# Financial Statistics
# Homework 2

Chenghua Liu
liuch18@mails.tsinghua.edu.cn
Department of Computer Science
Tsinghua University

## 1 2.7

After fitting an AR(3) model to the monthly log-return of CRSP data from Jan. 1926 to Dec. 1997 , it was obtained that

$$\hat{b}_0 = 0.0103, \hat{b}_1 = 0.104, \hat{b}_2 = -0.010, \hat{b}_3 = -0.120$$

with the estimated covariance matrix as follows:

$$\mathbf{S} = 1000^{-2} \begin{pmatrix} 2^2 & 34 & 0 & 0 \\ 34 & 34^2 & 0 & 0 \\ 0 & 0 & 34^2 & 0 \\ 0 & 0 & 0 & 34^2 \end{pmatrix}$$

(a) What is the standard error of $\hat{b}_1$?

(b) Test $H_0 : b_1 = 0$ at significance level 1%.

(c) The annual return is estimated as $\hat{r} = (1 + \hat{\mu})^{12} - 1$, where $\hat{\mu} = \hat{b}_0 / \left(1 - \hat{b}_1 - \hat{b}_2 - \hat{b}_3\right)$. Construct a 95% confidence interval of the annual return (computing directly from the information given above).

(d) Obtain directly the standard error of the annual return $\hat{r}$ if $\mathrm{SE}(\hat{\mu}) = 0.002186$ was already computed.

**Solution:**

**(a)**

We can get $\mathrm{var}(\hat{b}_1) = \frac{34^2}{1000^2}$ from the estimated covariance matrix, so the standard error of $\hat{b}_1$ is $\sqrt{\mathrm{var}(\hat{b}_1)} = 0.034$

**(b)**

To test the hypothesis $H_0 : b_1 = 0$ , the t-statisic is

$$T = \hat{b}_1/\mathrm{SE}(\hat{b}_1) = 0.104/0.034 = 3.058824$$

Against the two sided alternative $H_1 : b_1 \neq 0$ is $2(1 - \Phi(3.058824)) = 0.002222077$, so we conclude that $b_1 \neq 0$ at significance level 1%.

**(c)**

$$\hat{\mu} = \hat{b}_0 / \left(1 - \hat{b}_1 - \hat{b}_2 - \hat{b}_3\right) = \frac{0.0103}{1 - 0.104 + 0.010 + 0.120} = 0.01003899$$

The annual return is

$$\hat{r} = (1 + \hat{\mu})^{12} - 1 = \left(1 + \frac{\hat{b}_0}{1 - \hat{b}_1 - \hat{b}_2 - \hat{b}_3}\right)^{12} - 1 = 12.73473\%$$

Using the Delta method, we have

$$\mathrm{var}(\hat{r}) \approx \left[f(\boldsymbol{\theta})'\right]^T \boldsymbol{S} f(\boldsymbol{\theta})' = 857.5538 \times 10^{-6}$$

Where the $\hat{r} = f(\hat{\boldsymbol{\theta}})$ as a function of $\hat{\boldsymbol{\theta}} = (\hat{b_0}, \hat{b_1}, \hat{b_2}, \hat{b_3})^T$ . Therefore,

$$\mathrm{SE}(\hat{r}) = \sqrt{\mathrm{var}(\hat{\mu})} = 0.02928402$$

Consequently an approximate 95% confidence interval for the expected annual return is

$$12.73473\% \pm 1.96 \times 2.928402\% = [6.995062\%, 18.4744\%]$$

```
a=12*1.010039^11
b=1/(1-0.104+0.010+0.120)
c=0.0103/(1-0.104+0.010+0.120)^2
f=matrix(c(b,c,c,c),nrow=4, ncol=1)
D=matrix(c(4,34,0,0,34,34^2,0,0,0,0,34^2,0,0,0,0,34^2),nrow=4, ncol=4)
D=D*0.000001
var=t((a*f))%*%D%*%(a*f)
se=sqrt(var)
```

**(d)**

Using the Delta method, we have

$$\mathrm{var}(\hat{r}) \approx \left(12(1 + \hat{\mu})^{11}\right)^2 \mathrm{var}(\hat{\mu})$$

Therefore,

$$\mathrm{SE}(\hat{r}) = \left(12(1 + \hat{\mu})^{11}\right) \mathrm{SE}(\hat{\mu}) = 0.02927865$$

■

# 2 2.12

Suppose the prices of a stock follow the AR(1) model: $X_t = \mu + \rho X_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\} \sim_{\text{i.i.d.}} N(0, \sigma^2)$ and $\sigma$ is unknown.

