Time series models

Liming Feng

Dept. of Industrial & Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Time series

- Financial data are often time series data: data **indexed by time**
- One should take **autocorrelation** into account: value of the series at t may depend on the values before t
- Daily log return of a stock: independence that we have assumed so far often violated (e.g., large return often followed by other large returns)

1. Stationarity and autocorrelation

Ruppert 2011, §9.2; Hayashi 2000, §2.2

Stationarity

- Denote a time series by $Y = \{Y_1, \cdots, Y_t, \cdots\}$
- Y is a strictly stationary process if for any $t, s_1, \cdots, s_n$,

$$(Y_{s_1}, \cdots, Y_{s_n}) \sim (Y_{t+s_1}, \cdots, Y_{t+s_n})$$

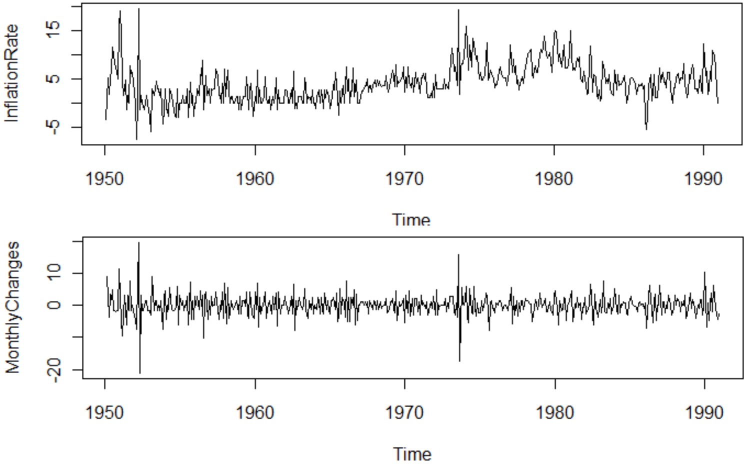
- Y is **weakly stationary** if (1) $\mathbb{E}[Y_t] = \mu$ doesn't depend on t; (2) $\mathrm{cov}(Y_t, Y_{t-h}) := \gamma(h)$ is finite and only depends on the lag h
  - $\mathrm{var}(Y_t) = \gamma(0), \forall t$
  - $\gamma(h) = \gamma(-h) = \mathrm{cov}(Y_t, Y_{t+h}), \forall t$
- Strictly stationary process with finite covariances is weakly stationary
- Stationary refers to weakly stationary in this course

Time series plot

- Stationarity may not hold if the time series systematically deviates away from a fixed level, or shows clustering
- If the original time series is not stationary, a transformed series might be stationary (difference, log, etc.)
- Annualized monthly inflation rate vs monthly changes in inflation rate

```
library(Ecdat)
data(Mishkin)
InflationRate=Mishkin[,1]
plot(InflationRate)
MonthlyChanges=diff(InflationRate)
plot(MonthlyChanges)
```

- It seems more reasonable to model the monthly change by a stationary process



White noise

- A weakly stationary process $\{\epsilon_t\}$ is a **white noise** if (1) $\mathbb{E}[\epsilon_t] = 0$, (2) $\text{var}(\epsilon_t) = \sigma^2$, and $\text{cov}(\epsilon_s, \epsilon_t) = 0, s \neq t$

$$\epsilon_t \sim WN(0, \sigma^2)$$

- It is a **Gaussian white noise** if each ϵ_t is normal
  - For Gaussian white noise, $\epsilon_t \sim N(0, \sigma^2)$ are i.i.d.
- White noise processes are the basic building block of stationary time series

Autocorrelation

- Autocovariance function of a stationary process

$$\gamma(h) = \mathrm{cov}(Y_t, Y_{t+h}) = \mathbb{E}[(Y_t - \mu)(Y_{t+h} - \mu)]$$

  Variance

$$\gamma(0) = \mathrm{cov}(Y_t, Y_t) = \mathbb{E}[(Y_t - \mu)^2]$$

- **Autocorrelation function** (ACF)

$$\rho(h) = \mathrm{corr}(Y_t, Y_{t+h}) = \gamma(h)/\gamma(0)$$

  - $\rho(h)$: lag h autocorrelation of the series
  - $\rho(0) = 1$

- Autocorrelation function is a basic tool for studying stationary time series
- **Correlogram**: the plot of $\{\rho(h)\}$ against $h = 0, 1, 2, \cdots$

Sample ACF

- Given a time series $\{Y_1, \cdots, Y_n\}$

$$\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t, \quad \hat{\gamma}(0) = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \bar{Y})^2$$

  converge to μ and $\gamma(0)$ under certain conditions

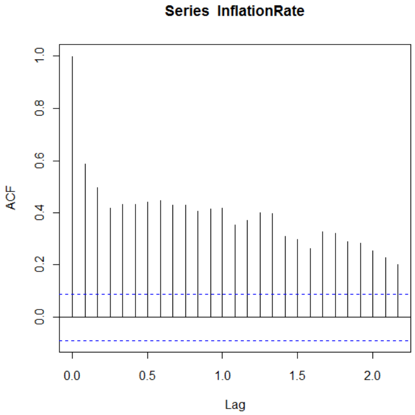
- Sample auto-covariance function

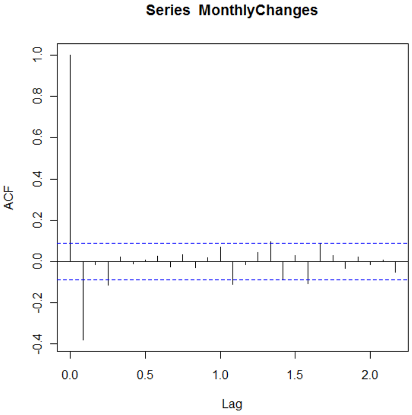
$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(Y_{t+h} - \bar{Y})(Y_t - \bar{Y})$$

- Sample ACF $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$ converges to $\rho(h)$ under similar conditions

Plotting sample ACF

- Shape of ACF helps identify features such as seasonality and appropriate time series models





2. Moving average models

Ruppert 2011, §9.7; Hayashi 2000, §6.1

Moving average processes

- Y_t is a first order **moving average** process if

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1},$$

  where $\epsilon_t \sim WN(0, \sigma^2)$; Y is a weighted average of the current and past values of a white noise process

- Denote $\theta(L) = 1 + \theta_1 L$, where L is the lag operator with $L\epsilon_t = \epsilon_{t-1}, L^2\epsilon_t = \epsilon_{t-2}$ etc.

$$Y_t - \mu = \theta(L)\epsilon_t$$

- MA(1) is **stationary for any $\theta_1 \in \mathbb{R}$**, with $\mathbb{E}[Y_t] = \mu$,

$$\gamma(0) = \text{var}(Y_t) = (1 + \theta_1^2)\sigma^2$$

$$\gamma(1) = \text{cov}(Y_t, Y_{t+1}) = \theta_1\sigma^2, \quad \gamma(h) = 0, \; h > 1$$

- ACF for MA(1)

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta_1}{1 + \theta_1^2}, \quad \rho(h) = 0, \; h > 1$$

- Correlogram determined by one parameter: θ_1

- **MA(q)** process

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

- Denote $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$

$$Y_t - \mu = \theta(L)\epsilon_t$$

- MA(q) is **stationary for any** $\theta_1, \cdots, \theta_q$. For MA(q), denote $\theta_0 = 1$

$$\gamma(h) = (\theta_h \theta_0 + \theta_{h+1}\theta_1 + \cdots + \theta_q \theta_{q-h})\sigma^2, h = 0, 1, \cdots, q$$

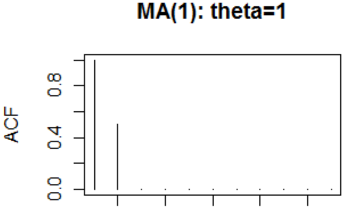
$$\gamma(h) = 0, \rho(h) = 0, \forall h > q$$

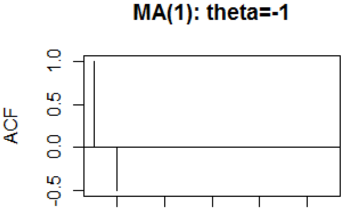
- Correlogram determined by q parameters: $\theta_1, \cdots, \theta_q$

- One may use the ACF plot to identify a MA(q) process
- For MA(1),

$$\rho(0) = 1, \rho(1) = \frac{\theta_1}{1+\theta_1^2}, \rho(2) = \rho(3) = \cdots = 0$$

```
> plot(ARMAacf(ma=c(1), lag.max = 10),type="h",xlab="Lag",ylab="ACF",main="MA(1): theta=1")
```

