lag order = 10, p-value = 0.1687
alternative hypothesis: stationary
```

3.3 Differensiasi

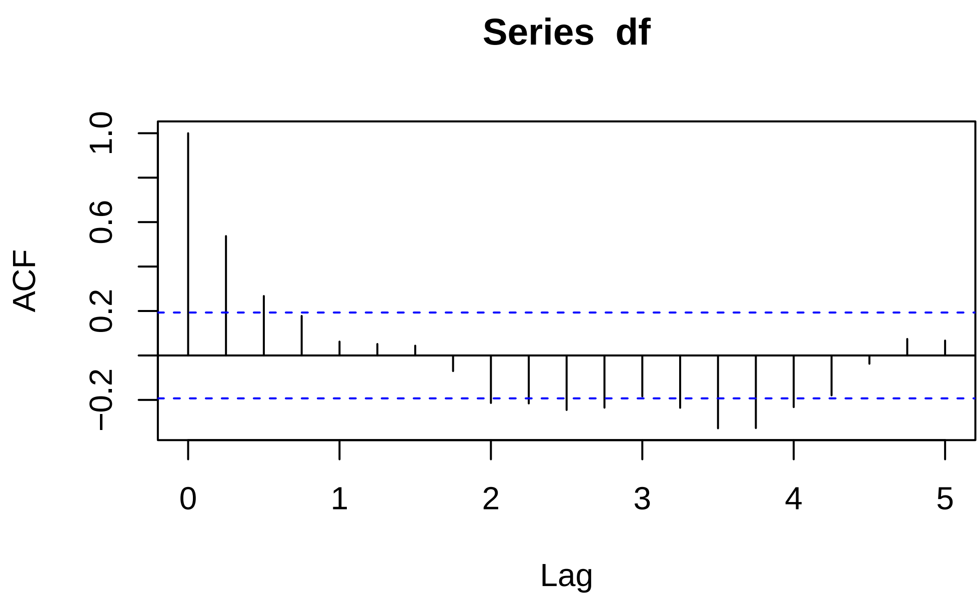
Konsep yang erat kaitannya dengan stasioneritas adalah orde integrasi, yaitu berapa kali kita perlu mendiferensiasikan suatu deret hingga deret tersebut stasioner.

- I(0) - stasioner dalam level
- I(1) jika deret tersebut tidak stasioner pada tingkat-tingkatnya, tetapi stasioner pada perbedaan pertamanya.

```r
df <- diff(usa.ts.df$f)
plot(df)
```



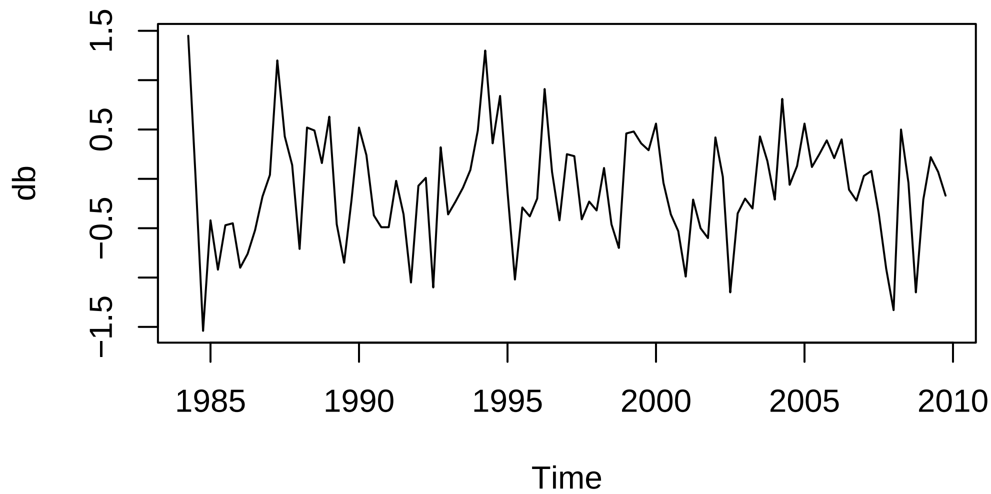
```r
acf(df)
```



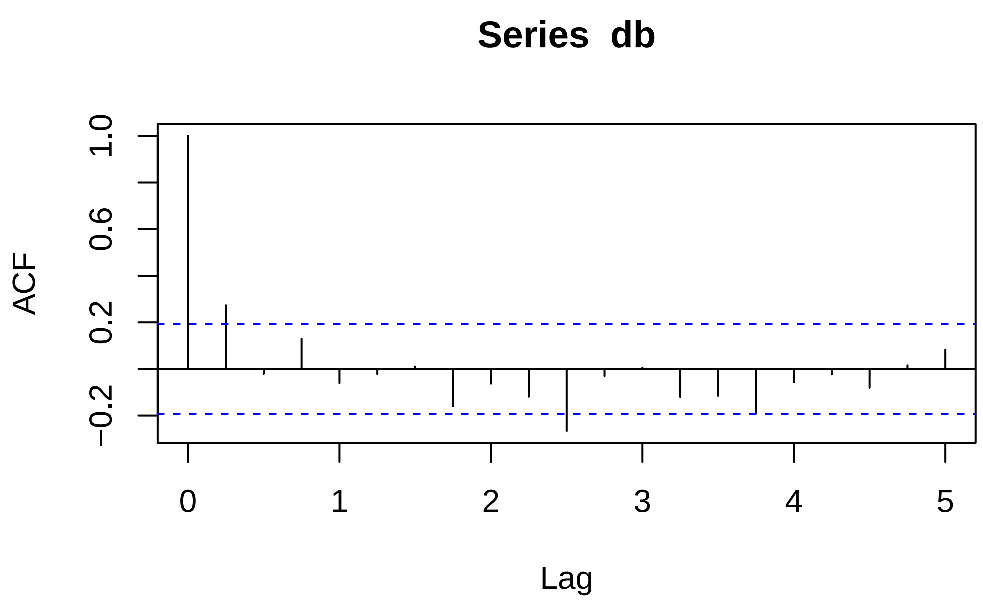