(a) Derive the conditional maximum likelihood estimator of $\tilde{\mu}$ and $\tilde{\rho}$ for $\mu$ and $\rho$. (Here, the conditional likelihood refers to the density of $(X_2, \cdots, X_T)$ given $X_1$ ).

(b) Use simulation ( 1000 times with the initial value $X_0 = 0$ ) to calculate the 95% tile and 99% -tile of the null distributions of the Dickey-Fuller tests (without drift and with drift) for $T = 100$ and 400 . For concreteness, set $\mu = 0$ and $\sigma = 1$ in your simulation experiment.

**Solution:**
**(a)** Given the data $(x_1, \ldots, x_n)$ , we seek the likelihood

$$L\left(\mu, \phi, \sigma^2\right) = f\left(x_1, \ldots, x_n \mid \mu, \phi, \sigma^2\right)$$

In this case, we may write the likelihood as

$$L\left(\mu, \phi, \sigma^2\right) = f\left(x_1\right) f\left(x_2 \mid x_1\right) \cdots f\left(x_n \mid x_{n-1}\right) = f\left(x_1\right) \prod_{t=2}^{n} f\left(x_t \mid x_{t-1}\right)$$

Since $\left.x_t\right|_{x_{t-1}} \sim \mathcal{N}\left(\mu + \rho x_{t-1}, \sigma^2\right)$ , we have

$$f\left(x_t \mid x_{t-1}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left[x_t - \mu - \rho x_{t-1}\right]^2}{2\sigma^2}\right\}$$

When given $X_1$, the conditional likelihood becomes

$$L\left(\mu, \phi, \sigma^2 \mid x_1\right) = \prod_{t=2}^{n} f\left(x_t \mid x_{t-1}\right) = \left(2\pi\sigma^2\right)^{-\frac{(n-1)}{2}} \exp\left\{-\frac{S_c(\mu, \rho)}{2\sigma^2}\right\}$$

where the conditional sum of square is

$$S_c(\mu, \rho) = \sum_{t=2}^{n} \left[x_t - \left(\mu + \rho x_{t-1}\right)\right]^2$$

The problem is the linear regression problem. Let

$$\bar{y}_{(1)} = \frac{1}{n-1} \sum_{t=1}^{n-1} y_t \qquad \bar{y}_{(2)} = \frac{1}{n-1} \sum_{t=2}^{n} y_t$$

Then we have

$$\hat{\mu} = \bar{y}_{(2)} - \hat{\rho}\bar{y}_{(1)} \qquad \hat{\rho} = \frac{\sum_{t=2}^{n}\left(y_t - \bar{y}_{(2)}\right)\left(y_{t-1} - \bar{y}_{(1)}\right)}{\sum_{t=2}^{n}\left(y_{t-1} - \bar{y}_{(1)}\right)^2}$$

**(b)**
**Without drift:**

The limiting null distribution of the "Dickey-Fuller t-statistic" is that of a nonstandard random variable which may be expressed through a functional of a Brownian motion $W$.

Denote by $\hat{\rho}_T$ the OLS estimate of a regression of $x_t$ on $x_{t-1}$, and by $t_T$ the standard t-ratio for the null that $\rho = 1$. In the simplest case without constant or trend in the test regression, we have

$$\begin{aligned} t_T &= \frac{\hat{\rho}_T - 1}{s.e.\left(\hat{\rho}_T\right)} \\ &= \frac{T\left(\hat{\rho}_T - 1\right)}{\left\{s_T^2\right\}^{1/2}} \left\{T^{-2} \sum_{t=1}^{T} x_{t-1}^2\right\}^{1/2} \\ &\Rightarrow \frac{1/2\left\{W(1)^2 - 1\right\}}{\int_0^1 W(r)^2 dr} \frac{1}{\sigma} \left\{\sigma^2 \int_0^1 W(r)^2 dr\right\}^{1/2} \\ &= \frac{W(1)^2 - 1}{2\left\{\int_0^1 W(r)^2 dr\right\}^{1/2}} \end{aligned}$$

Noting that, for a suitable distribution of $u$ like the standard normal, $1/\sqrt{T}\sum_{t=1}^{[sT]} u_t$ will behave like $W(s)$ for $T$ "large", where $[sT]$ denotes the integer part of $sT$. Hence, the DF distribution can be simulated as follows:

```
T <- 100
reps <- 1000
DFstats <- rep(NA, reps)
for (i in 1:reps){
        u <- rnorm(T)
        W <- 1/sqrt(T)*cumsum(u)
        DFstats[i] <- (W[T]^2-1)/(2*sqrt(mean(W^2)))
}
quantile(DFstats, probs = c(0.01,0.05))
```