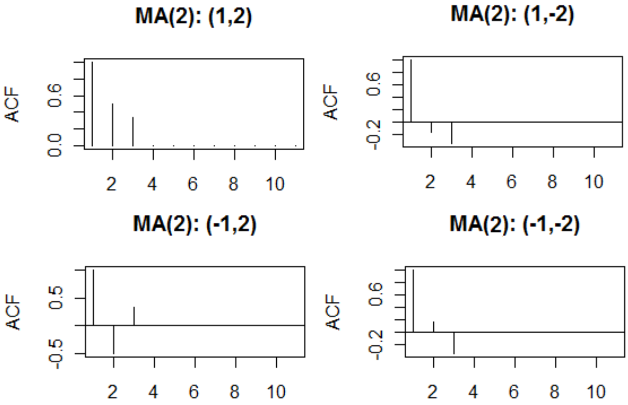




- For MA(2),

$$\rho(0) = 1, \rho(1) = \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2}, \rho(2) = \frac{\theta_2}{1+\theta_1^2+\theta_2^2}, \rho(3) = \cdots = 0$$

```
> plot(ARMAacf(ma=c(1,2), lag.max = 10),type="h",xlab="Lag",ylab="ACF",main="MA(1): (1,2)")
>
```



- Recall the ACF plot for monthly changes in inflation, MA(1) or MA(3) may fit well

```
> arima(MonthlyChanges,order=c(0,0,1))
Coefficients:
          ma1  intercept
      -0.7932     0.0010
s.e.   0.0355     0.0281

sigma^2 estimated as 8.882:  log likelihood = -1230.85,  aic = 2467.7
> arima(MonthlyChanges,order=c(0,0,2))
Coefficients:
          ma1      ma2  intercept
      -0.6621  -0.1660     0.0001
s.e.   0.0419   0.0407     0.0230

sigma^2 estimated as 8.595:  log likelihood = -1222.85,  aic = 2453.7
> arima(MonthlyChanges,order=c(0,0,3))
Coefficients:
         ma1      ma2      ma3  intercept
      -0.633  -0.1027  -0.1082    -0.0002
s.e.   0.046   0.0514   0.0470     0.0209

sigma^2 estimated as 8.505:  log likelihood = -1220.26,  aic = 2450.52
> arima(MonthlyChanges,order=c(0,0,4))
Coefficients:
          ma1      ma2      ma3     ma4  intercept
      -0.6434  -0.1114  -0.1399  0.0625     0.0000
s.e.   0.0462   0.0548   0.0522  0.0487     0.0224

sigma^2 estimated as 8.476:  log likelihood = -1219.45,  aic = 2450.89
```

MA(∞)

- MA(q): serial correlation dies out after q lags
- When data show that serial correlation doesn't die out, MA(∞) might be the case

$$Y_t = \mu + \psi_0 \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \cdots$$

where we assume that $\sum_{j=0}^{\infty} |\psi_j| < \infty$

- MA(∞) is stationary with mean $\mathbb{E}[Y_t] = \mu$ and **absolutely summable** autocovariances

$$\gamma(h) = (\psi_h \psi_0 + \psi_{h+1} \psi_1 + \cdots)\sigma^2, \quad \sum_{h=0}^{\infty} |\gamma(h)| < \infty$$

3. Filtering

Hayashi 2000, §6.1

- If $\{X_t\}$ is stationary and $\sum_{j=0}^{\infty} |h_j| < \infty$, then

$$Y_t = h_0 X_t + h_1 X_{t-1} + h_2 X_{t-2} + \cdots$$

  is stationary; if $\{X_t\}$ has absolutely summable autocovariances, so does $\{Y_t\}$
- **Filtering**: taking a weighted average of successive values of a process
- **Filter**:

$$h(L) = h_0 + h_1 L + h_2 L^2 + \cdots$$

  is an absolutely summable filter

Product of filters

- Product of $\alpha(L) = 1 + \alpha_1 L$ and $\beta(L) = 1 + \beta_1 L$

$$\begin{aligned}\alpha(L)\beta(L) &= (1 + \alpha_1 L)(1 + \beta_1 L) \\ &= 1 + (\alpha_1 + \beta_1)L + \alpha_1\beta_1 L^2\end{aligned}$$

- Products for arbitrary filters defined similarly
- If $\{X_t\}$ is stationary, $\alpha(L)$ and $\beta(L)$ are absolutely summable filters, then $\alpha(L)\beta(L)$ is absolutely summable,

$$\alpha(L)\beta(L)X_t$$

  is well defined and stationary

- If $\alpha(L)\beta(L) = 1$, $\beta(L)$ is called the **inverse** of $\alpha(L)$:

$$\beta(L) = \alpha(L)^{-1}$$

- Inverse can be easily found: the inverse of $\alpha(L) = 1 - \phi_1 L$

$$(1 - \phi_1 L)(\beta_0 + \beta_1 L + \beta_2 L^2 + \cdots) = 1$$

$$\beta_0 = 1,$$

$$(\beta_1 - \beta_0\phi_1)L = 0 \Rightarrow \beta_1 = \beta_0\phi_1 = \phi_1,$$

$$(\beta_2 - \beta_1\phi_1)L^2 = 0 \Rightarrow \beta_2 = \beta_1\phi_1 = \phi_1^2, \cdots$$

- Therefore

$$(1 - \phi_1 L)^{-1} = 1 + \phi_1 L + \phi_1^2 L^2 + \cdots$$

- $(1-\phi_1 L)^{-1}$ is absolutely summable if $|\phi_1| < 1$. That is, when the root of the following polynomial is great than 1 in absolute value

$$1 - \phi_1 z = 0$$

- Inverse of $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$: find $\psi(L)$ such that $\phi(L)\psi(L) = 1$

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)(\psi_0 + \psi_1 L + \psi_2 L^2 + \cdots) = 1$$

- $\psi(L)$ is absolutely summable if the roots of the following polynomial are great than 1 in absolute value (**stability condition**)

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

4. Autoregressive models

Ruppert 2011, §9.4, 9.5, 9.6; Hayashi 2000, §6.2

Autoregressive processes

- An AR(1) process is given by

$$Y_t = a + \phi_1 Y_{t-1} + \epsilon_t,$$

  where $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ for some $\sigma_\epsilon > 0$
- Y_t consists of (1) feedback from the past Y_{t-1}; (2) new information ϵ_t
- Would like to find Y_t in terms of $\{\epsilon_t\}$ that is stationary and satisfies the above equation: not always possible for an arbitrary ϕ_1

$\phi_1 = 1$

- When $\phi_1 = 1$,

$$
\begin{aligned}
Y_t &= a + Y_{t-1} + \epsilon_t \\
&= a + (a + Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \;=\; 2a + Y_{t-2} + \epsilon_t + \epsilon_{t-1} \\
&= \cdots \\
&= ha + Y_{t-h} + \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_{t-h+1}, \forall h \geq 1
\end{aligned}
$$

- If $\{Y_t\}$ were stationary,

$$\mathrm{var}(Y_t - Y_{t-h}) = 2\gamma(0) - 2\gamma(h) = h\sigma_\epsilon^2, \;\; \rho(h) = 1 - \frac{h\sigma_\epsilon^2}{2\gamma(0)}$$

  $\rho(h)$ will become less than -1 for large h, contradiction!

- Similar contradiction when $\phi_1 = -1$

Lévy process models

- Asset price $S_t = S_0 e^{X_t}$, where $\{X_t\}$ is a Lévy process with independent and stationary increments
- For asset prices observed at times $\{0, \Delta, 2\Delta, \cdots\}$

$$
\begin{aligned}
S_\Delta &= S_0 e^{X_\Delta - X_0}, \ X_0 = 0 \\
S_{2\Delta} &= S_0 e^{X_{2\Delta}} = S_\Delta e^{X_{2\Delta} - X_\Delta} \quad \cdots \\
S_{k\Delta} &= S_{(k-1)\Delta} e^{X_{k\Delta} - X_{(k-1)\Delta}}
\end{aligned}
$$

- $X_{k\Delta} - X_{(k-1)\Delta}$ are i.i.d; denote its mean by a and variance by σ_ϵ^2; then $\epsilon_k := X_{k\Delta} - X_{(k-1)\Delta} - a \sim WN(0, \sigma_\epsilon^2)$

$$
\ln(S_{k\Delta}) = a + \ln(S_{(k-1)\Delta}) + \epsilon_k, \ \ln(S_{k\Delta}/S_{(k-1)\Delta}) = a + \epsilon_k
$$

- Log price process is non-stationary AR(1); the log return process is stationary

$|\phi_1| > 1$

- For $\phi_1 \neq 1$, define μ such that $\mu = a + \phi_1 \mu$; $Y_t = a + \phi_1 Y_{t-1} + \epsilon_t$ becomes

$$Y_t - \mu = \phi_1 (Y_{t-1} - \mu) + \epsilon_t$$

- When $|\phi_1| > 1$, can verify that the following is the stationary solution of the AR(1) model

$$Y_t = \mu - \sum_{j=1}^{\infty} \phi_1^{-j} \epsilon_{t+j}$$

- This is not a useful model: current value depends on all future information

$|\phi_1| < 1$

- When $|\phi_1| < 1$,

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \epsilon_t \Leftrightarrow (1 - \phi_1 L)(Y_t - \mu) = \epsilon_t$$

- $(1 - \phi_1 L)^{-1}$ is absolutely summable since $|\phi_1| < 1$
- If $\{Y_t\}$ is the stationary solution of the above, it should satisfy