```r
tseries::adf.test(df, k=2)
```

```
	Augmented Dickey-Fuller Test

data:  df
Dickey-Fuller = -4.1782, Lag order = 2, p-value = 0.01
alternative hypothesis: stationary
```

```r
db <- diff(usa.ts.df$b)
plot(db)
```


1.5
0.5
−0.5
−1.5
db
1985
1990
1995
2000
2005
2010
Time

```
acf(db)
```


Series db
1.0
0.6
0.2
−0.2
ACF
0
1
2
3
4
5
Lag

```r
tseries::adf.test(db, k=2)
```

```
	Augmented Dickey-Fuller Test

data:  db
Dickey-Fuller = -4.3976, Lag order = 2, p-value = 0.01
alternative hypothesis: stationary
```

3.4 Kointegrasi

Dua seri terkointegrasi ketika trennya tidak berbeda jauh dan dalam beberapa hal serupa. Uji kointegrasi pada kenyataannya adalah uji stasioneritas Dickey-Fuler terhadap residu, dan hipotesis nolnya adalah nonkointegrasi. Dengan kata lain, kita ingin menolak hipotesis nol dalam uji kointegrasi, seperti yang kita inginkan dalam uji stasioneritas.

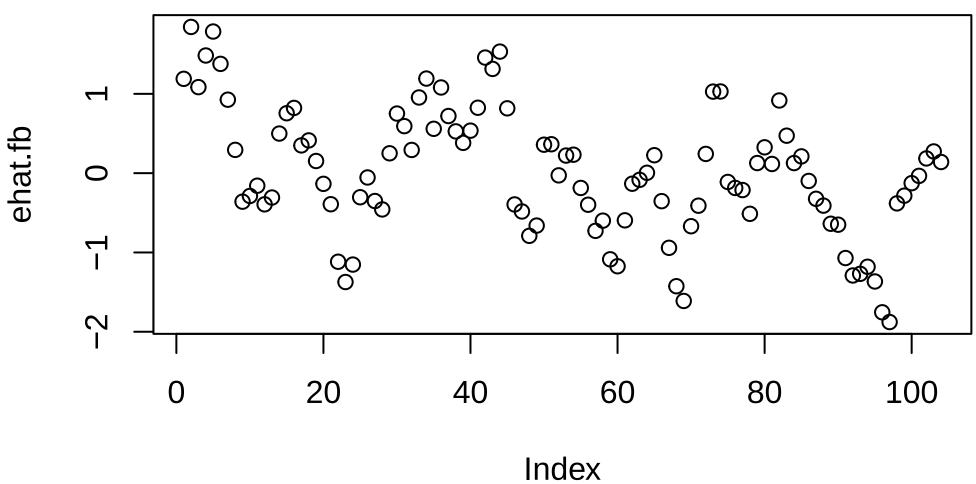
Mari kita terapkan metode ini untuk menentukan keadaan kointegrasi antara rangkaian dan dalam kumpulan data

