Case Study I

Case 1. Weekly egg prices at a German agricultural market between April 1967 and May 1990

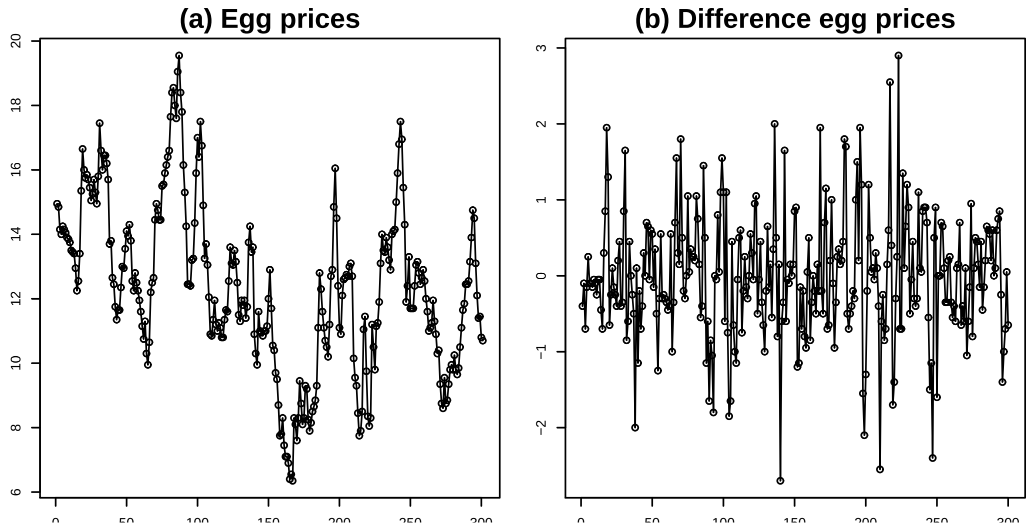
Case 2. GE daily returns for GE common stock from December 1999 to December 2000.

Case 3. The log series of quarterly earning per share of Johnson and Johson from 1960 to 1980.

Case 4. The monthly simple returns of the CRSP Decile 1 index from January 1960 to December 2003 for 528 observations.

Case 5. The 1-year and 3-year U.S. treasury constant maturity rates.

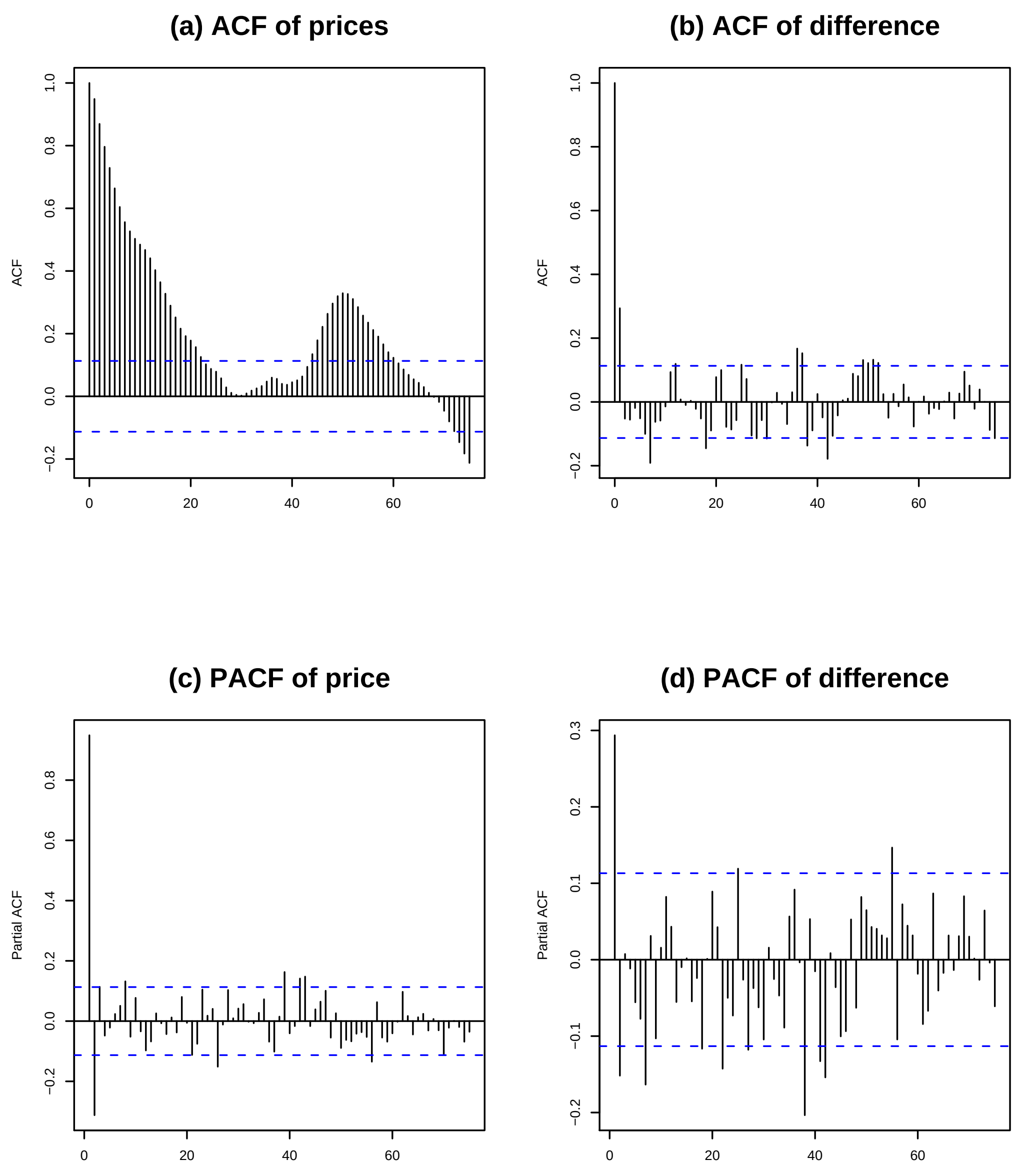
1. German Egg Prices



- The sample mean and variance are 12.38 and 6.77, respectively.

- The data exhibit a clear nonstationary feature. Take the first-order difference of the series, which looks more stationary like.

- Figures of autocorrelation and partial autocorrelation suggests that ARMA(p,q) model with $p \leq 7$ and $q \leq 7$.


(a) ACF of prices
ACF
(b) ACF of difference
ACF
(c) PACF of price
Partial ACF
(d) PACF of difference
Partial ACF

- The optimal AR model based on AICC is AR(7) with the AICC-value 698.24. The estimated AR-coefficient $\hat{b}_1, \ldots, \hat{b}_7$ are

	1	2	3	4	5	6	7
$\hat{b}_j$	0.322	-0.159	0.021	-0.004	-0.055	-0.023	-0.163
$\hat{b}_j/\{\mathrm{SE}(\hat{b}_j)\}$	5.651	-2.666	0.035	-0.071	-0.906	-0.378	-2.869

- The fitted model

$$X_t = 0.321X_{t-1} - 0.160X_{t-2} - 0.057X_{t-5} - 0.023X_{t-6} - 0.165X_{t-7} + \varepsilon_t,$$

where $\varepsilon_t \sim \mathcal{N}(0, 0.567)$ and the corresponding AICC-value is 694.34

- The MA(7) model is also used to fit the data. The estimated MA-coefficient $\hat{a}_1, \ldots, \hat{b}_7$ are

	1	2	3	4	5	6	7
$\hat{a}_j$	0.320	-0.038	-0.054	-0.023	-0.048	-0.046	-0.195
$\hat{a}_j/\{\text{SE}(\hat{a}_j)\}$	5.541	-0.629	-0.896	-0.386	-0.790	-0.757	-3.210

- The fitted model

$$X_t = \varepsilon_t - 0.345\varepsilon_{t-1} - 0.173\varepsilon_{t-7},$$

where $\hat{\sigma}^2 = 0.570$ and the corresponding AICC-value is 689.34. The standard errors of the two coefficients in the model above is 0.054 and 0.051.