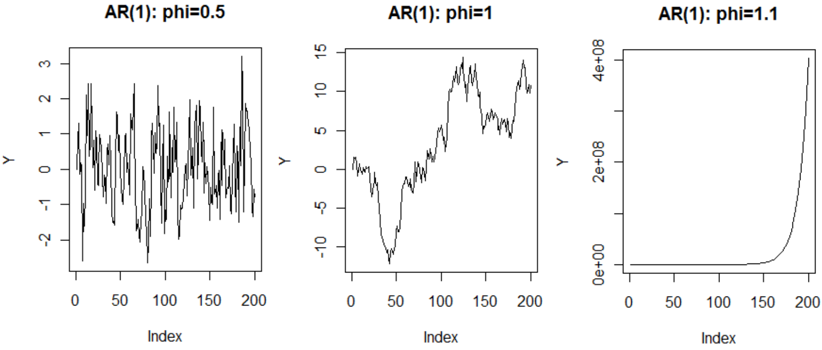
$$(1 - \phi_1 L)^{-1}(1 - \phi_1 L)(Y_t - \mu) = Y_t - \mu = (1 - \phi_1 L)^{-1}\epsilon_t$$

$$Y_t = \mu + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}$$

- AR(1) process is stationary with mean μ and admits a MA(∞) representation when the following **stationarity condition** is satisfied: root of $1 - \phi_1 z = 0$ is great than 1 in absolute value

Simulated AR(1)

- Consider an AR(1) process with $a = 0$, $Y_0 = 0$ and Gaussian white noise ϵ_t
$$Y_t = \phi_1 Y_{t-1} + \epsilon_t$$
- $|\phi_1| < 1$: stationary
- $\phi_1 = \pm 1$: non-stationary
- $\phi_1 = 1$: $Y_t = \sum_{j=1}^{t} \epsilon_j$ is a random walk, $\text{var}(Y_t) = t\sigma_\epsilon^2$ increases over time
- $|\phi_1| > 1$ and assuming ϵ_t represents new information and is uncorrelated with Y_{t-1}: non-stationary and explosive
- Simulate AR(1) for $\phi_1 = 0.5, 1, 1.1$ and plot the sample paths


AR(1): phi=0.5
AR(1): phi=1
AR(1): phi=1.1
Y
Index
0
50
100
150
200
3
2
1
0
-1
-2
15
10
5
-5
-10
4e+08
2e+08
0e+00

ACF of AR(1)

- Consider a stationary AR(1) process with $|\phi_1| < 1$

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \epsilon_t$$

- Take variance on both sides (Y_{t-1} uncorrelated with ϵ_t)

$$\gamma(0) = \phi_1^2\gamma(0) + \sigma_\epsilon^2 \Rightarrow \gamma(0) = \sigma_\epsilon^2/(1 - \phi_1^2)$$

- To compute the autocovariance $\text{cov}(Y_t, Y_{t+h})$, we note that

$$\begin{aligned} Y_{t+h} - \mu &= \phi_1(Y_{t+h-1} - \mu) + \epsilon_{t+h} = \cdots \\ &= \phi_1^h(Y_t - \mu) + \phi_1^{h-1}\epsilon_{t+1} + \cdots + \epsilon_{t+h} \end{aligned}$$

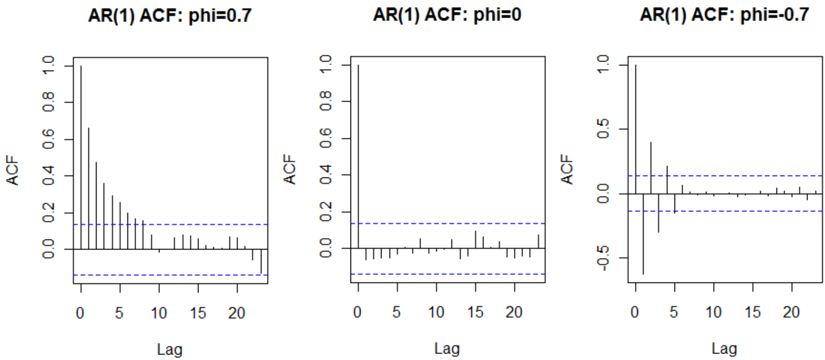
- $\gamma(h) = \phi_1^h\gamma(0)$, and ACF $\rho(h) = \gamma(h)/\gamma(0) = \phi_1^h$, $h \geq 0$

- Simulate AR(1) and plot ACF for $\phi_1 = -0.7, 0, 0.7$

```
n=200
Y=rep(0,n)
epsilon=rnorm(n) # Gaussian white noise
phi=0.7
for (i in 2:n)Y[i]=phi*Y[i-1]+epsilon[i]
acf(Y)
```

- $\phi_1 < 0$: ACF alternates signs as h increases
- $\phi_1 = 0$: one gets a white noise; ACF close to zero for $h > 0$

Typical shapes of AR(1) ACF



AR(p)

- AR(p) incorporates more past values into the model

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t$$

- Denote $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$

$$\phi(L)(Y_t - \mu) = \epsilon_t$$

- Under **stationarity condition** (all roots of $\phi(z) = 0$ are greater than 1 in absolute value): AR(p) is stationary; admits MA(∞) representation with absolutely summable coefficients

$$Y_t = \mu + \psi(L)\epsilon_t, \quad \phi(L)\psi(L) = 1$$

has mean μ and absolutely summable autocovariances

ACF of AR(2)

- Consider a stationary AR(2)

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \epsilon_t$$

- **Yule-Walker equations**: multiply $Y_{t-h} - \mu$ on both sides and take expectation

$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2), h \geq 1$$

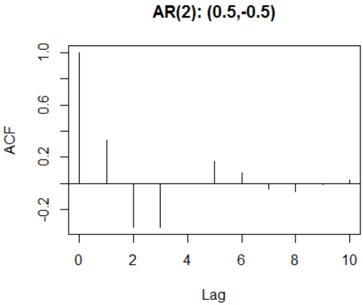
  Dividing $\gamma(0)$ on both sides

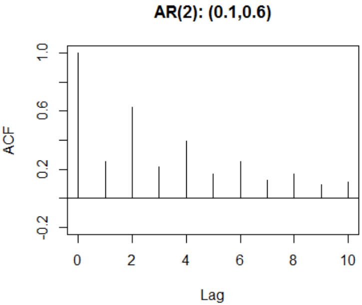
$$\rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2), h \geq 1$$

- Using $\rho(0) = 1$ and $\rho(h) = \rho(-h)$, one can solve all $\rho(h)$

- AR(p) allows more flexibility in acf

```
x=ARMAacf(ar=c(0.5, -0.5), lag.max = 10)
plot(0:10,x,type="h",xlab="Lag",ylab="ACF",main="AR(2): (0.5,-0.5)")
abline(h=0) #add a horizontal line at y=0
```





- ACF decreases to zero geometrically

AR(p) fit for monthly changes in inflation

- Are autoregressive models better for monthly changes in inflation?

```
> arima(MonthlyChanges,order=c(1,0,0))

Call:
arima(x = MonthlyChanges, order = c(1, 0, 0))

Coefficients:
          ar1  intercept
      -0.3849     0.0038
s.e.   0.0420     0.1051

sigma^2 estimated as 10.37:  log likelihood = -1268.34,  aic = 2542.68
```

- The following table shows that there is no parsimonious AR(p) fit for monthly changes in inflation

Model	AIC	Model	AIC
$p = 1$	2542.68	$p = 6$	2467.32
$p = 2$	2526.87	$p = 7$	2465.6
$p = 3$	2499.68	$p = 8$	2462.4
$p = 4$	2486.1	$p = 9$	2463.32
$p = 5$	2475.73	$p = 10$	2462.21

- MA(3) remains the best among all AR(p) and MA(q) models, with AIC 2450.52

5. ARMA models

Ruppert 2011, §9.8; Hayashi 2000, §6.2

ARMA models

- ARMA models combine AR and MA and can be more parsimonious than pure AR or MA models
- An ARMA(p, q) model is of the following form

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

- Using the lag operator

$$(1 - \phi_1 L - \cdots - \phi_p L^p)(Y_t - \mu) = (1 + \theta_1 L + \cdots + \theta_q L^q)\epsilon_t$$

$$\text{or simply} \quad \phi(L)(Y_t - \mu) = \theta(L)\epsilon_t$$

Stationarity

- When roots of $\phi(z) = 0$ are great than 1 in absolutely value, the ARMA(p,q) process is stationary
- It has mean μ
- It admits MA(∞) representation $Y_t = \mu + \psi(L)\epsilon_t$, $\psi(L) = \phi(L)^{-1}\theta(L)$ is absolutely summable
- Autocovariances decreases to zero geometrically and are absolutely summable

Special cases

- With $q = 0$, one obtains an AR(p) model
- With $p = 0$, one obtains a MA(q) model
- With $p = q = 0$, one obtains a white noise process
- Stationary ARMA(1,1) with $p = q = 1$

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1}$$

- Yule-Walker equation for autocovariances: take covariance with $Y_{t-h} - \mu$ on both sides of

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \epsilon_t + \theta_1 \epsilon_{t-1}$$

$$\gamma(h) = \phi_1 \gamma(h-1), \;\; h \geq 2$$

$$\rho(h) = \phi_1 \rho(h-1), \;\; h \geq 2$$

- It suffices to compute $\gamma(0), \gamma(1)$

- Note that

$$\text{cov}(Y_t - \mu, \epsilon_t) = \text{cov}(\phi_1(Y_{t-1} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1}, \epsilon_t) = \sigma_\epsilon^2$$