```r
fb.dyn <- dynlm(b~f, data = usa)
ehat.fb <- resid(fb.dyn)
summary(fb.dyn)
```

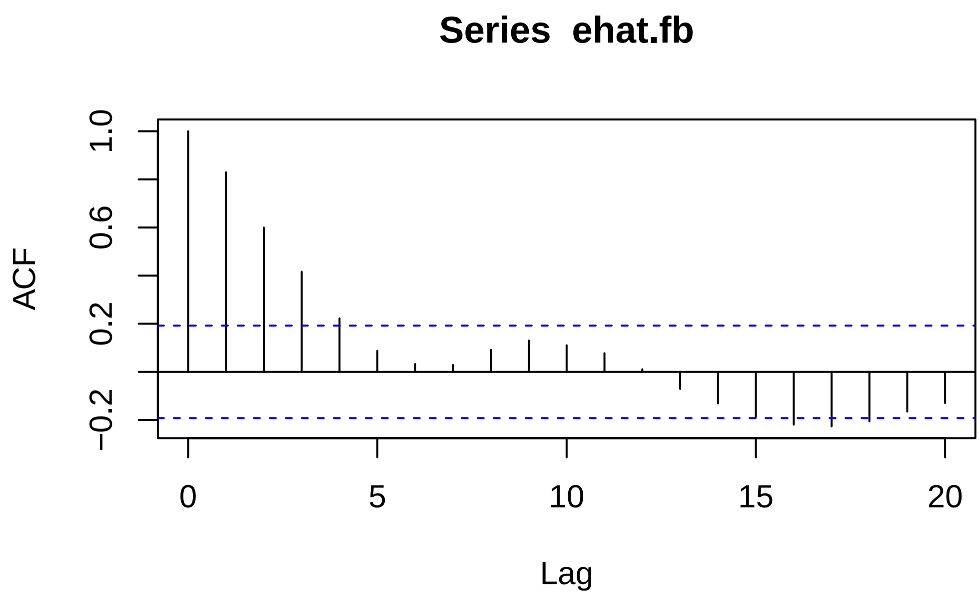
```

Time series regression with "numeric" data:
Start = 1, End = 104

Call:
dynlm(formula = b ~ f, data = usa)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.8777 -0.4220 -0.0445  0.5062  1.8440 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.13983    0.17408   6.548  2.4e-09 ***
f            0.91441    0.03108  29.421  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.8102 on 102 degrees of freedom
Multiple R-squared:  0.8946,	Adjusted R-squared:  0.8936
F-statistic: 865.6 on 1 and 102 DF,  p-value: < 2.2e-16
```

```
#db <- diff(usa.ts.df$b)
plot(ehat.fb)
```




```
acf(ehat.fb)
```




```
tseries::adf.test(ehat.fb, k=4)
```

```
	Augmented Dickey-Fuller Test

data:  ehat.fb
Dickey-Fuller = -4.0009, Lag order = 4, p-value = 0.01184
alternative hypothesis: stationary
```

```
b=usa.ts.df$b
f=usa.ts.df$f
bfx <- as.matrix(cbind(b,f), demean=FALSE)
tseries::po.test(bfx)
```

```
	Phillips-Ouliaris Cointegration Test

data:  bfx
Phillips-Ouliaris demeaned = -20.508, Truncation lag parameter = 1,
p-value = 0.04986
```

3.5 Error Correction Model (ECM)

- An Error Correction Model (ECM) merupakan metode standard untuk memodelkan data time series.
- The ECM makes it possible to deal with nonstationary data series and separates the long and short run.

3.5.1 Two Steps Engle Granger

Spesifikasi Model

Tahap 1

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

$$\hat{u_t} = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t$$

Dimana:

- β_1 merupakan Koefisien Model Long-run

Tahap 2

$$\Delta y_t = \beta_2 + \beta_3 \Delta x_t - \pi_1 \hat{u}_{t-1} + \varepsilon_t$$

Dimana:

- π_1 is the feedback effect, or the adjustment effect, or error correction coefficient and shows how much of the disequilibrium is being corrected.
- ε_t merupakan white noise error term

** Step 1 **

```
b=usa.ts.df$b
f=usa.ts.df$f
fb.dyn <- dynlm(b~f)
summary(fb.dyn)
```

```
Time series regression with "ts" data:
Start = 1984(1), End = 2009(4)

Call:
dynlm(formula = b ~ f)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.8777 -0.4220 -0.0445  0.5062  1.8440 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.13983    0.17408   6.548  2.4e-09 ***
f            0.91441    0.03108  29.421  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8102 on 102 degrees of freedom
Multiple R-squared:  0.8946,	Adjusted R-squared:  0.8936 
F-statistic: 865.6 on 1 and 102 DF,  p-value: < 2.2e-16
```

```
ect=fb.dyn$residuals
tseries::adf.test(ect)
```

```

	Augmented Dickey-Fuller Test

data:  ect
Dickey-Fuller = -4.0009, Lag order = 4, p-value = 0.01184
alternative hypothesis: stationary
```

Signifikansi Menunjukkan adanya kointegrasi

** Step 2 **