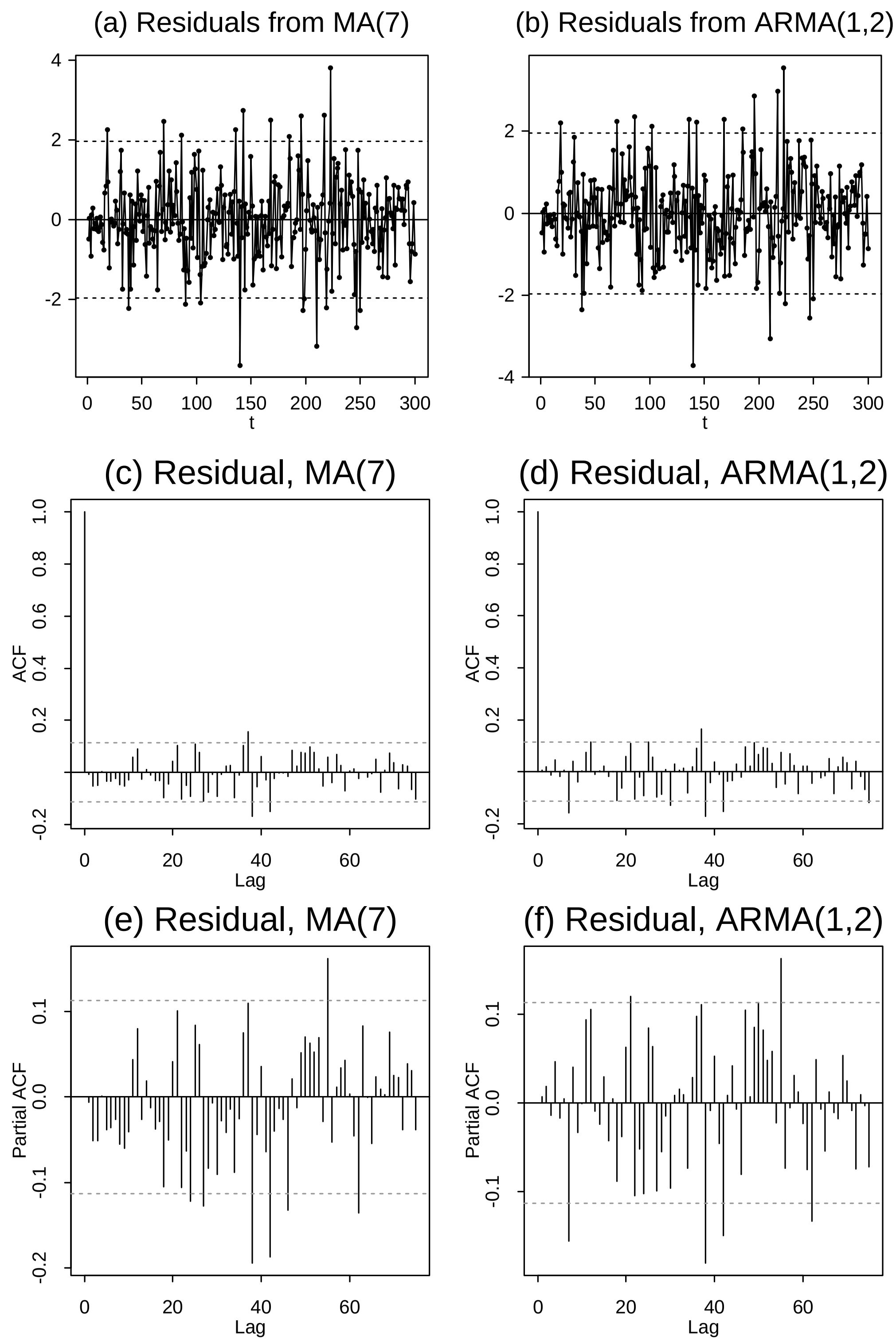
- The optimal ARMA model with $p = 1$ or 2 and $1 \leq q \leq 7$ based on AICC is the ARMA (1,2)

$$X_t = 0.906X_{t-1} + \varepsilon_t - 0.619\varepsilon_{t-1} - 0.381\varepsilon_{t-2},$$

with $\hat{\sigma}^2 = 0.563$ and the corresponding AICC-value is 690.58. The standard error for three coefficients in the model are 0.022, 0.053 and 0.052.

- According to AICC, both MA(7) and ARMA(1,2) are comparable with each other.

- From standard residuals and their ACF and PACF plots, slightly more than 5% (but$< 6\%$) of residuals from both modelsare beyond the bound ± 1.96. But ACF and PACF plots show that there still exists weak but significant autocorrelation in the residuals at some discrete lags. They failed in Portmanteau χ^2 Test.


(a) Residuals from MA(7)
(b) Residuals from ARMA(1,2)
t
(c) Residual, MA(7)
(d) Residual, ARMA(1,2)
ACF
Lag
(e) Residual, MA(7)
(f) Residual, ARMA(1,2)
Partial ACF

- One possible remedy is to include variables at the lags at which (partial) autocorrelation is significant. However it is in general difficult to interpret the resulting model.

- Converting MA(7) and ARMA(1,2) to the original egg price data $\{Y_t\}$, two competitive ARIMA models are obtained.

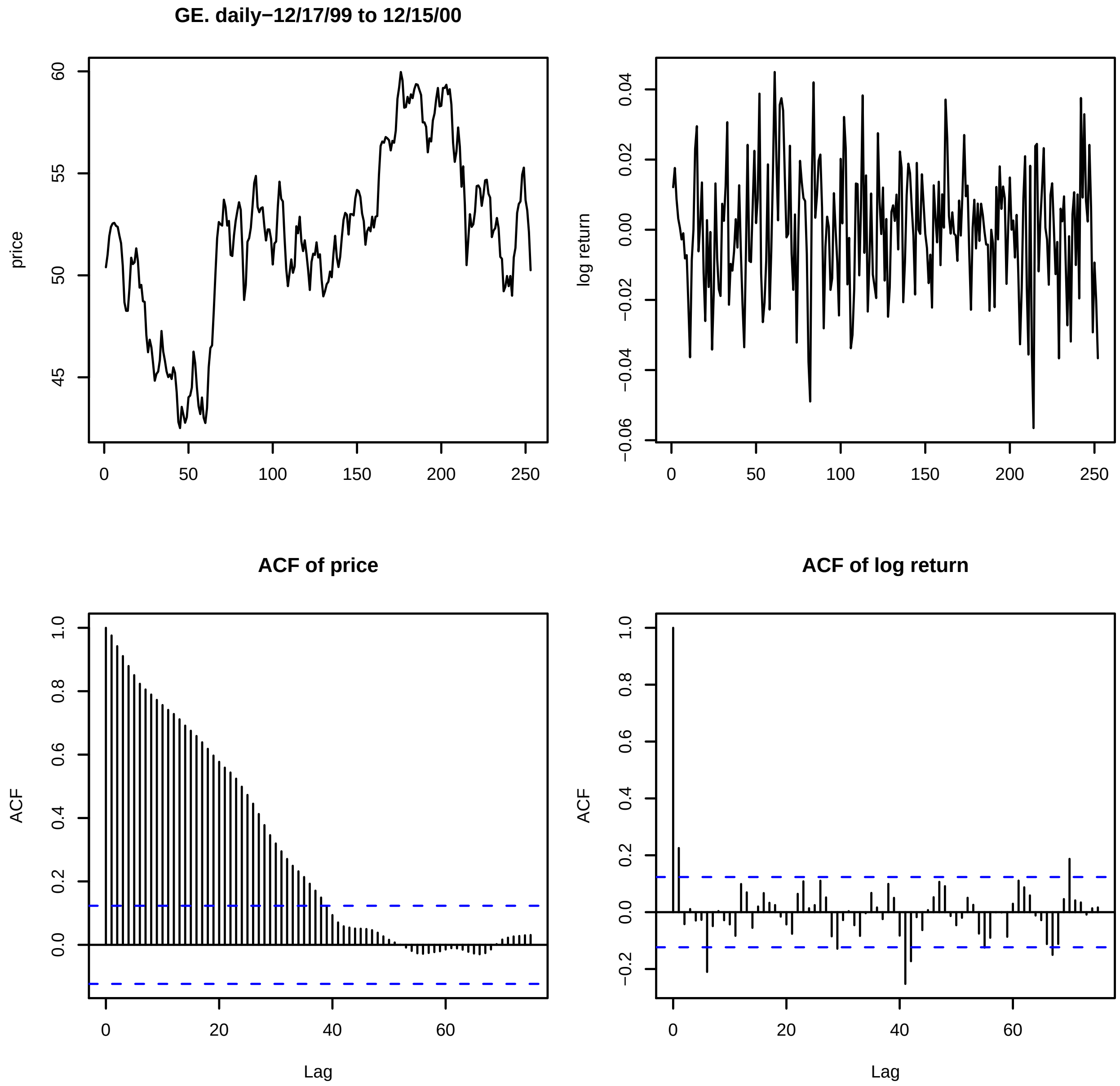
$$Y_t = Y_{t-1} + \varepsilon_t + 0.345\varepsilon_{t-1} - 0.173\varepsilon_{t-7}, \quad \{\varepsilon_t\} \sim \mathsf{WN}(0, 0.570),$$