- Take variance on both sides of

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1}$$

$$\gamma(0) = \phi_1^2\gamma(0) + (1 + \theta_1^2 + 2\phi_1\theta_1)\sigma_\epsilon^2$$

$$\gamma(0) = \frac{1 + \theta_1^2 + 2\phi_1\theta_1}{1 - \phi_1^2}\sigma_\epsilon^2$$

- Take covariance with $Y_{t-1} - \mu$ on both sides of

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \epsilon_t + \theta_1 \epsilon_{t-1}$$

$$\gamma(1) = \phi_1 \gamma(0) + \theta_1 \sigma_\epsilon^2$$

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{(\phi_1 + \theta_1)(1 + \phi_1 \theta_1)}{1 + \theta_1^2 + 2\phi_1 \theta_1}$$

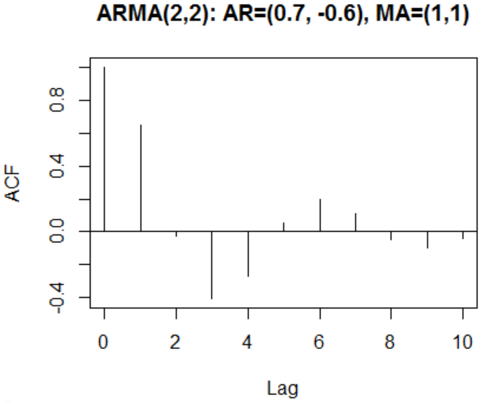
- For $h \geq 2$

$$\rho(h) = \phi_1^{h-1} \rho(1), \;\; h \geq 2$$

- ACF for ARMA(p,q) computed similarly

- Plot the ACF of an AMRA(2,2)

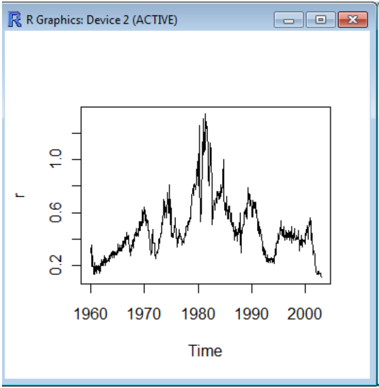
```
> x=ARMAacf(ar=c(0.7, -0.6), ma=c(1, 1), lag.max = 10)
> plot(0:10,x,type="h",xlab="Lag",ylab="ACF",main="ARMA(2,2): AR=(0.7, -0.6), MA=(1,1)")
> abline(h=0) #add a horizontal line at y=0
```

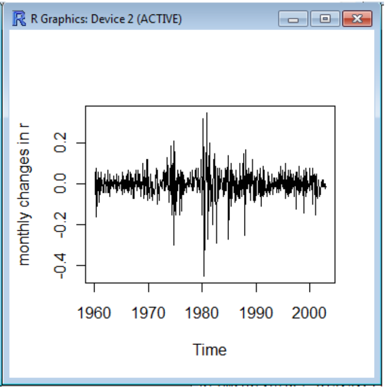


Example: monthly risk free rates

- Monthly risk free interest rates from Jan 1960 to Dec 2002

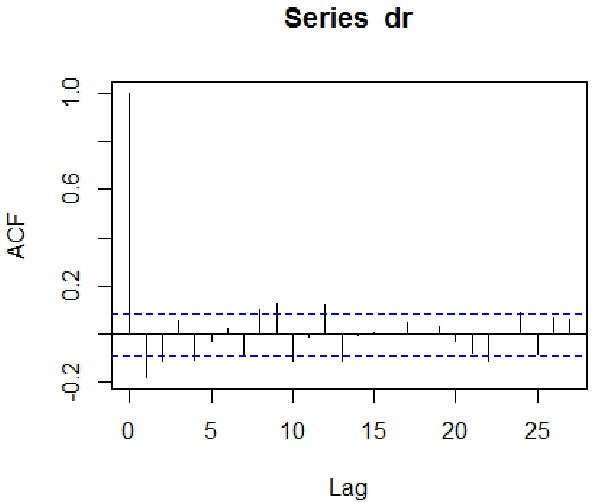
```
> library(Ecdat)
> data(Capm)
> dr=diff(Capm$rf)
> plot(ts(Capm$rf,start=c(1960,1),frequency=12),ylab="r")
> plot(ts(dr,start=c(1960,2),frequency=12),ylab="monthly changes in r")
>
```





Monthly change in risk free rate

- Sample autocorrelation: acf(dr)



ARMA fit

- Monthly change in interest rate seems stationary; fit ARMA(p,q) to monthly changes (using no more than 3 parameters)

Model	AIC
AR(1)	-1275.14
AR(2)	**-1284.39**
AR(3)	-1282.42
MA(1)	-1280.79
MA(2)	**-1287**
MA(3)	-1285
ARMA(1,1)	**-1287.33**
ARMA(2,1)	-1283.02
ARMA(1,2)	-1285.33

- Using no more than 3 parameters, ARMA(1,1) provides the best fit

ARMA(1,1) fit

- ARMA(1,1) fit for monthly change in interest rate

```
> fit=arima(dr,order=c(1,0,1))
> fit
Series: dr
ARIMA(1,0,1) with non-zero mean

Coefficients:
         ar1      ma1  intercept
      0.4546  -0.6758    -0.0004
s.e.  0.1105   0.0903     0.0018

sigma^2 estimated as 0.004733:  log likelihood=647.67
AIC=-1287.33   AICc=-1287.25   BIC=-1270.36
```

- The ARMA(1,1) process for the monthly change in interest rate

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1}$$

with estimates $\hat{\phi}_1 = 0.4546$, $\hat{\theta}_1 = -0.6758$, $\hat{\mu} = -0.0004$, ϵ_t is a Gaussian white noise with $\hat{\sigma}^2_\epsilon = 0.004733$

6. Parameter estimation

Ruppert 2011, §9.5.2; Hayashi 2000, §6.4

Maximum likelihood estimation for AR(1)

- Consider AR(1) with Gaussian noise $\epsilon_t \sim N(0, \sigma_\epsilon^2)$

$$Y_t = a + \phi_1 Y_{t-1} + \epsilon_t$$

- Find maximum likelihood estimates

$$\hat{a}, \hat{\phi}_1, \hat{\sigma}_\epsilon; \quad \hat{\mu} = \frac{\hat{a}}{1 - \hat{\phi}_1}$$

- Key identity in deriving the likelihood function: joint pdf of (X, Y, Z) satisfies

$$f_{XYZ}(x, y, z) = f_X(x) f_{Y|X}(y|x) f_{Z|X,Y}(z|x, y)$$

Conditional MLE

- Given data $\{y_1, \cdots, y_n\}$. Starting from the initial observation y_1, what's the likelihood of observing $Y_2 = y_2, \cdots, Y_n = y_n$
- Given the initial observation y_1, distribution of Y_2 is $N(a + \phi_1 y_1, \sigma_\epsilon^2)$ with density

$$f_{Y_2}(y_2) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e^{-\frac{(y_2 - a - \phi_1 y_1)^2}{2\sigma_\epsilon^2}}$$

- Given $Y_2 = y_2$, distribution of Y_3 is $N(a + \phi_1 y_2, \sigma_\epsilon^2)$ with density

$$f_{Y_3|Y_2}(y_3|y_2) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e^{-\frac{(y_3 - a - \phi_1 y_2)^2}{2\sigma_\epsilon^2}}$$

- The likelihood function

$$L(a, \phi_1, \sigma_\epsilon) = f_{Y_2}(y_2) f_{Y_3|Y_2}(y_3|y_2) \cdots f_{Y_n|Y_{n-1}}(y_n|y_{n-1})$$

- Log likelihood function

$$\log(L(a, \phi_1, \sigma_\epsilon)) = -\frac{n-1}{2} \log(2\pi\sigma_\epsilon^2) - \sum_{i=2}^{n} \frac{(y_i - a - \phi_1 y_{i-1})^2}{2\sigma_\epsilon^2}$$

- **Conditional maximum likelihood estimates** for a, ϕ_1 can be found by minimizing

$$\sum_{i=2}^{n} (y_i - a - \phi_1 y_{i-1})^2$$

$\hat{a}$ and $\hat{\phi}_1$ obtained through **ordinary least squares** (OLS) regression

- $\hat{\sigma}_\epsilon$ obtained by solving $\partial L(\hat{a}, \hat{\phi}_1, \sigma_\epsilon)/\partial \sigma_\epsilon = 0$

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-1} \sum_{i=2}^{n} (y_i - \hat{a} - \hat{\phi}_1 y_{i-1})^2$$

  That is, $\hat{\sigma}_\epsilon^2$ is the **mean squared error** in OLS

- For AR(p), given p initial observations $\{y_1, \cdots, y_p\}$, one can derive conditional maximum likelihood estimates similarly from an OLS regression of y on p of its lagged values

Maximum likelihood estimation for MA(1)