```
# Set ECT dan Variabel penting lainnya menjadi time Series
ect=ts(ect,start=c(1984,1), end=c(2009,4),
               frequency=4)
ect1=ts(ect,start=c(1984,2), end=c(2009,4),
               frequency=4)
```

```r
b=ts(b,start=c(1984,1), end=c(2009,4),
               frequency=4)

f=ts(f,start=c(1984,1), end=c(2009,4),
               frequency=4)

L1.b=stats::lag(b,-1)
L1.f=stats::lag(f,-1)
L1.b=ts(L1.b,start=c(1984,2), end=c(2009,4),
               frequency=4)
L1.f=ts(L1.f,start=c(1984,2), end=c(2009,4),
               frequency=4)

tsdata=ts.union(b,f,L1.b,L1.f,ect,ect1)
head(tsdata)
```

```
            b     f  L1.b  L1.f      ect     ect1
1984 Q1 11.19  9.69    NA    NA 1.189524       NA
1984 Q2 12.64 10.56 11.19  9.69 1.843986 1.189524
1984 Q3 12.64 11.39 12.64 10.56 1.085025 1.843986
1984 Q4 11.10  9.27 12.64 11.39 1.483577 1.085025
1985 Q1 10.68  8.48 11.10  9.27 1.785962 1.483577
1985 Q2  9.76  7.92 10.68  8.48 1.378032 1.785962
```

```r
regECM1=lm(diff(b)~diff(f)+ect1)
summary(regECM1)
```

```

Call:
lm(formula = diff(b) ~ diff(f) + ect1)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.03574 -0.30858 -0.03668  0.24341  1.06092 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.03039    0.04265  -0.713   0.4778    
diff(f)      0.69873    0.08128   8.597 1.16e-13 ***
ect1        -0.12307    0.05502  -2.237   0.0275 *  
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4261 on 100 degrees of freedom
Multiple R-squared:  0.426, Adjusted R-squared:  0.4145
F-statistic: 37.11 on 2 and 100 DF,  p-value: 8.811e-13
```

3.5.2 One Step ECM

$$\Delta y_t = \beta_2 + \beta_3 \Delta x_t - \pi_1 \hat{u}_{t-1} + \varepsilon_t$$
$$\Delta y_t = \beta_2 + \beta_3 \Delta x_t - \pi_1 (y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 x_{t-1}) + \varepsilon_t$$

$$\Delta y_t = \beta_2 + \beta_3 \Delta x_t - \pi_1 (y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 x_{t-1}) + \varepsilon_t$$

$$\Delta y_t = \beta_2 + \beta_3 \Delta x_t - \pi_1 y_{t-1} + \pi_1 \hat{\beta}_0 + \pi_1 \hat{\beta}_1 x_{t-1} + \varepsilon_t$$

Susun kembali, dan misalkan jika $\pi_1 = -\alpha_1$ dan $\pi_1 \beta_1 = \alpha_2$

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_{t-1} + \alpha_3 \Delta x_t + \varepsilon_t$$

Maka diperoleh hubungan

$$\Delta y_t = \alpha_0 - \alpha_1 (y_{t-1} + \frac{\alpha_2}{\alpha_1} x_{t-1}) + \alpha_3 \Delta x_t + \varepsilon_t$$

Long-Run Coefficient Model

$$\hat{\beta}_1 = -\frac{\alpha_2}{\alpha_1}$$

```r
regECM2=lm(diff(b)~L1.b+L1.f+diff(f))
summary(regECM2)
```

```
Call:
lm(formula = diff(b) ~ L1.b + L1.f + diff(f))

Residuals:
     Min       1Q   Median       3Q      Max
```

```
-1.06697 -0.30745 -0.05507  0.26760  1.13987

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.22636    0.11166   2.027   0.0453 *
L1.b        -0.12001    0.05476  -2.191   0.0308 *
L1.f         0.08565    0.05338   1.605   0.1118
diff(f)      0.68582    0.08133   8.433 2.82e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4238 on 99 degrees of freedom
Multiple R-squared:  0.4379,	Adjusted R-squared:  0.4209
F-statistic: 25.71 on 3 and 99 DF,  p-value: 2.214e-12
```

```r
# Homoskedastisitas Check
library(car)
ncvTest(regECM2)
```

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.000515675, Df = 1, p = 0.98188
```

```r
# Homoskedastisitas Check
library(lmtest)
bgtest(regECM2)
```

```

	Breusch-Godfrey test for serial correlation of order up to 1

data:  regECM2
LM test = 2.2827, df = 1, p-value = 0.1308
```