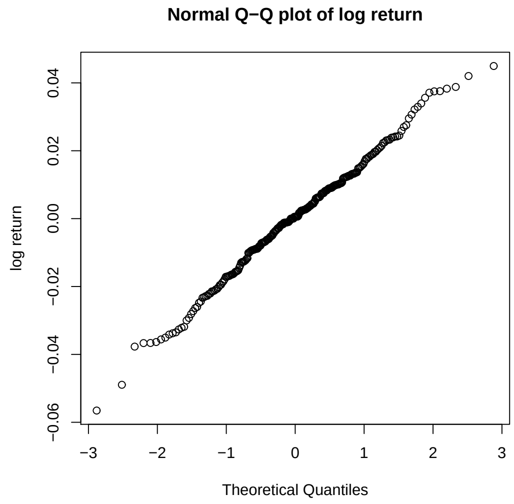
and

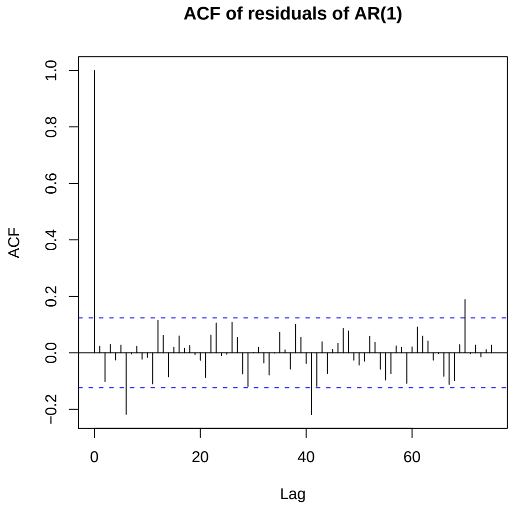
$$Y_t = -0.001 + 1.906Y_{t-1} - 0.906Y_{t-2} + \varepsilon_t - 0.619\varepsilon_{t-1} - 0.318\varepsilon_{t-2},$$

$\{\varepsilon_t\} \sim \mathsf{WN}(0, 0.563)$.

2. GE daily returns





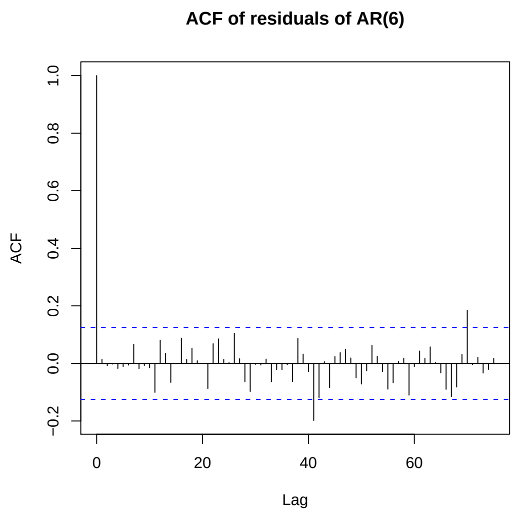


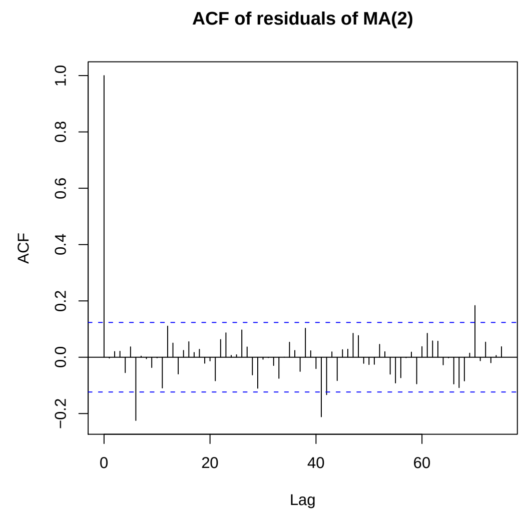
- AR(1) model

$$Y_t = \beta_0 - \phi Y_{t-1} + \varepsilon_t$$

- The estimate of β_0 is -0.00000361 and its standard deviation is 0.0014009. t-ratio is -0.03.
  The estimate of ϕ is 0.22943, the standard deviation is 0.06213, and t-ratio is 3.69.

- The Ljung-Box "simultaneous" χ^2 test that $\rho(1) = \cdots = \rho(12) = 0$ has $p = 0.0179$. Hence the AR(1) model does not fit well.





- AR(6) model

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_6(Y_{t-6} - \mu) + \varepsilon_t$$

	μ	1	2	3	4	5	6
$\hat{\phi}_j$	-2.106E-6	-0.2531	0.1257	-0.0714	-0.0748	-0.0520	0.2227
$\hat{\phi}_j/\{\mathrm{SE}(\hat{\phi}_j)\}$	-0.00	-4.06	1.96	-1.11	1.16	-0.81	-3.58

- The total R-square of AR(1) is 0.0000 and for AR(6) it is 0.1139.

- MA(2) model

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}.$$

- The estimate of MA(2) model

$$\begin{aligned}
\hat{\mu} &= -0.0000247(0.0012775) \\
\hat{\theta}_1 &= -0.26477(0.006362) \\
\hat{\theta}_2 &= 0.07617(0.06385)
\end{aligned}$$

and the Ljung-Box χ_p^2 statistics is 14.47 to lag 6, 21.25 to lag 12 and 24.12 to lag 18. The corresponding p-value is 0.0059, 0.0194 and 0.0868.

- ARMA(2,1) model

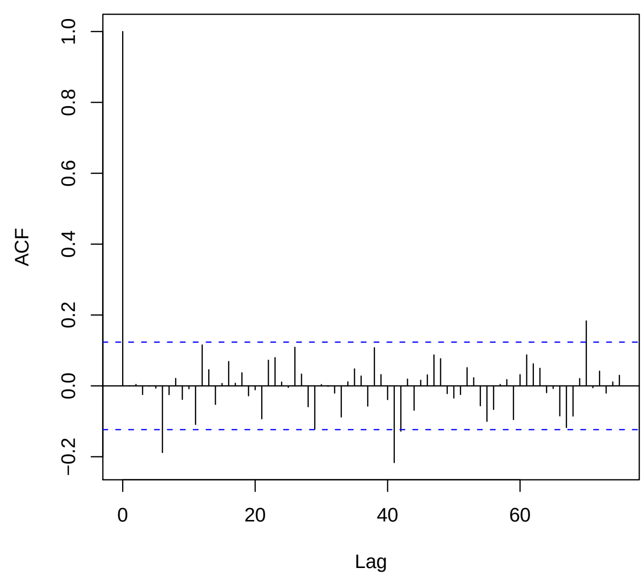
$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1}.$$

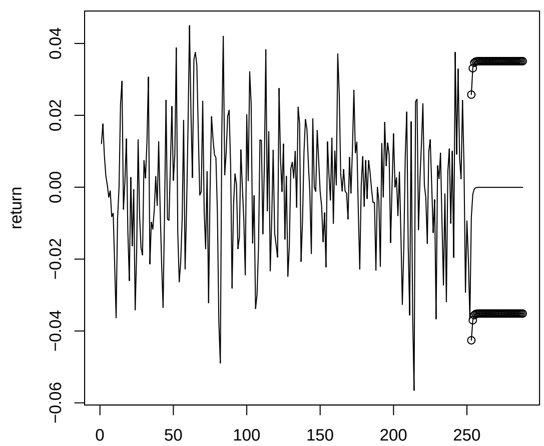
- The estimate of ARMA(2,1) model is

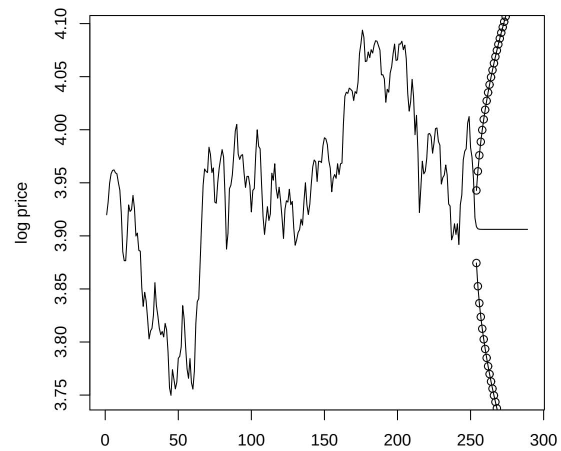
$$\begin{aligned}\hat{\mu} &= -0.0000217(0.0013272),\\ \hat{\phi}_1 &= -0.53313(0.16319),\\ \hat{\phi}_2 &= 0.07806(0.08953),\\ \hat{\theta}_1 &= -0.80566(0.14654)\end{aligned}$$

and the Ljung-Box χ_p^2 statistics is 9.31 to lag 6, 16.72 to lag 12 and 19.68 to lag 18. The corresponding p-value is 0.0254, 0.0534 and 0.1847.