- Consider MA(1) with Gaussian noise $\epsilon_t \sim N(0, \sigma_\epsilon^2)$

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}$$

- Note that $(Y_1, \cdots, Y_n)^\top$ has a multivariate normal distribution with mean vector $(\mu, \cdots, \mu)^\top$ and covariance matrix

$$\Sigma = \begin{pmatrix} \gamma(0) & \gamma(1) & & \\ \gamma(1) & \gamma(0) & \gamma(1) & \\ & & \ddots & \\ & & \gamma(1) & \gamma(0) \end{pmatrix}$$

where $\gamma(0) = (1 + \theta_1^2)\sigma_\epsilon^2$, $\gamma(1) = \theta_1 \sigma_\epsilon^2$

- Denote $\boldsymbol{\mu} = (\mu, \cdots, \mu)^\top$, $\mathbf{y} = (y_1, \cdots, y_n)^\top$
- Joint density of $(Y_1, \cdots, Y_n)^\top$

$$\frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$

- Log-likelihood function

$$\log(L(\mu, \theta_1, \sigma_\epsilon)) = -\frac{1}{2}\log((2\pi)^n|\Sigma|) - \frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})$$

- The **exact maximum likelihood estimates** $\hat{\mu}, \hat{\theta}_1, \hat{\sigma}_\epsilon$ can be obtained numerically by maximizing the above log-likelihood
- The exact MLE method can be used for MA(q), AR(p), ARMA(p,q) in general

7. ARIMA models

Ruppert 2011, §9.9, 9.11

ARIMA models

- Recall that the inflation/interest rate processes are not so stationary; but their monthly changes seem stationary
- Given a non-stationary series $\{Y_t, t \geq 1\}$, one may obtain a stationary series $\{\Delta Y_t, t \geq 2\}$ by differencing

$$\Delta Y_t = Y_t - Y_{t-1} = (1 - L)Y_t$$

- E.g., $\{Y_t, t \geq 1\}$ is a random walk: $Y_t = Y_{t-1} + \epsilon_t$

$$\Delta Y_t = \epsilon_t$$

$\{\Delta Y_t, t \geq 2\}$ is a white noise and hence stationary

Differencing

- If first order differencing not enough, one may use higher order differencing

$$
\begin{aligned}
\Delta^2 Y_t &= \Delta(\Delta Y_t) \\
&= Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2}) \\
&= Y_t - 2Y_{t-1} + Y_{t-2} \\
&= (1 - 2L + L^2) Y_t \\
&= (1 - L)^2 Y_t, \;\; t \geq 2
\end{aligned}
$$

- In general, $\Delta^d Y_t = (1 - L)^d Y_t$

ARIMA(p,d,q)

- Series Y_t is an ARIMA(p,d,q) process if $\Delta^d Y_t$ is ARMA(p,q)
- MA(3) fits monthly changes in inflation well: inflation itself is ARIMA(0,1,3)
- ARMA(1,1) fits monthly changes in interest rate well: interest rate itself is ARIMA(1,1,1)
- Random walk: ARIMA(0,1,0)

Differencing vs integrating

- Suppose Y_t is ARIMA(p,1,q) so that $X_t = \Delta Y_t$ is ARMA(p,q)

$$X_t = Y_t - Y_{t-1}$$

$$Y_t = X_t + Y_{t-1} = X_t + X_{t-1} + Y_{t-2} = X_t + \cdots + X_2 + Y_1$$

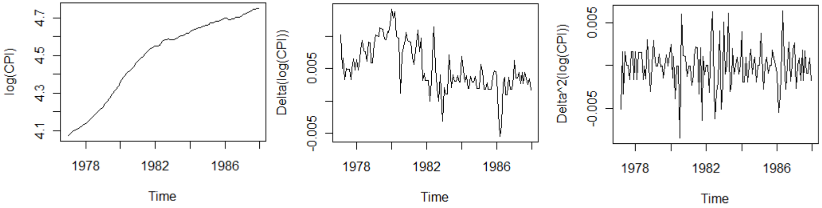
- X_t is obtained by differencing Y_t; Y_t is obtained by integrating X_t; I in ARIMA refers to integrating
- E.g., Log returns are obtained by differencing log prices; log prices are obtained by integrating log returns

Fitting U.S. consumer price index

- Obtain monthly U.S. CPI from Jan 1977 to Dec 1987

```
> library(fEcofin)
> y=log(CPI.dat$CPI[769:900])
> dy=diff(y)
> ddy=diff(dy)
> plot(ts(y,start=c(1977,1),frequency=12),ylab="log(CPI)")
> plot(ts(dy,start=c(1977,2),frequency=12),ylab="Delta(log(CPI))")
> plot(ts(ddy,start=c(1977,3),frequency=12),ylab="Delta^2(log(CPI))")
>
```

- Time series plots show that log CPI becomes stationary after second order differencing; i.e., log CPI is integrated to order 2
- Log CPI is likely ARIMA(p,2,q)



Determining p, d, q in R

- R can automatically determine p, d, q; it confirms that $d = 2$, and chooses ARMA(2,2,2)

```
> auto.arima(y)
ARIMA(2,2,2)
> fit=arima(ddy,order=c(2,0,2))
> fit
Coefficients:
          ar1     ar2     ma1      ma2  intercept
      -0.3847  0.3911  0.1265  -0.8735      0e+00
s.e.   0.1157  0.1143  0.0678   0.0667      1e-04

sigma^2 estimated as 4.683e-06:  log likelihood=611.66
AIC=-1211.32   AICc=-1210.64   BIC=-1194.12
```

8. Forecasting

Ruppert 2011, §9.12

Forecasting

- Predicting the future based on the values of the time series up to the present
- Consider an ARMA(1,1) process

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \epsilon_t + \theta_1 \epsilon_{t-1},$$

- Based on information known up to time t (e.g., Y_t, ϵ_t, etc.), predict $Y_{t+1}, Y_{t+2}, \cdots$

One-step-ahead forecast

- According to the ARMA(1,1) model

$$Y_{t+1} = \mu + \phi_1(Y_t - \mu) + \epsilon_{t+1} + \theta_1\epsilon_t$$

- Best prediction for ϵ_{t+1} is its mean: 0
- One-step-ahead forecast is then

$$\hat{Y}_{t+1} = \mu + \phi_1(Y_t - \mu) + \theta_1\epsilon_t$$

- One-step-ahead forecasting error

$$Y_{t+1} - \hat{Y}_{t+1} = \epsilon_{t+1}$$

Prediction interval

- Suppose that the white noise process is Gaussian: $\epsilon_{t+1} \sim N(0, \sigma_\epsilon^2)$

  $$\mathbb{P}(|Y_{t+1} - \hat{Y}_{t+1}| \leq z_{\alpha/2}\sigma_\epsilon) = 1 - \alpha$$

  For $\alpha = 5\%$, $z_{\alpha/2} \approx 1.96$
- With probability 95%, the true value $Y_{t+1} \in \hat{Y}_{t+1} \pm 1.96\sigma_\epsilon$
- $[\hat{Y}_{t+1} - z_{\alpha/2}\sigma_\epsilon, \hat{Y}_{t+1} + z_{\alpha/2}\sigma_\epsilon]$ is the prediction interval for Y_{t+1}

Two-step-ahead forecast

- Since

$$\begin{aligned} Y_{t+2} &= \mu + \phi_1(Y_{t+1} - \mu) + \epsilon_{t+2} + \theta_1 \epsilon_{t+1} \\ &= \mu + \phi_1^2(Y_t - \mu) + \phi_1\theta_1\epsilon_t + \epsilon_{t+2} + (\theta_1 + \phi_1)\epsilon_{t+1} \end{aligned}$$

- Two-step-ahead forecast

$$\begin{aligned} \hat{Y}_{t+2} &= \mu + \phi_1(\hat{Y}_{t+1} - \mu) \\ &= \mu + \phi_1^2(Y_t - \mu) + \phi_1\theta_1\epsilon_t \end{aligned}$$

with forecasting error

$$Y_{t+2} - \hat{Y}_{t+2} = \epsilon_{t+2} + (\theta_1 + \phi_1)\epsilon_{t+1} \sim N(0, (1 + (\theta_1 + \phi_1)^2)\sigma_\epsilon^2)$$

Multi-step-ahead forecast

- In general, k-step-ahead forecast

$$\hat{Y}_{t+k} = \mu + \phi_1^k (Y_t - \mu) + \phi_1^{k-1} \theta_1 \epsilon_t$$

Forecasting error when $k \geq 2$

$$N\Big(0, \big(1 + (\theta_1 + \phi_1)^2 (1 + \phi_1^2 + \phi_1^4 + \cdots + \phi_1^{2k-4})\big)\sigma_\epsilon^2\Big)$$

- Note that $|\phi_1| < 1$ for the ARMA(1,1) process to be stationary
- (1) Multi-step-ahead forecast converges to μ; (2) variance of forecasting error gradually increase; (3) it converges to

$$\frac{1 + \theta_1^2 + 2\phi_1\theta_1}{1 - \phi_1^2}\sigma_\epsilon^2 = \gamma(0)$$