3.5.3 ARDL

```r
library(ARDL)
ardl1=ardl(b~f, data=usa.ts.df, order=c(1,1))
summary(ardl1)
```

```
Time series regression with "ts" data:
Start = 2, End = 104

Call:
dynlm::dynlm(formula = full_formula, data = data, start = start,
    end = end)

Residuals:
     Min       1Q   Median       3Q      Max
-1.06697 -0.30745 -0.05507  0.26760  1.13987

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.22636    0.11166   2.027   0.0453 *
L(b, 1)      0.87999    0.05476  16.070  < 2e-16 ***
f            0.68582    0.08133   8.433 2.82e-13 ***
L(f, 1)     -0.60017    0.08077  -7.431 3.89e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4238 on 99 degrees of freedom
Multiple R-squared:  0.9706,    Adjusted R-squared:  0.9697
F-statistic:  1089 on 3 and 99 DF,  p-value: < 2.2e-16
```

```r
bounds_f_test(ardl1, case=3)
```

```
    Bounds F-test (Wald) for no cointegration

data:  d(b) ~ L(b, 1) + L(f, 1) + d(f)
F = 3.5753, p-value = 0.2269
alternative hypothesis: Possible cointegration
null values:
   k    T
   1 1000
```

```r
bounds_t_test(ardl1, case=3)
```

```
    Bounds t-test for no cointegration

data:  d(b) ~ L(b, 1) + L(f, 1) + d(f)
t = -2.1915, p-value = 0.3339
alternative hypothesis: Possible cointegration
null values:
   k    T
   1 1000
```

```r
ecm=uecm(ardl1)
summary(ecm)
```

```
Time series regression with "ts" data:
Start = 2, End = 104

Call:
dynlm::dynlm(formula = full_formula, data = data, start = start,
    end = end)

Residuals:
     Min       1Q   Median       3Q      Max
-1.06697 -0.30745 -0.05507  0.26760  1.13987

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.22636    0.11166   2.027   0.0453 *
L(b, 1)     -0.12001    0.05476  -2.191   0.0308 *
L(f, 1)      0.08565    0.05338   1.605   0.1118
d(f)         0.68582    0.08133   8.433 2.82e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4238 on 99 degrees of freedom
Multiple R-squared:  0.4379,    Adjusted R-squared:  0.4209
F-statistic: 25.71 on 3 and 99 DF,  p-value: 2.214e-12
```

4 Model Persamaan Simultan

4.1 Definisi

Persamaan simultan merupakan model yang mempunyai lebih dari satu variabel respon, dimana penyelesaiannya ditentukan oleh kesetimbangan antara gaya-gaya yang berlawanan.

Contoh umum dari masalah persamaan simultan ekonomi adalah model penawaran dan permintaan, dimana harga dan kuantitas saling bergantung dan ditentukan oleh interaksi antara penawaran dan permintaan.

4.2 Spesifikasi Model

4.2.1 Persamaan Struktural

model ekonomi seperti persamaan permintaan dan penawaran mencakup beberapa variabel dependen (endogen) dalam setiap persamaan. Model seperti ini disebut bentuk struktural model.

4.2.2 Persamaan reduced

Jika bentuk strukturalnya ditransformasikan sedemikian rupa sehingga setiap persamaan menunjukkan satu variabel terikat sebagai fungsi dari variabel bebas eksogen saja, maka bentuk baru tersebut disebut bentuk tereduksi . Bentuk tereduksi dapat diperkirakan dengan kuadrat terkecil, sedangkan bentuk struktural tidak dapat diperkirakan karena memuat variabel endogen di sisi kanannya.

4.3 Permasalahan Identifikasi

Kondisi yang diperlukan untuk identifikasi mensyaratkan bahwa, agar masalah memiliki solusi, setiap persamaan dalam bentuk struktural sistem harus melewatkan setidaknya satu variabel eksogen yang ada dalam persamaan lainnya.

4.4 Langkah-langkah dengan Contoh 1

Persamaan struktural permintaan dan penawaran (Persamaan 1 Dan 2) dirumuskan berdasarkan teori ekonomi; kuantitas dan harga bersifat endogen, dan semua variabel lainnya dianggap eksogen.