Change T to 400, we get

| *Leval* $\alpha$ | $\alpha = 0.01$ | $\alpha = 0.05$ |
|---|---|---|
| $T = 100$ | $-2.512817$ | $-1.965358$ |
| $T = 400$ | $-2.456848$ | $-1.890859$ |

**With drift:**

In this case ,we use OLS to simulate by fitting $y_t = \mu + \rho x_t + \epsilon_t$ .The OLS regression is $b = \left(X'X\right)^{-1} X'y$ where $X$ is a matrix with a column of 1's for the intercept and with $y_t$ as the second column.

```
T = 100
DFstats <- rep(NA, 1000)
mu<- rep(NA, 1000)
for (i in 1:1000){
        yt <- cumsum(rnorm(T))
        X <- cbind(1,yt[-T])
```

```
        # remove last element of yt and form predictor matrix
        beta <- solve(t(X) %*% X) %*% t(X) %*% yt[-1]
        # get ols estimate, note observation vector is yt with first element removed.
        ssr <- sum((yt[-1] - X %*% beta)^2)
        se <- sqrt(ssr / (T-3)) / sqrt(sum((X[,2] - mean(X[,2]))^2))
        # note (T - 3) instead of (T - 2) as we're only using (T - 1) data points
        DFstats[i] <- (beta[2] - 1) / se
        mu[i] <- beta[1]
}
quantile(DFstats,probs = c(0.01,0.05))
```

Change T to 400, we get

| *Leval* $\alpha$ | $\alpha = 0.01$ | $\alpha = 0.05$ |
|---|---|---|
| $T = 100$ | $-3.452491$ | $-2.787520$ |
| $T = 400$ | $-3.245447$ | $-2.907946$ |

■

## 3 2.13

Download the daily closing prices from January 1, 2019 to March 31,2021 of the Shanghai Composite Index. Analyze the daily data with necessary supporting tables and figures. Answer the following questions.

(a) Do the log-prices follow a random walk with a drift?

(b) Are the log-returns predictable? Use the Ljung-Box test with lags 5 and 10 at significance level 1%

(c) Fit an ARMA(p,q) ($p + q \leqslant 2$) model to the data with order chosen by AIC.

(d) Check if the residual series is white noise.

(e) Predict the time series for the first two weeks in 2014 .

**Solution:**

**(a)**

```
library(quantmod)
getSymbols("000001.ss",src="yahoo",from="2019-1-1",to="2021-3-31")
data=get("000001.SS")
close_data=data[,4]
log_close=log(data[,4])
library(tseries)
adf.test(log_close)
```

```
#        Augmented Dickey-Fuller Test
#
#data:  log_close
#Dickey-Fuller = -2.8432, Lag order = 8, p-value = 0.2214
#alternative hypothesis: stationary
```

We conclude that the log-prices follow a random walk with a drift at significance level 5%.

**(b)**

```
log_returns <- diff(log_close, lag=1)
library(stats)
Box.test(log_returns,lag=5,type = "Ljung-Box")
#        Box-Ljung test
#
#data:  log_returns
#X-squared = 4.223, df = 5, p-value = 0.5178
Box.test(log_returns,lag=10,type = "Ljung-Box")
#        Box-Ljung test
#
#data:  log_returns
#X-squared = 10.517, df = 10, p-value = 0.3963
```

We conclude that the log-returns data at each lag are i.i.d. with lags5 and 10 at significance level 10%,which means the log-returns are predictable.

**(c)**

For log-prices,we search under $p + q \leq 2$.Then, we get the best model is ARMA(1,0). We

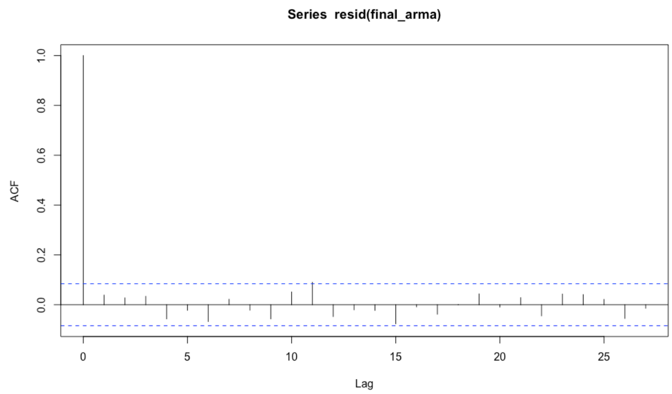
```
final_aic <- Inf
final_order <- c(0,0,0)
for (i in 0:2) for (j in 0:2) {
  current_aic <- AIC(arima(log_close, order=c(i, 0, j)))
   if (current_aic < final_aic) {
     final_aic <- current_aic
     final_order <- c(i, 0, j)
     final_arma <- arima(log_close, order=final_order)
   }
}
```

```
> final_arma
Call:
arima(x = log_close, order = final_order)
Coefficients:
            ar1   intercept
         0.9940      8.0260
s.e.     0.0051      0.0705

sigma^2 estimated as 0.0001502:  log likelihood = 1614.54,   aic = -3223.08
```