Forecasting a non-stationary process

- Consider a random walk

$$Y_t = Y_{t-1} + \epsilon_t$$

- Given information up to time t, k-step-ahead forecast

$$\hat{Y}_{t+k} = Y_t$$

  Forecasting error

$$\epsilon_{t+1} + \cdots + \epsilon_{t+k} \sim N(0, k\sigma_\epsilon^2)$$

  Variance converges to infinity

Replacing the unknown parameters

- Fit the model to find estimates $\hat{\phi}_1$, $\hat{\theta}_1$, $\hat{\mu}$, $\hat{\sigma}_\epsilon^2$
- Obtain the residuals $\{\hat{\epsilon}_1, \cdots, \hat{\epsilon}_t\}$

$$Y_t - \hat{\mu} = \hat{\phi}_1(Y_{t-1} - \hat{\mu}) + \epsilon_t + \hat{\theta}_1\epsilon_{t-1}$$

$$\Downarrow$$

$$\hat{\epsilon}_t = Y_t - \hat{\mu} - \hat{\phi}_1(Y_{t-1} - \hat{\mu}) - \hat{\theta}_1\hat{\epsilon}_{t-1}$$

- k-step-ahead forecast

$$\hat{Y}_{t+k} = \hat{\mu} + \hat{\phi}_1^k(Y_t - \hat{\mu}) + \hat{\phi}_1^{k-1}\hat{\theta}_1\hat{\epsilon}_t$$

Forecasting monthly change in interest rate

- Recall that ARMA(1,1) fits monthly changes in interest rate well

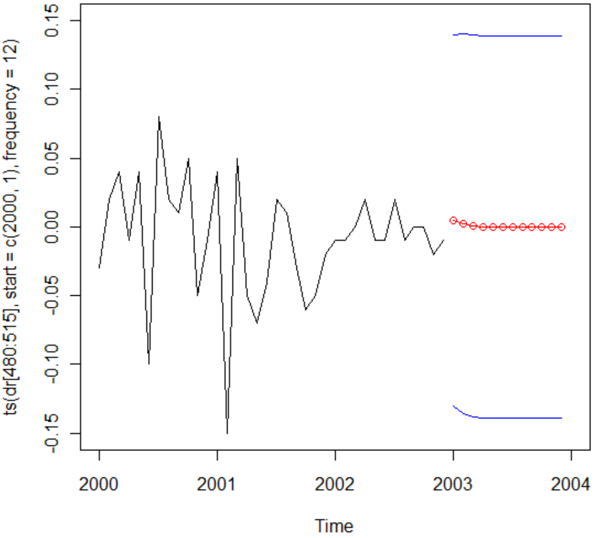
```
> library(Ecdat)
> data(Capm)
> dr=diff(Capm$rf)
> fit=arima(dr,order=c(1,0,1))
> forecasts=predict(fit,12)
> forecasts
> plot(ts(dr[480:515],start=c(2000,1),frequency=12),xlim=c(2000,2004),ylim=c(-0.15,0.15))
> lines(seq(from=2003,by=1/12,length=12),forecasts$pred,col="red",type="o")
> lines(seq(from=2003,by=1/12,length=12),forecasts$pred+1.96*forecasts$se,col="blue")
> lines(seq(from=2003,by=1/12,length=12),forecasts$pred-1.96*forecasts$se,col="blue")
```

- Predicted future values and standard deviations

```
> forecasts
$pred
Time Series:
Start = 516
End = 527
Frequency = 1
 [1]  0.0048615455  0.0020142320  0.0007197037  0.0001311477
 [5] -0.0001364387 -0.0002580967 -0.0003134083 -0.0003385557
 [9] -0.0003499889 -0.0003551871 -0.0003575504 -0.0003586249

$se
Time Series:
Start = 516
End = 527
Frequency = 1
 [1] 0.06879341 0.07045505 0.07079365 0.07086344 0.07087786
 [6] 0.07088084 0.07088145 0.07088158 0.07088161 0.07088161
[11] 0.07088161 0.07088161
```

Predicted values and prediction interval

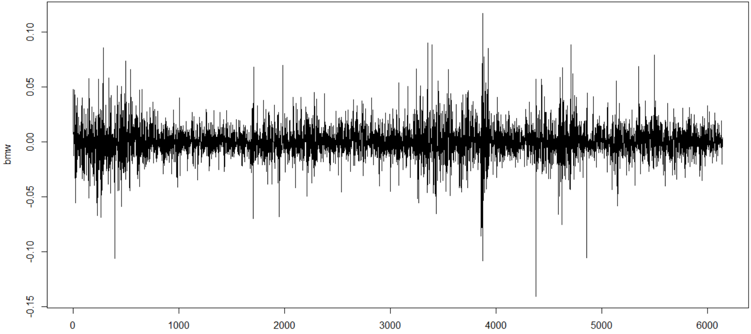


9. GARCH

Ruppert 2011, Chapter 18

Volatility clustering

- Financial time series often exhibit **volatility clustering** (extreme values followed by extreme values, large variance followed by large variance)
- Daily log return of BMW



- In an ARMA(p,q) model with **Gaussian white noise**,

$$\phi(L)(Y_t - \mu) = \theta(L)a_t, \quad a_t = \sigma\epsilon_t$$

where $\epsilon_t \sim N(0,1)$ are i.i.d. and σ is a constant

$$\mathbb{E}[a_t | a_{t-1}, a_{t-2}, \cdots] = 0$$

$$\mathrm{var}(a_t | a_{t-1}, a_{t-2}, \cdots) = \sigma^2$$

Large variance yesterday has no impact on the variance today

- In an ARMA(p,q) model with GARCH volatility, the constant σ is replaced by σ_t **that depends on previous values of a_t and σ_t**

ARCH(1)

- In an ARCH(1) model

$$a_t = \sigma_t \epsilon_t$$

$$\sigma_t = \sqrt{w + \alpha_1 a_{t-1}^2}$$

where $w > 0, 0 \leq \alpha_1 < 1$ and $\epsilon_t \sim N(0, 1)$ is a Gaussian white noise

- Denote $\mathbb{E}[\cdot|\mathcal{F}_{t-1}] := \mathbb{E}[\cdot|a_{t-1}, \epsilon_{t-1}, \sigma_{t-1}, a_{t-2}, \epsilon_{t-2}, \sigma_{t-2}, \cdots]$

$$\begin{aligned} \mathbb{E}[a_t|\mathcal{F}_{t-1}] &= \mathbb{E}[\sigma_t \epsilon_t|\mathcal{F}_{t-1}] \\ &= \sigma_t \mathbb{E}[\epsilon_t|\mathcal{F}_{t-1}] = 0 \\ \text{var}(a_t|\mathcal{F}_{t-1}) &= \mathbb{E}[a_t^2|\mathcal{F}_{t-1}] \\ &= \sigma_t^2 \mathbb{E}[\epsilon_t^2|\mathcal{F}_{t-1}] = \sigma_t^2 \end{aligned}$$

- ARCH refers to **A**uto**R**egressive **C**onditional **H**eteroscedasticity ("different variance")

$$a_t = \sqrt{w + \alpha_1 a_{t-1}^2} \epsilon_t$$

- If a_{t-1} is extreme, $\sigma_t^2 = w + \alpha_1 a_{t-1}^2$ is large, $a_t = \sigma_t \epsilon_t$ tends to be extreme as well, $\sigma_{t+1}^2 = w + \alpha_1 a_t^2$ also tends to be large
- Extreme values followed by extreme values; large variance followed by large variance (volatility clustering)
- Clustering gradually disappears since $0 \leq \alpha_1 < 1$
- a_t is stationary when $0 \leq \alpha_1 < 1$

Unconditional mean/variance

- Unconditional mean of a_t

$$\mathbb{E}[a_t] = \mathbb{E}[\mathbb{E}[a_t|\mathcal{F}_{t-1}]] = 0$$

- Unconditional variance of a_t

$$\begin{aligned} \text{var}(a_t) &= \mathbb{E}[a_t^2] = \mathbb{E}[\mathbb{E}[a_t^2|\mathcal{F}_{t-1}]] = \mathbb{E}[\sigma_t^2] \\ &= \mathbb{E}[w + \alpha_1 a_{t-1}^2] = w + \alpha_1 \text{var}(a_{t-1}) \end{aligned}$$

- Denote the unconditional variance by $\gamma_a(0)$

$$\gamma_a(0) = w + \alpha_1 \gamma_a(0), \quad \gamma_a(0) = \frac{w}{1 - \alpha_1}$$

- Lag h autocovariance of a_t

$$\begin{aligned}\gamma_a(h) &= \mathrm{cov}(a_t, a_{t-h}) \\ &= \mathbb{E}[a_t a_{t-h}] \\ &= \mathbb{E}[\mathbb{E}[a_t a_{t-h}|\mathcal{F}_{t-1}]] \\ &= \mathbb{E}[a_{t-h}\mathbb{E}[a_t|\mathcal{F}_{t-1}]] \\ &= 0, \;\; h \geq 1 \\ \rho_a(h) &= 0, h \geq 1\end{aligned}$$

- The ARCH process a_t is a white noise (zero autocorrelation), but dependent

Fourth moment

- Assume that $0 \leq \alpha_1 < 1/\sqrt{3}$ and a_t is fourth-order stationary

$$\begin{aligned}\mathbb{E}[a_t^4] &= \mathbb{E}[\mathbb{E}[(w + \alpha_1 a_{t-1}^2)^2 \epsilon_t^4 | \mathcal{F}_{t-1}]] \\ &= \mathbb{E}[(w + \alpha_1 a_{t-1}^2)^2 \mathbb{E}[\epsilon_t^4 | \mathcal{F}_{t-1}]] \\ &= 3(w^2 + 2w\alpha_1\gamma_a(0) + \alpha_1^2 \mathbb{E}[a_{t-1}^4])\end{aligned}$$