$$q_d = \alpha_1 + \alpha_2 p + \alpha_3 ps + \alpha_4 di + e_d$$
$$q_s = \beta_1 + \beta_2 p + \beta_3 pf + e_s$$

```r
rm(list=ls()) #Removes all items in Environment!
library(systemfit)
library(broom) #for `glance(`) and `tidy()`
library(PoEdata) #for PoE4 dataset
library(knitr) #for kable()
```

```r
data("truffles", package="PoEdata")
D <- q~p+ps+di
S <- q~p+pf
sys <- list(D,S)
instr <- ~ps+di+pf
truff.sys <- systemfit(sys, inst=instr,
                       method="2SLS", data=truffles)
summary(truff.sys)
```

```
systemfit results
method: 2SLS

        N DF     SSR detRCov   OLS-R2 McElroy-R2
system 60 53 692.472 49.8028 0.438964   0.807408

     N DF      SSR      MSE    RMSE        R2    Adj R2
eq1 30 26 631.9171 24.30450 4.92996 -0.023950 -0.142098
eq2 30 27  60.5546  2.24276 1.49758  0.901878  0.894610

The covariance matrix of the residuals
         eq1     eq2
eq1 24.30451 2.16943
eq2  2.16943 2.24276
```

```
The correlations of the residuals
        eq1     eq2
eq1 1.00000 0.29384
eq2 0.29384 1.00000


2SLS estimates for 'eq1' (equation 1)
Model Formula: q ~ p + ps + di
Instruments: ~ps + di + pf

             Estimate Std. Error  t value  Pr(>|t|)
(Intercept) -4.279471   5.543884 -0.77193 0.4471180
p           -0.374459   0.164752 -2.27287 0.0315350 *
ps           1.296033   0.355193  3.64881 0.0011601 **
di           5.013977   2.283556  2.19569 0.0372352 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.92996 on 26 degrees of freedom
Number of observations: 30 Degrees of Freedom: 26
SSR: 631.917143 MSE: 24.304505 Root MSE: 4.92996
Multiple R-Squared: -0.02395 Adjusted R-Squared: -0.142098


2SLS estimates for 'eq2' (equation 2)
Model Formula: q ~ p + pf
Instruments: ~ps + di + pf

              Estimate Std. Error  t value   Pr(>|t|)
(Intercept) 20.0328022  1.2231148  16.3785 1.5543e-15 ***
p            0.3379816  0.0249196  13.5629 1.4344e-13 ***
pf          -1.0009094  0.0825279 -12.1281 1.9456e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.497585 on 27 degrees of freedom
Number of observations: 30 Degrees of Freedom: 27
SSR: 60.554565 MSE: 2.242762 Root MSE: 1.497585
Multiple R-Squared: 0.901878 Adjusted R-Squared: 0.89461
```

```
Q.red <- lm(q~ps+di+pf, data=truffles)
P.red <- lm(p~ps+di+pf, data=truffles)
kable(tidy(Q.red), digits=4,
      caption="Reduced form for quantity")
```

Table 4.1: Reduced form for quantity

term	estimate	std.error	statistic	p.value
(Intercept)	7.8951	3.2434	2.4342	0.0221
ps	0.6564	0.1425	4.6051	0.0001
di	2.1672	0.7005	3.0938	0.0047
pf	-0.5070	0.1213	-4.1809	0.0003

```
kable(tidy(P.red), digits=4,
      caption="Reduced form for price")
```

Table 4.2: Reduced form for price

term	estimate	std.error	statistic	p.value
(Intercept)	-32.5124	7.9842	-4.0721	4e-04
ps	1.7081	0.3509	4.8682	0e+00
di	7.6025	1.7243	4.4089	2e-04
pf	1.3539	0.2985	4.5356	1e-04

4.5 Contoh 2

```
data("fultonfish", package="PoEdata")
fishQ.ols <- lm(lquan~mon+tue+wed+thu+stormy, data=fultonfish)
kable(tidy(fishQ.ols), digits=4,
      caption="Reduced 'Q' equation for the fultonfish example")
```

Table 4.3: Reduced 'Q' equation for the fultonfish example

term	estimate	std.error	statistic	p.value
(Intercept)	8.8101	0.1470	59.9225	0.0000
mon	0.1010	0.2065	0.4891	0.6258

term	estimate	std.error	statistic	p.value
tue	-0.4847	0.2011	-2.4097	0.0177
wed	-0.5531	0.2058	-2.6875	0.0084
thu	0.0537	0.2010	0.2671	0.7899
stormy	-0.3878	0.1437	-2.6979	0.0081

```
fishP.ols <- lm(lprice~mon+tue+wed+thu+stormy, data=fultonfish)
kable(tidy(fishP.ols), digits=4,
      caption="Reduced 'P' equation for the fultonfish example")
```