ACF of residuals of ARMA(2,1)







- Model selection criterions:
  Akaike's information criterion (AIC) and Schwarz's Bayesian Criterion (SBC or BIC)

$$-2\log(L) + 2(p+q) \approx n\log(\hat{\sigma}^2) + 2(p+q) \quad \text{(AIC)},$$

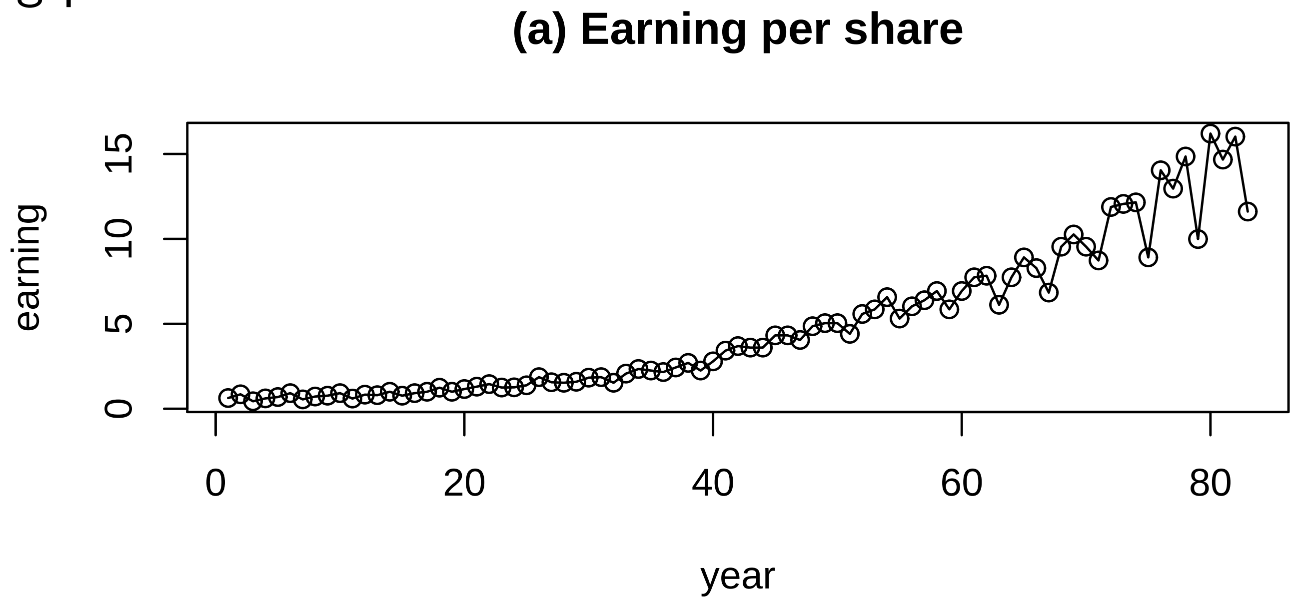
$$-2\log(L) + \log(n)(p+q) \approx n\log(\hat{\sigma}^2) + \log(n)(p+q) \quad \text{(SBC)}.$$

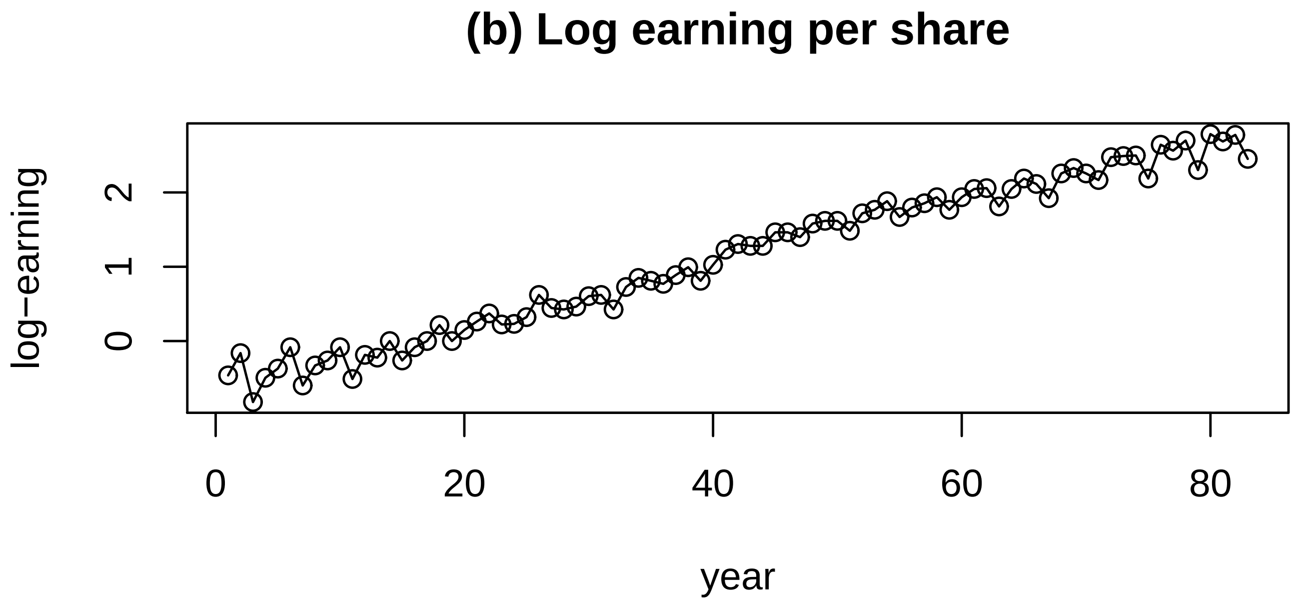
- GE daily log returns: choosing the AR order