- Note that $\gamma_a(0) = w/(1-\alpha_1)$

$$\mathbb{E}[a_t^4] = \frac{3(w^2 + 2w\alpha_1\gamma_a(0))}{1 - 3\alpha_1^2} = \frac{3w^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

a_t is heavy tailed

- a_t is conditionally normal

$$a_t | a_{t-1} \sim N(0, w + \alpha_1 a_{t-1}^2)$$

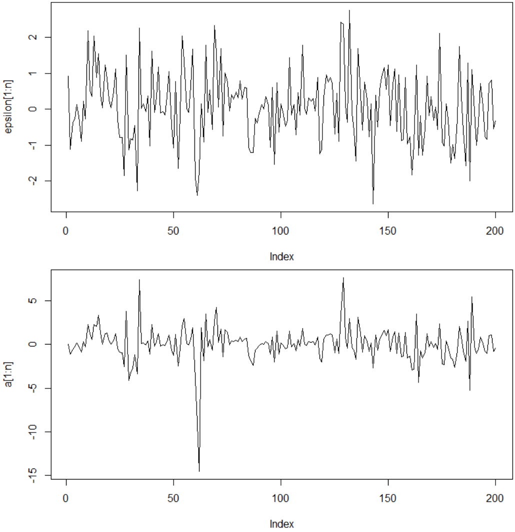
- $a_t = \sqrt{w + \alpha_1 a_{t-1}^2} \epsilon_t$ is non-Gaussian with kurtosis greater than 3 when $0 < \alpha_1 < 1/\sqrt{3}$
- Note that $\text{var}(a_t) = \gamma_a(0) = w/(1 - \alpha_1)$. Kurtosis of a_t

$$\frac{\mathbb{E}[a_t^4]}{\text{var}(a_t)^2} = \frac{3(1 - \alpha_1^2)}{1 - 3\alpha_1^2} > 3, \quad 0 < \alpha_1 < 1/\sqrt{3}$$

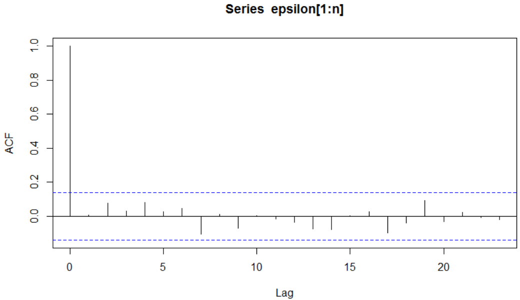
- Consider ARCH(1) with $w = 1, \alpha_1 = 0.85$

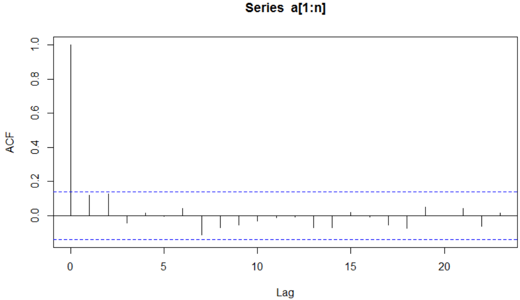
```
n=200
a=rep(0,n)
epsilon=rnorm(n)
w=1
alpha=0.85
for (i in 2:n){a[i]=sqrt(w+alpha*(a[i-1]^2))*epsilon[i]}
plot(epsilon[1:n],type="l")
plot(a[1:n],type="l")
acf(epsilon[1:n])
acf(a[1:n])
```

- a_t shows volatility clustering: the 60th, 61st, 62nd values are -3.93, -9.05, -14.52



- Both ϵ_t and a_t are not autocorrelated





ARMA/ARCH(1)

- As a white noise, the ARCH(1) process a_t can be incorporated into any ARMA(p,q) model
- Consider an **AR(1)/ARCH(1)** model

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + a_t, a_t = \sqrt{w + \alpha_1 a_{t-1}^2}\epsilon_t, \epsilon_t \sim N(0,1)$$

- The AR(1) component models conditional mean; the ARCH(1) models conditional variance

$$\mathbb{E}[Y_t | \mathcal{F}_{t-1}] = \mu + \phi_1(Y_{t-1} - \mu)$$

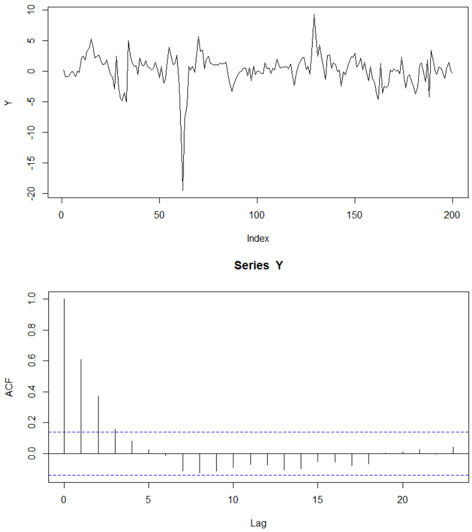
$$\text{var}(Y_t | \mathcal{F}_{t-1}) = \sigma_t^2 = w + \alpha_1 a_{t-1}^2$$

Simulated AR(1)/ARCH(1)

- Simulate AR(1)/ARCH(1) with $\mu = 0.2, \phi_1 = 0.5$ (ARCH(1) as before)

```
mu=0.2
phi=0.5
Y=rep(mu,n)
for (i in 2:n){Y[i]=mu+phi*(Y[i-1]-mu)+a[i]}
plot(Y,type="l")
acf(Y)
```

- The AR(1)/ARCH(1) process shows volatility clustering; autocorrelation structure of the AR(1)/ARCH(1) process remains the same as before



ARCH($\bar{p}$)

- ARCH($\bar{p}$) process:

$$a_t = \sigma_t \epsilon_t$$

$$\sigma_t = \sqrt{w + \alpha_1 a_{t-1}^2 + \cdots + \alpha_{\bar{p}} a_{t-\bar{p}}^2}$$

where $\epsilon_t \sim N(0, 1)$ is a Gaussian white noise, $w > 0$, $\alpha_i \geq 0$, $\alpha_1 + \cdots + \alpha_{\bar{p}} < 1$

- Zero conditional and unconditional means

$$\mathbb{E}[a_t | \mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[\epsilon_t | \mathcal{F}_{t-1}] = 0$$

$$\mathbb{E}[a_t] = \mathbb{E}[\mathbb{E}[a_t | \mathcal{F}_{t-1}]] = 0$$

- Conditional variance

$$
\begin{aligned}
\text{var}(a_t|\mathcal{F}_{t-1}) &= \mathbb{E}[a_t^2|\mathcal{F}_{t-1}] \\
&= \sigma_t^2 \mathbb{E}[\epsilon_t^2|\mathcal{F}_{t-1}] \\
&= \sigma_t^2 \;=\; w + \alpha_1 a_{t-1}^2 + \cdots + \alpha_{\bar{p}} a_{t-\bar{p}}^2
\end{aligned}
$$

- Unconditional variance

$$
\begin{aligned}
\gamma_a(0) &= \text{var}(a_t) \;=\; \mathbb{E}[a_t^2] \\
&= \mathbb{E}[\mathbb{E}[a_t^2|\mathcal{F}_{t-1}]] \\
&= w + \alpha_1 \mathbb{E}[a_{t-1}^2] + \cdots + \alpha_{\bar{p}} \mathbb{E}[a_{t-\bar{p}}^2] \\
&= w + \alpha_1 \gamma_a(0) + \cdots + \alpha_{\bar{p}} \gamma_a(0) \\
\gamma_a(0) &= \frac{w}{1 - (\alpha_1 + \cdots + \alpha_{\bar{p}})}
\end{aligned}
$$

- ARCH models often are not able to generate persistent volatility
- GARCH($\bar{p}$, $\bar{q}$)

$$a_t = \sigma_t \epsilon_t$$

$$\sigma_t = \sqrt{w + \alpha_1 a_{t-1}^2 + \cdots + \alpha_{\bar{p}} a_{t-\bar{p}}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_{\bar{q}} \sigma_{t-\bar{q}}^2}$$

where $\epsilon \sim N(0,1)$ is a Gaussian white noise, $w > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$, $\alpha_1 + \cdots + \alpha_{\bar{p}} + \beta_1 + \cdots + \beta_{\bar{q}} < 1$

- Consider a GARCh(1,1) model

$$a_t = \sigma_t \epsilon_t$$

$$\sigma_t = \sqrt{w + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2}$$

where $\epsilon \sim N(0,1)$ is a Gaussian white noise, $w > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, $\alpha_1 + \beta_1 < 1$