Table 4.4: Reduced 'P' equation for the fultonfish example

term	estimate	std.error	statistic	p.value
(Intercept)	-0.2717	0.0764	-3.5569	0.0006
mon	-0.1129	0.1073	-1.0525	0.2950
tue	-0.0411	0.1045	-0.3937	0.6946
wed	-0.0118	0.1069	-0.1106	0.9122
thu	0.0496	0.1045	0.4753	0.6356
stormy	0.3464	0.0747	4.6387	0.0000

```
fish.D <- lquan~lprice+mon+tue+wed+thu
fish.S <- lquan~lprice+stormy
fish.eqs <- list(fish.D, fish.S)
fish.ivs <- ~mon+tue+wed+thu+stormy
fish.sys <- systemfit(fish.eqs, method="2SLS",
              inst=fish.ivs, data=fultonfish)
summary(fish.sys)
```

```
systemfit results
method: 2SLS

         N  DF     SSR  detRCov   OLS-R2 McElroy-R2
system 222 213 109.612 0.107301 0.094242  -0.597812

      N  DF     SSR      MSE     RMSE       R2   Adj R2
eq1 111 105 52.0903 0.496098 0.704342 0.139124 0.098130
eq2 111 108 57.5218 0.532610 0.729801 0.049360 0.031755
```

```
The covariance matrix of the residuals
         eq1      eq2
eq1 0.496098 0.396138
eq2 0.396138 0.532610

The correlations of the residuals
         eq1      eq2
eq1 1.000000 0.770653
eq2 0.770653 1.000000


2SLS estimates for 'eq1' (equation 1)
Model Formula: lquan ~ lprice + mon + tue + wed + thu
Instruments: ~mon + tue + wed + thu + stormy

              Estimate Std. Error  t value   Pr(>|t|)
(Intercept)  8.5059113  0.1661669 51.18896 < 2.22e-16 ***
lprice      -1.1194169  0.4286450 -2.61152  0.0103334 *
mon         -0.0254022  0.2147742 -0.11827  0.9060766
tue         -0.5307694  0.2080001 -2.55177  0.0121574 *
wed         -0.5663511  0.2127549 -2.66199  0.0089895 **
thu          0.1092673  0.2087866  0.52334  0.6018373
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.704342 on 105 degrees of freedom
Number of observations: 111 Degrees of Freedom: 105
SSR: 52.090321 MSE: 0.496098 Root MSE: 0.704342
Multiple R-Squared: 0.139124 Adjusted R-Squared: 0.09813


2SLS estimates for 'eq2' (equation 2)
Model Formula: lquan ~ lprice + stormy
Instruments: ~mon + tue + wed + thu + stormy

               Estimate  Std. Error  t value Pr(>|t|)
(Intercept)  8.62835440  0.38897023 22.18256  < 2e-16 ***
lprice       0.00105931  1.30954697  0.00081  0.99936
stormy      -0.36324606  0.46491248 -0.78132  0.43632
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.729801 on 108 degrees of freedom
```

```
Number of observations: 111 Degrees of Freedom: 108
SSR: 57.521843 MSE: 0.53261 Root MSE: 0.729801
Multiple R-Squared: 0.04936 Adjusted R-Squared: 0.031755
```

5 Dynamic Panel Data Model

5.1 Introduction

Linear dynamic panel data models account for dynamics and unobserved individual-specific heterogeneity. Due to the presence of lagged dependent variables, applying ordinary least squares including individual-specific dummy variables is inconsistent.

5.2 Model Specification

$$y_{it} = \alpha y_{i,t-1} + \beta x_{i,t} + u_{i,t}$$
$$u_{i,t} = \eta_i + \varepsilon_{i,t}$$
$$y_{it} = \alpha y_{i,t-1} + \beta x_{i,t} + \eta_i + \varepsilon_{i,t}$$

First Difference to elliminate η_i

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \beta \Delta x_{i,t} + \Delta \varepsilon_{i,t}$$

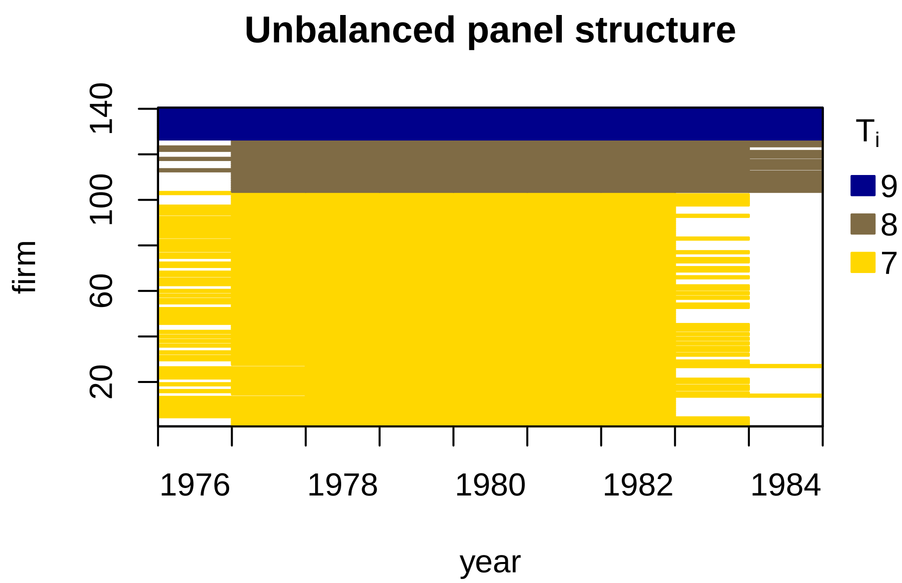
5.3 R implementation for Arrelano and Bond (1991)