p	AIC	SBC
1	0	0
2	-0.41	3.12
3	1.03	8.09
4	2.44	13.03
5	4.43	18.54
6	-7.04	10.61
7	-6.06	15.11
8	4.50	20.20

3. Seasonal Models

- Log earning per share of Johnson and Johnson



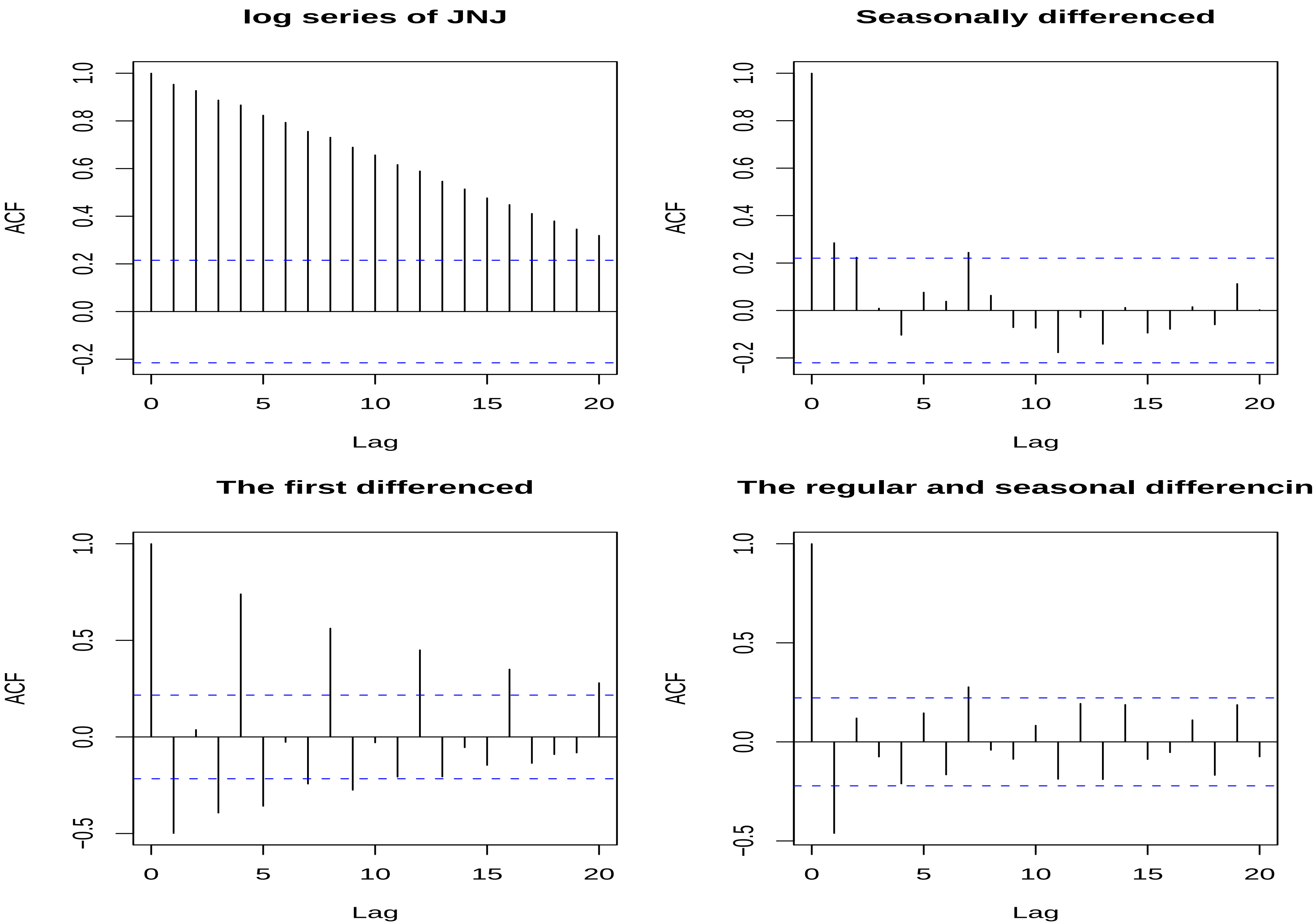


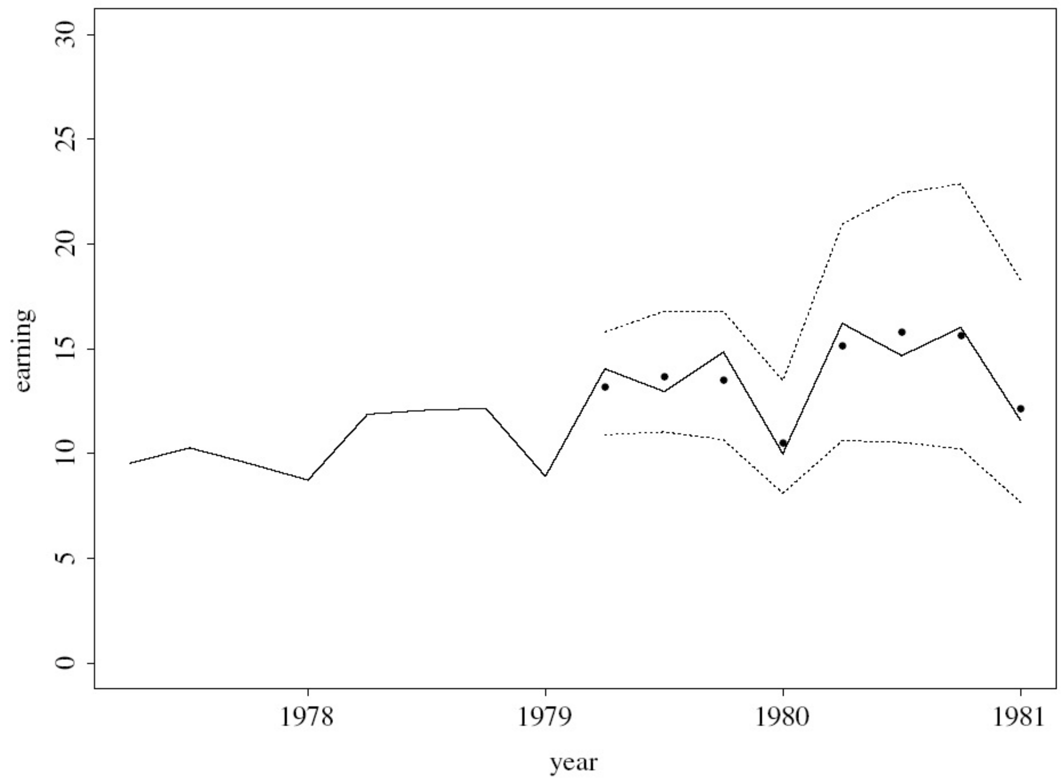
- Seasonal Differencing and Multiplicative Seasonal Models, for example,

$$(1 - B^s)(1 - B)X_t = (1 - \theta B)(1 - \Theta B^s)a_t$$

- Seasonal Model for Log earning per share of Johnson and Johnson.

$$(1 - B)(1 - B^4)X_t = (1 - 0.678B)(1 - 0.314B^4)a_t, \quad \hat{\sigma}_a = 0.089.$$


log series of JNJ
Seasonally differenced
The first differenced
The regular and seasonal differencin
ACF
Lag
0
5
10
15
20
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
-0.5
0.5



earning
0
5
10
15
20
25
30
1978
1979
1980
1981
year

- The monthly simple returns of the CRSP Decile 1 index from January 1960 to December 2003 for 528 observations.

- The fitted seasonal ARMA models by the conditional likelihood

$$(1 - 0.25B)(1 - 0.99B^{12})R_t = 0.0004 + (1 - 0.92B^{12})a_t, \hat{\sigma}_a = 0.071.$$

- The fitted seasonal ARMA models by the exact likelihood

$$(1 - 0.264B)(1 - 0.996B^{12})R_t = 0.0002 + (1 - 0.999B^{12})a_t, \hat{\sigma}_a = 0.067.$$