**(d)**

We plot ACF and do Ljung-Box test check if the residual series is white noise.




```
acf(resid(final_arma))
Box.test(resid(final_arma), lag=20, type="Ljung-Box")
#        Box-Ljung test
#
#data:   resid(final_arma)
#X-squared = 21.696, df = 20, p-value = 0.3573
```

We conclude that the residual series is white noise at significance level 10%.

**(e)**

We predict the time series(log prices) for the first two weeks in March of 2021.

```
 pred<-predict(final_arma,n.ahead = 10)
```

```
Time Series:
Start = 543
End = 552
Frequency = 1
[1] 8.147333 8.146607 8.145886 8.145168 8.144455 8.143746 8.143042 8.142341 8.141645 8.140953
```

■

# 4 3.1

For ARCH(1) model, if $X_t$ is strong stationary and $EX_t^4 < \infty$, show that

$$\mathrm{E}X_t^4 = \frac{a_0^2\,(1+a_1)\,\mathrm{E}\varepsilon_t^4}{(1-a_1)\,(1-a_1^2\mathrm{E}\varepsilon_t^4)}$$

and the kurtosis

$$\kappa_x = \frac{(1-a_1^2)\,\kappa_\varepsilon}{1-a_1^2\mathrm{E}\varepsilon_t^4}$$

Therefore a necessary condition for the existence of the fourth moment is that $a_1 < 1/\sqrt{E\varepsilon_t^4}$

**Solution:**
Indeed,

$$\mathrm{E}\left(X_t^4 \mid I_{t-1}\right)=\mathrm{E}\left(\sigma_t^4 \varepsilon_t^4 \mid I_{t-1}\right)=\mathrm{E}\left(\varepsilon_t^4 \mid I_{t-1}\right) E\left[\left(\sigma_t^2\right)^2 \mid I_{t-1}\right]=\left(a_0+a_1 X_{t-1}^2\right)^2 \mathrm{E} \varepsilon_t^4$$

Applying the law of iterated expectations, we have

$$\begin{aligned}
\mathrm{E}\left(X_t^4\right) & =E\left[\mathrm{E}\left(X_t^4 \mid I_{t-1}\right)\right] \\
& =\mathrm{E}\left(a_0+a_1 X_{t-1}^2\right)^2 \mathrm{E} \varepsilon_t^4 \\
& =\left[a_0^2+2 a_0 a_1 \mathrm{E}\left(X_{t-1}^2\right)+a_1^2 \mathrm{E}\left(X_{t-1}^4\right)\right] \mathrm{E} \varepsilon_t^4 \\
& =\left[a_0^2+2 a_0 a_1 \frac{a_0}{1-a_1}+a_1^2 \mathrm{E}\left(X_{t-1}^4\right)\right] \mathrm{E} \varepsilon_t^4
\end{aligned}$$

Where the last equal sign is because of $\mathrm{E} X_t^2=\frac{a_0}{1-\sum_{j=1}^p a_j}$ for ARCH$(p)$.Noticed that $X_t$ is strong stationary and $E X_t^4<\infty$, we have $\mathrm{E}\left(X_t^4\right)=\mathrm{E}\left(X_{t-1}^4\right)$.Then,reorganize the above equation we get

$$\mathrm{E} X_t^4=\frac{a_0^2\left(1+a_1\right) E \varepsilon_t^4}{\left(1-a_1\right)\left(1-a_1^2 E \varepsilon_t^4\right)}$$

Simple algebra then reveals that the kurtosis is

$$\kappa_x=\frac{\mathrm{E} X_t^4}{\left(\mathrm{E} X_t^2\right)^2}=\frac{\left(1-a_1^2\right) \mathrm{E} \varepsilon_t^4}{1-a_1^2 \mathrm{E} \varepsilon_t^4}=\frac{\left(1-a_1^2\right) \frac{\mathrm{E} \varepsilon_t^4}{\left(\mathrm{E} \varepsilon_t^2\right)^2}}{1-a_1^2 \mathrm{E} \varepsilon_t^4}=\frac{\left(1-a_1^2\right) \kappa_{\varepsilon}}{1-a_1^2 \mathrm{E} \varepsilon_t^4}$$

■