- a_t has zero conditional and unconditional means; conditional variance of a_t

$$\text{var}(a_t | \mathcal{F}_{t-1}) = \sigma_t^2$$

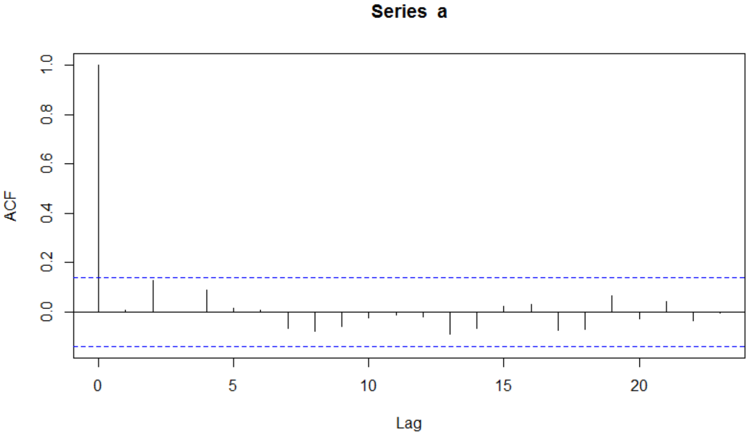
- Unconditional variance of GARCH(1,1)

$$
\begin{aligned}
\mathrm{var}(a_t) &= \mathbb{E}[a_t^2] \\
&= \mathbb{E}[\mathbb{E}[a_t^2|\mathcal{F}_{t-1}]] \\
&= \mathbb{E}[\sigma_t^2] \\
&= \mathbb{E}[w + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2] \\
&= w + \alpha_1 \mathrm{var}(a_{t-1}) + \beta_1 \mathrm{var}(a_{t-1}) \\
\gamma_a(0) &= \frac{w}{1-(\alpha_1+\beta_1)}
\end{aligned}
$$

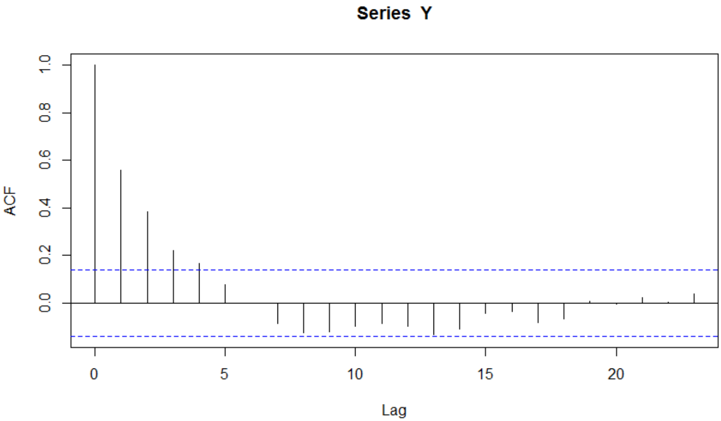
- Simulate AR(1)/GARCH(1,1) with $w = 1$, $\alpha_1 = 0.35$, $\beta_1 = 0.5$, AR(1) parameters as before

```
> n=200
> a=rep(0,n)
> epsilon=rnorm(n)
> w=1
> alpha=0.35
> beta=0.5
> sigma=rep(sqrt(2),n)
> mu=0.2
> phi=0.5
> Y=rep(mu,n)
> for (i in 2:n){
+ sigma[i]=sqrt(w+alpha*(a[i-1]^2)+beta*(sigma[i-1]^2))
+ a[i]=sigma[i]*epsilon[i]
+ Y[i]=mu+phi*(Y[i-1]-mu)+a[i]
+ }
> acf(a)
> acf(Y)
> plot(a,type="l")
> plot(Y,type="l")
```

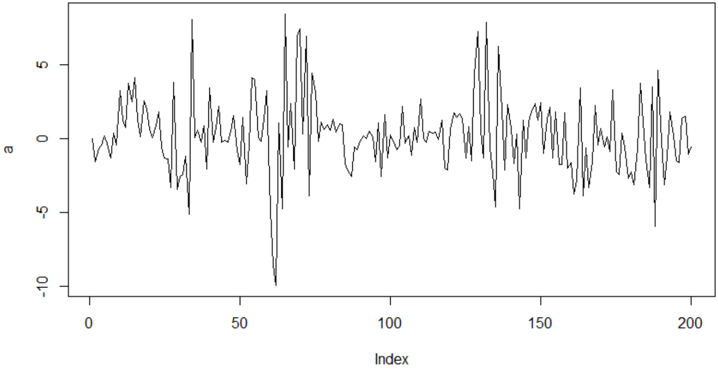
- a_t is white noise with zero autocorrelation



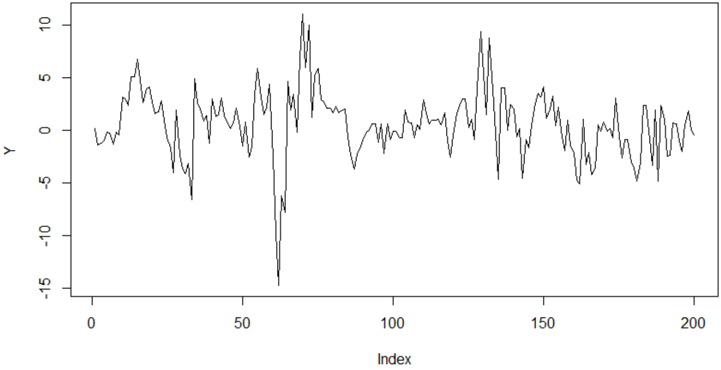
- Autocorrelation structure of Y_t is still that of an AR(1)



- a_t shows more persistent volatility clustering



- Y_t shows more persistent volatility clustering



Fitting BMW data

- Without GARCH, AR(2) is preferred

```
library(evir)
data(bmw)
library(forecast)
auto.arima(bmw)
fit=garchFit(formula = arma(1,0) + garch(1,1), data =
bmw, cond.dist="norm")
fit
```

```
Series: bmw
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1      ar2  intercept
      0.0833  -0.0273      3e-04
s.e.  0.0128   0.0128      2e-04

sigma^2 estimated as 0.0002161:  log likelihood=17214.63
AIC=-34421.27   AICc=-34421.26   BIC=-34394.37
```

- With GARCH, AR(1)/GARCH(1,1) provides the best fit

Model	log-likelihood	# of parameters	AIC
AR(2)	17214.6	4	-34421.3
AR(2)/GARCH(1,1)	17757.4	6	-35502.8
AR(1)/GARCH(1,1)	**17757.2**	**5**	**-35504.3**

```
Call:
 garchFit(formula = ~arma(1, 0) + garch(1, 1), data = bmw, cond.dist = "norm")

Mean and Variance Equation:
 data ~ arma(1, 0) + garch(1, 1)
<environment: 0x0622451c>
 [data = bmw]

Conditional Distribution:
 norm

Coefficient(s):
        mu          ar1        omega       alpha1        beta1
4.0092e-04   9.8596e-02   8.9043e-06   1.0210e-01   8.5944e-01

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu      4.009e-04   1.579e-04    2.539   0.0111 *
ar1     9.860e-02   1.431e-02    6.888 5.65e-12 ***
omega   8.904e-06   1.449e-06    6.145 7.97e-10 ***
alpha1  1.021e-01   1.135e-02    8.994  < 2e-16 ***
beta1   8.594e-01   1.581e-02   54.348  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 17757.16    normalized:  2.889222
```

- Can AR(1)/GARCH(1,1) be further improved? What else are missing from the model?
- Check the residuals: for AR(1)/GARCH(1,1)

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + a_t$$

$$a_t = \sigma_t \epsilon_t, \quad \sigma_t = \sqrt{w + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2}$$

- Residuals and standardized residuals

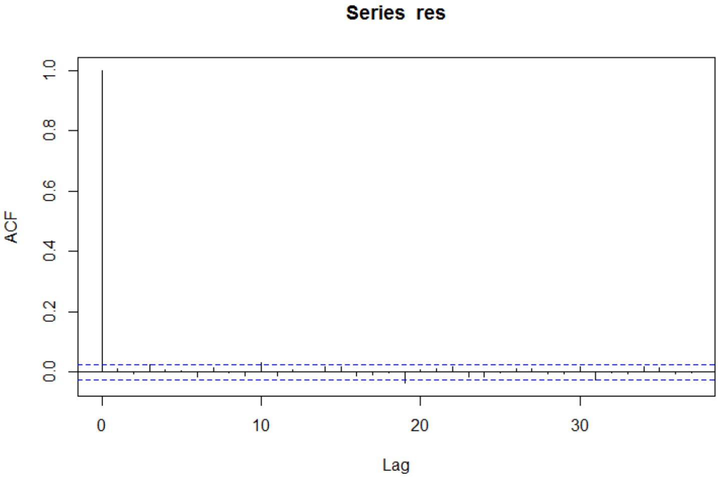
$$\hat{a}_t = Y_t - \hat{\mu} - \hat{\phi}_1(Y_{t-1} - \hat{\mu})$$

$$\hat{\epsilon}_t = \hat{a}_t / \hat{\sigma}_t, \quad \hat{\sigma}_t = \sqrt{\hat{w} + \hat{\alpha}_1 \hat{a}_{t-1}^2 + \hat{\beta}_1 \hat{\sigma}_{t-1}^2}$$

- Does the standardized residual still show autocorrelation? Is it normal?

```
res=residuals(fit, standardize = T)
acf(res)
acf(res*res)
qqnorm(res)
qqline(res)
```

- ACF plot shows that the standardized residual is not autocorrelated (also for squared standardized residual)



- Normal probability plot shows that $\epsilon_t \sim N(0,1)$ is not sufficient; a heavy tailed distribution is preferred