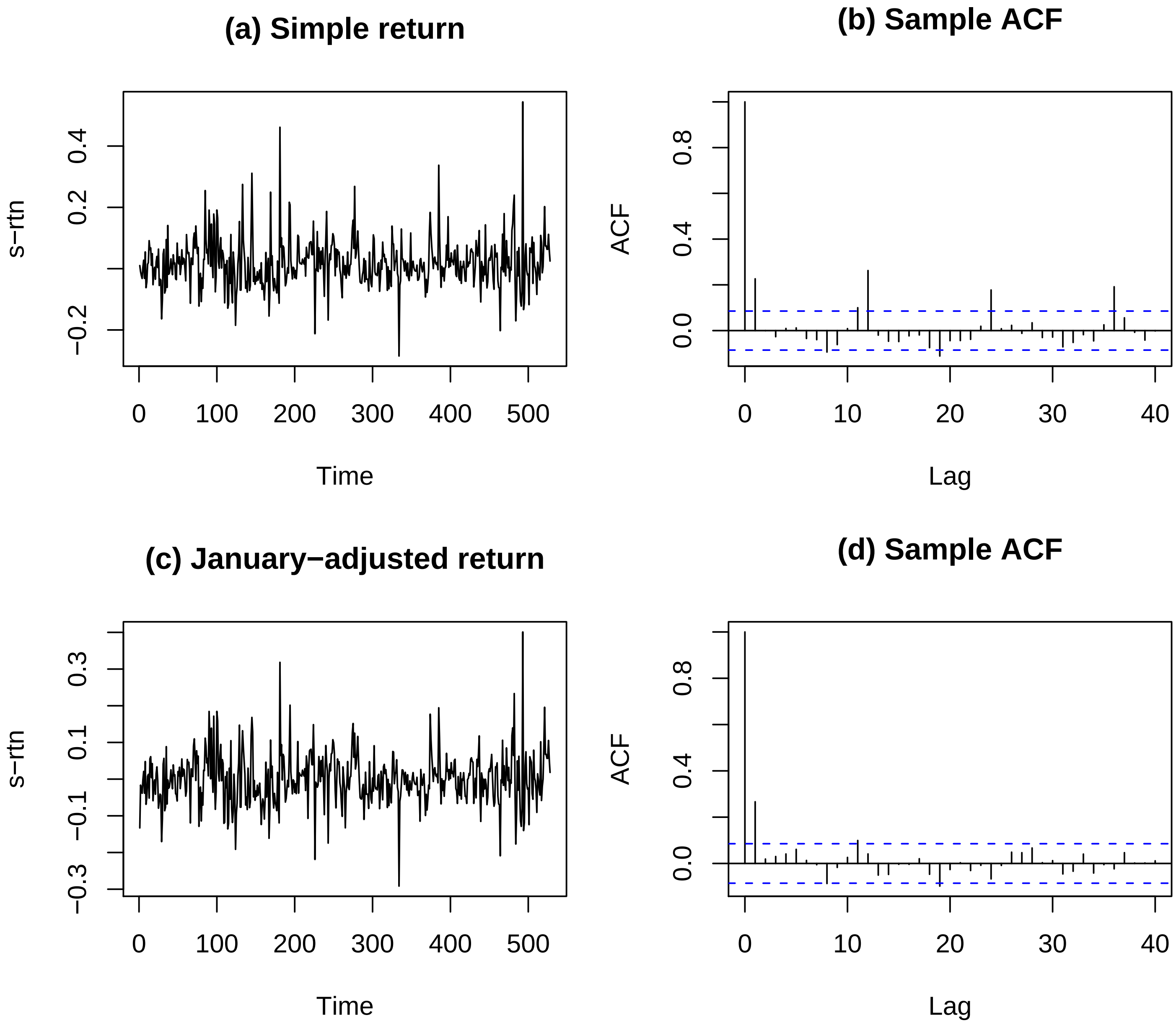
- The cancellation between seasonal AR and MA factors is clearly seen. The estimation results suggests that the seasonal behavior might be deterministic.

- January Effect: Employ the simple linear regression

$$R_t = \beta_0 + \beta_1 \mathsf{Jan}_t + e_t$$

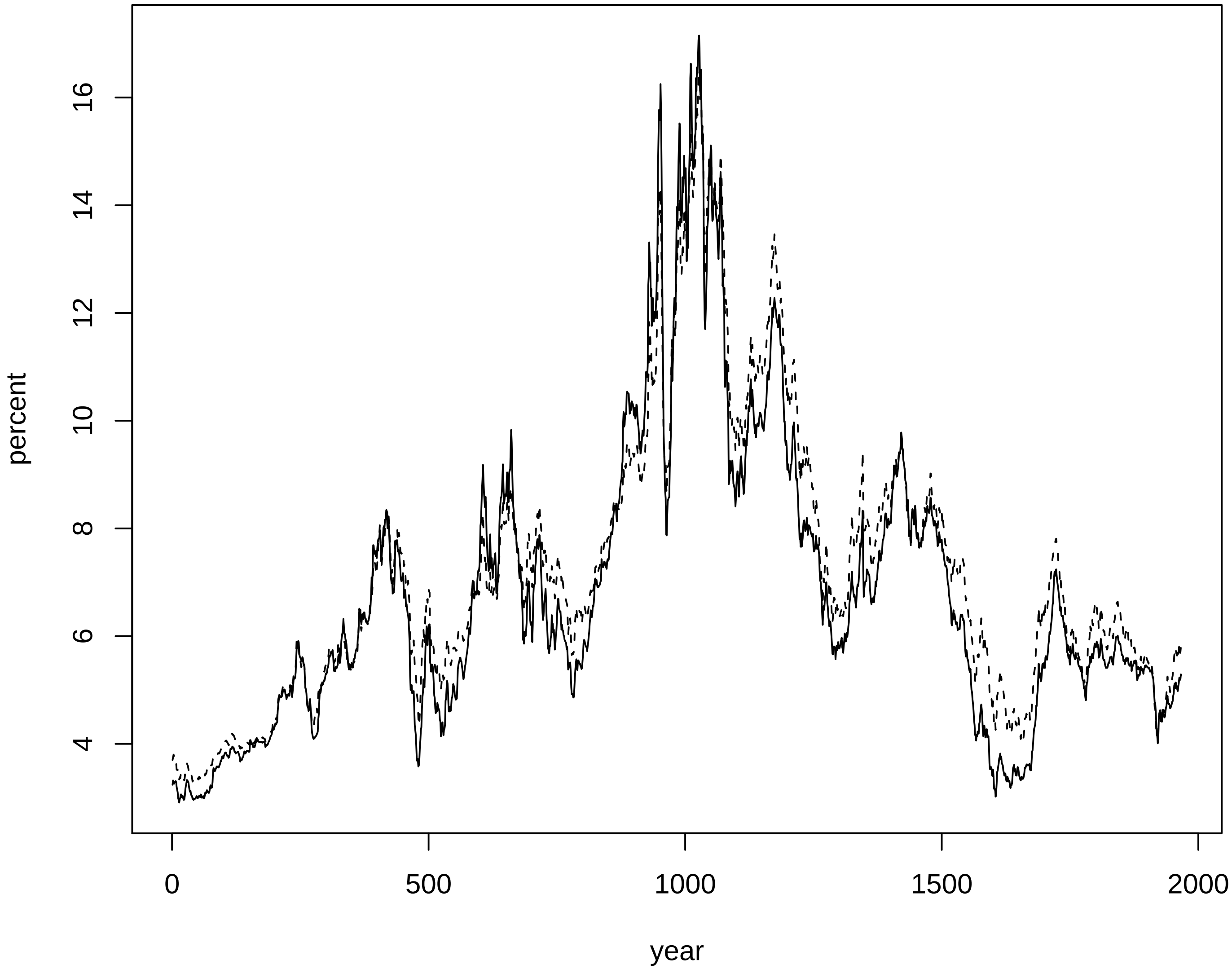
where

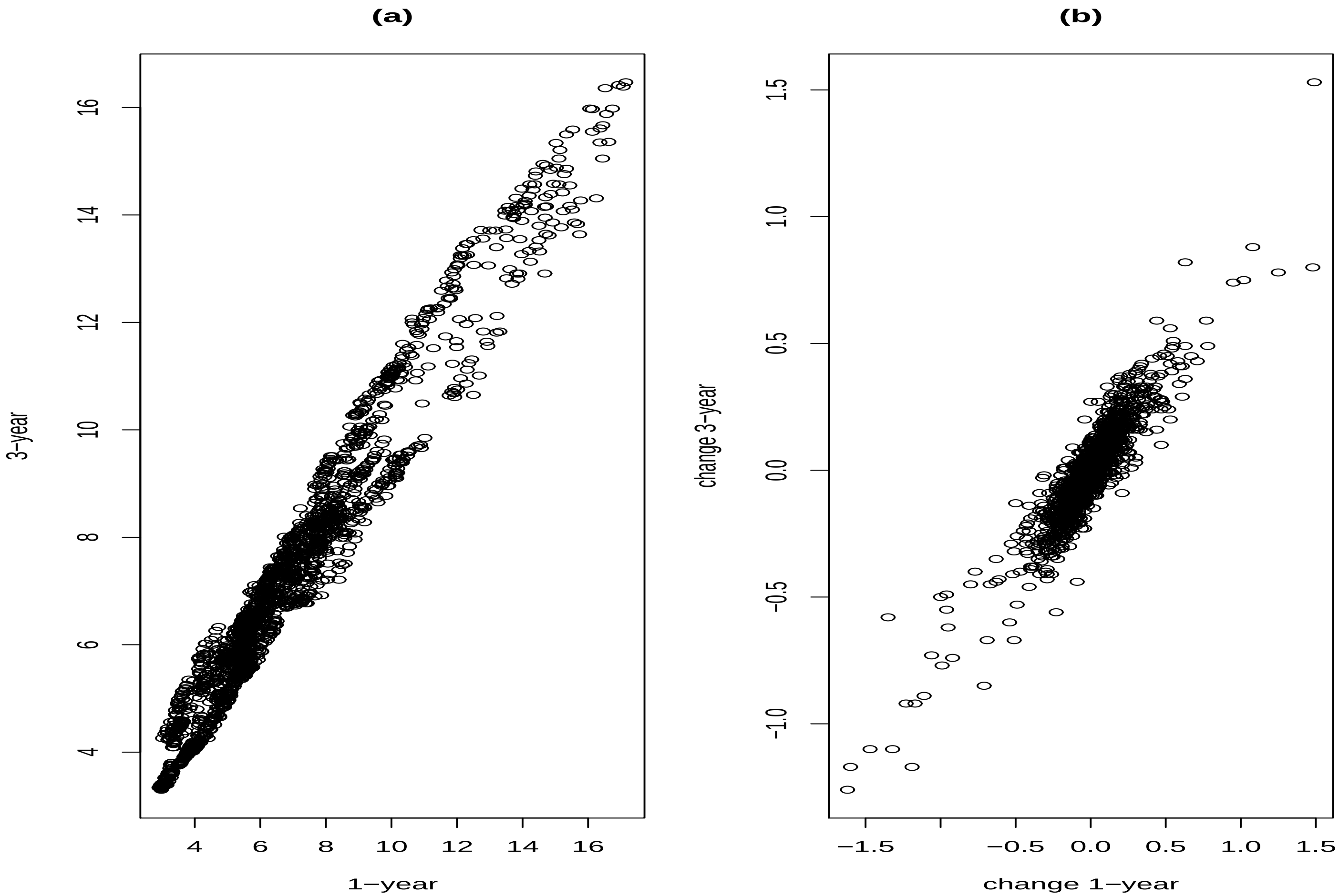
$$\mathsf{Jan}_t = \begin{cases} 1 & \text{if } t \text{ is January} \\ 0 & \text{otherwise.} \end{cases}$$