n=140 firms T=9

```
library(plm)
library(pdynmc)

# Load Data
data(EmplUK, package = "plm")
dat <- EmplUK
dat[,c(4:7)] <- log(dat[,c(4:7)])
names(dat)[4:7] <- c("n", "w", "k", "ys")
data.info(dat,i.name = "firm",t.name = "year")
```

```
Unbalanced panel data set with 1031 rows and the following time period frequencies:

1976 1977 1978 1979 1980 1981 1982 1983 1984
  80  138  140  140  140  140  140   78   35
```

```
strucUPD.plot(dat,i.name = "firm",t.name = "year")
```



5.4 Estimation

```
m1 <- pdynmc(
    dat = dat,
    varname.i = "firm",
    varname.t = "year",
    use.mc.diff = TRUE,
    use.mc.lev = FALSE,
    use.mc.nonlin = FALSE,
    include.y = TRUE,
    varname.y = "n",
    lagTerms.y = 2,
    fur.con = TRUE,
```

```r
    fur.con.diff = TRUE,
    fur.con.lev = FALSE,
    varname.reg.fur = c("w", "k", "ys"),
    lagTerms.reg.fur = c(1,2,2),
    include.dum = TRUE,
    dum.diff = TRUE,
    dum.lev = FALSE,
    varname.dum = "year",
    w.mat = "iid.err",
    std.err = "corrected",
    estimation = "twostep",
    opt.meth = "none"
)
summary(m1)
```

```
Dynamic linear panel estimation (twostep)
GMM estimation steps: 2

Coefficients:
      Estimate Std.Err.rob z-value.rob Pr(>|z.rob|)
L1.n   0.62871     0.19341       3.251      0.00115 **
L2.n  -0.06519     0.04505      -1.447      0.14790
L0.w  -0.52576     0.15461      -3.401      0.00067 ***
L1.w   0.31129     0.20300       1.533      0.12528
L0.k   0.27836     0.07280       3.824      0.00013 ***
L1.k   0.01410     0.09246       0.152      0.87919
L2.k  -0.04025     0.04327      -0.930      0.35237
L0.ys  0.59192     0.17309       3.420      0.00063 ***
L1.ys -0.56599     0.26110      -2.168      0.03016 *
L2.ys  0.10054     0.16110       0.624      0.53263
1979   0.01122     0.01168       0.960      0.33706
1980   0.02307     0.02006       1.150      0.25014
1981  -0.02136     0.03324      -0.642      0.52087
1982  -0.03112     0.03397      -0.916      0.35967
1983  -0.01799     0.03693      -0.487      0.62626
1976  -0.02337     0.03661      -0.638      0.52347
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 41 total instruments are employed to estimate 16 parameters
 27 linear (DIF)
```

```
 8 further controls (DIF)
 6 time dummies (DIF)

J-Test (overid restrictions):  31.38 with 25 DF, pvalue: 0.1767
F-Statistic (slope coeff):  269.16 with 10 DF, pvalue: <0.001
F-Statistic (time dummies):  15.43 with 6 DF, pvalue: 0.0172
```

5.5 Hypotheses Testing

5.5.1 Arrelano Bond Serial Correlation

```r
mtest.fct(m1,t.order = 2)
```

```

    Arellano and Bond (1991) serial correlation test of degree 2

data:  2step GMM Estimation
normal = -0.36744, p-value = 0.7133
alternative hypothesis: serial correlation of order 2 in the error terms
```

The test does not reject the null hypothesis at any plausible significance level and provides no indication that the model specification might be inadequate.

5.5.2 Hansen J-test

```r
jtest.fct(m1)
```

```

    J-Test of Hansen

data:  2step GMM Estimation
chisq = 31.381, df = 25, p-value = 0.1767
alternative hypothesis: overidentifying restrictions invalid
```

5.5.3 Wald Test

```
wald.fct(m1,param = "all")
```

```

	Wald test

data:  2step GMM Estimation
chisq = 1104.7, df = 16, p-value < 2.2e-16
alternative hypothesis: at least one time dummy and/or slope coefficient is not equal to zer
```

6 Summary

- Modelling itu merupakan seni
- "All Models are Wrong, but some are useful" - George Box (Statistician)
- Data terdiri atas 3:
    - Cross Section
    - Time Series
    - Panel
- Penggunaan Jenis data yang berbeda akan membawa implikasi penggunaan Alat Statistik yang berbeda pula.

References