(a) Simple return
s−rtn
−0.2
0.2
0.4
0
100
200
300
400
500
Time
(b) Sample ACF
ACF
0.0
0.4
0.8
0
10
20
30
40
Lag
(c) January−adjusted return
s−rtn
−0.3
−0.1
0.1
0.3
0
100
200
300
400
500
Time
(d) Sample ACF
ACF
0.0
0.4
0.8
0
10
20
30
40
Lag

Regression Models with Time series errors

The 1-year and 3-year U.S. treasury constant maturity rates.

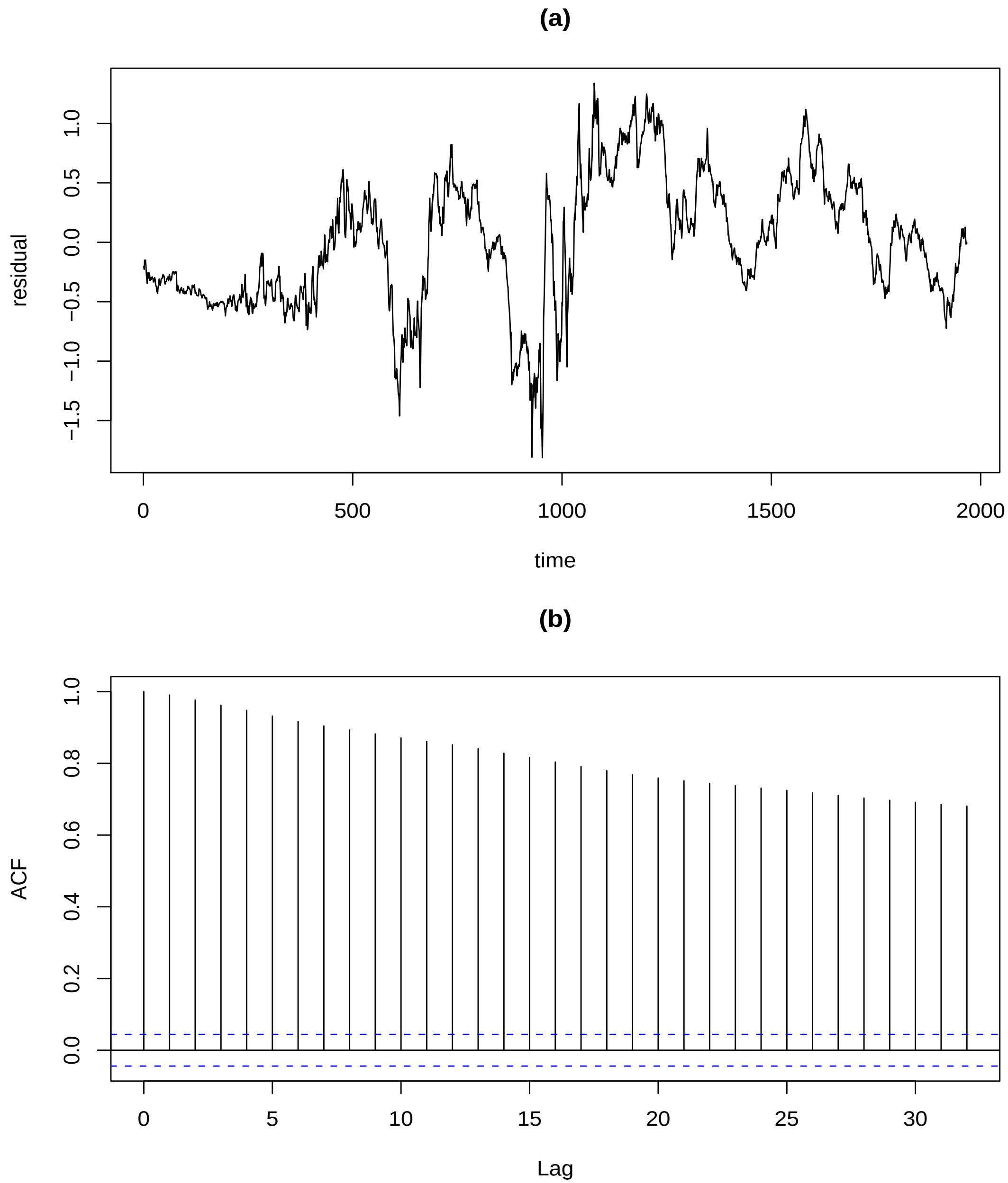


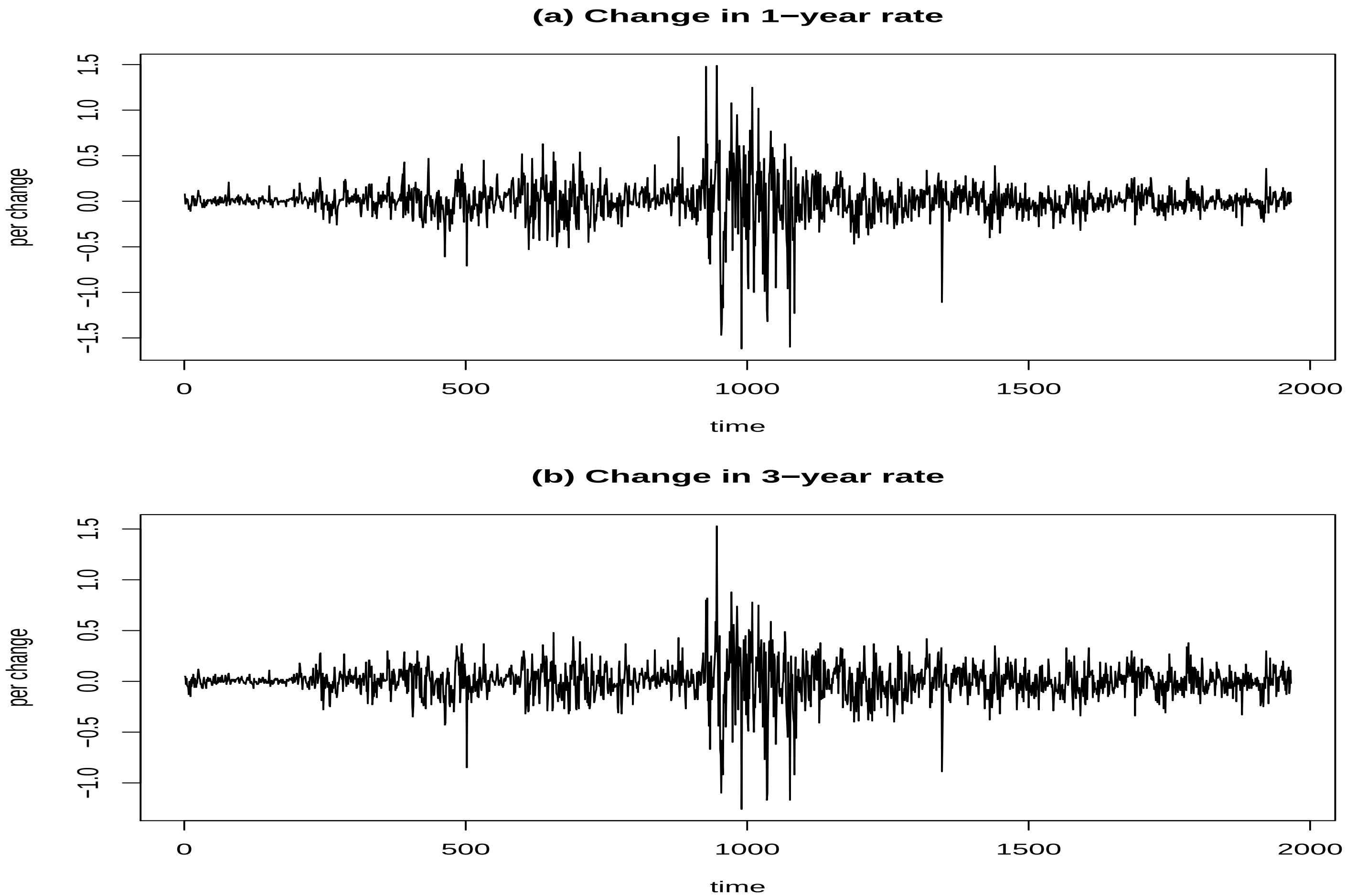


- The simple regression model between two rates $r_{3t} = \alpha + \beta r_{1t} + e_t$ results in a fitted model

$$r_{3t} = 0.911 + 0.924 r_{1t} + e_t, \quad \hat{\sigma}_e = 0.538,$$

with $R^2 = 95.8\%$, where the standard errors of the two coefficients are 0.032 and 0.004, respectively.


(a)
residual
−1.5
−1.0
−0.5
0.0
0.5
1.0
0
500
1000
1500
2000
time
(b)
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Lag



- Nonstationary of both interest rate and the residuals leads to consideration of the change series of interest rate.

$$
\begin{aligned}
c_{1t} &= r_{1t} - r_{1,t-1} = (1-B)r_{1t}\text{for}t \geq 2 : \text{changes in the 1-year interest rate} \\
c_{3t} &= r_{3t} - r_{3,t-1} = (1-B)r_{3t}\text{for}t \geq 2 : \text{changes in the 3-year interest rate}
\end{aligned}
$$

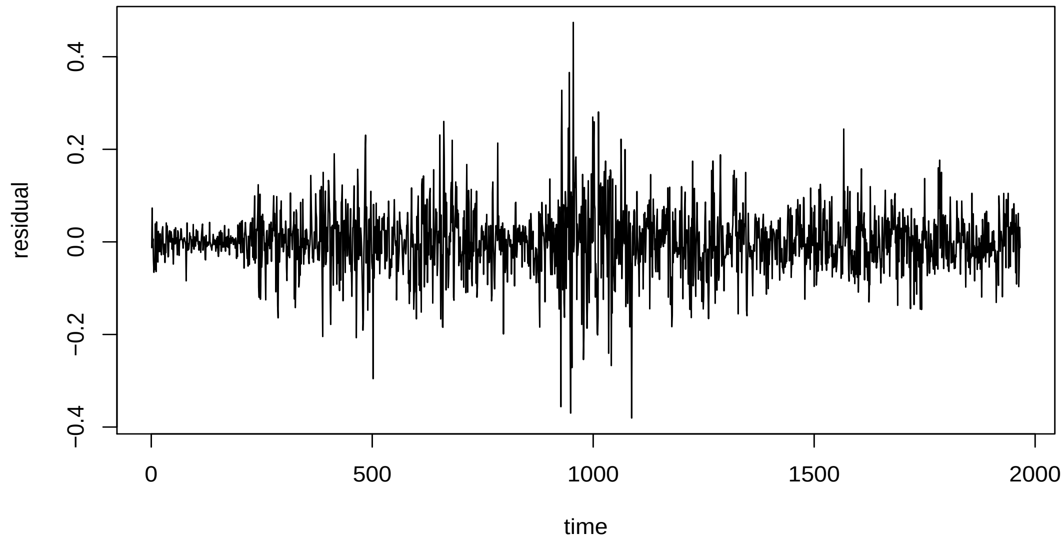
- Consider the linear regression $c_{3t} = \alpha + \beta c_{1t} + e_t$, the change series remain highly correlated with a fitted linear regression model given by

$$c_{3t} = 0.00002 + 0.7811 c_{1t} + e_t, \quad \hat{\sigma}_e = 0.0682,$$

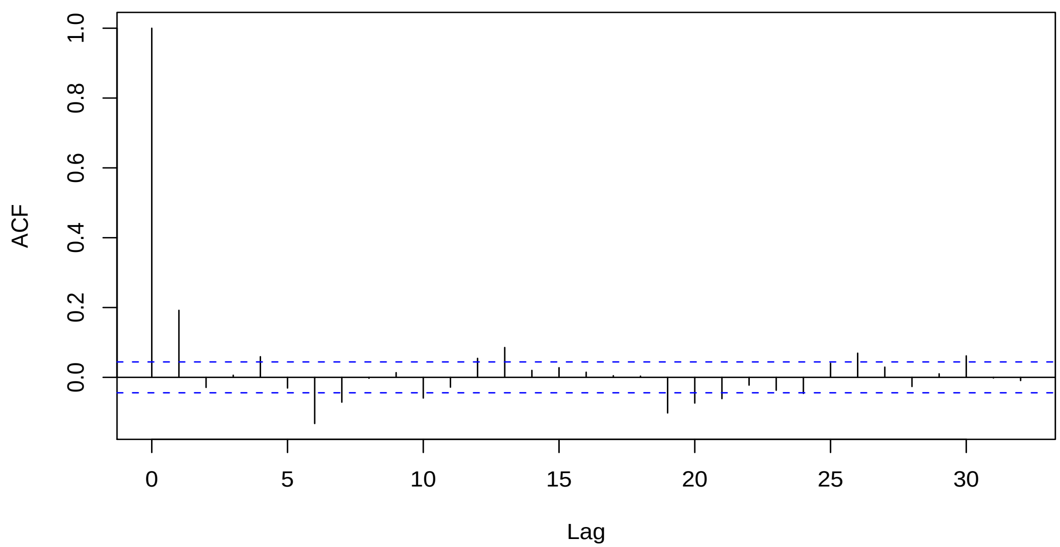
with $R^2 = 84.8\%$. The standard errors of the two coefficients are 0.0015 and 0.0075.

- The model further confirms the strong linear dependence between interest rates. However, the ACF shows some significant serial correlation in the residuals, but the magnitude of the correlation is much smaller.

(a)


residual
0.4
0.2
0.0
−0.2
−0.4
0
500
1000
1500
2000
time

(b)


ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

- The weak serial dependence in the residuals can be modeled by using the simple time series models. Because residuals of the mdoel are serial correlated, we shall identify a simple ARMA model for the residuals. From the figure of ACF of residual, MA(1) model was used for residuals, the linear regression model has been modified as

$$c_{3t} = \alpha + \beta c_{1t} + e_t, \quad e_t = a_t - \theta_1 a_{t-1}.$$

- The fitted version of the model is given by

$$c_{3t} = 0.0002 + 0.782 c_{1t} + e_t, \quad e_t = a_t + 0.2115 a_{t-1}, \quad \hat{\sigma}_a = 0.0668,$$

with $R^2 = 85.4\%$. The standard errors of the parameters are 0.0018, 0.0077 and 0.0221 respectively.

- The model no longer has a significant lag-1 residual ACF, even though some minor serial correlation remain at lag 4 and 6.

- the high R^2 and coefficient 0.924 of the first model are misleading because the residuals of the model show strong serial correlation.

- For the change series, R^2 and the coefficient of c_{1t} of the models are close. In this particular instance, adding MA(1) model to the change series only provides a marginal improvement. This is not surprising because the estimated MA coefficient is small numerically, even though it is statistically highly significant.

- The analysis demonstrates that it is important to check residual serial dependence in linear regression analysis.

- Because the constant term in the above equation is insignificant, the model shows that the two weekly interest rate series are related as

$$r_{3t} = r_{3,t-1} + 0.782(r_{1t} - r_{1,t-1}) + a_t + 0.212a_{t-1}.$$

The interest rates are concurrently and serially correlated.