Chapter 7 Conditional Heteroscedastic Models (Part A)

For time series $\{y_t\}$, denote the information up to time t by

$$I_t = \sigma\{y_t, y_{t-1}, y_{t-2}, ...\}.$$

(A) The conditional mean $\mu_t = E(y_t|I_{t-1})$ can be used to predict y_t based on past information I_{t-1}.

(B) Note that the auto-correlation coefficients in y_t are "almost" 0; the prediction is almost impossible

(C) How about the conditional variance (risk, or volatility)

$$\sigma_t^2 = \mathbf{cov}(y_t|I_{t-1}) = E\{(y_t - \mu_t)^2|I_{t-1}\}$$

(D) Definition: asset **volatility** is conditional standard deviation of the asset returns.

(E) Importance of σ_t

(a) Option (derivative) pricing, e.g., Black-Scholes formula

(b) Risk management, e.g. value at risk (VaR)

(c) Asset allocation, e.g., minimum-variance portfolio.

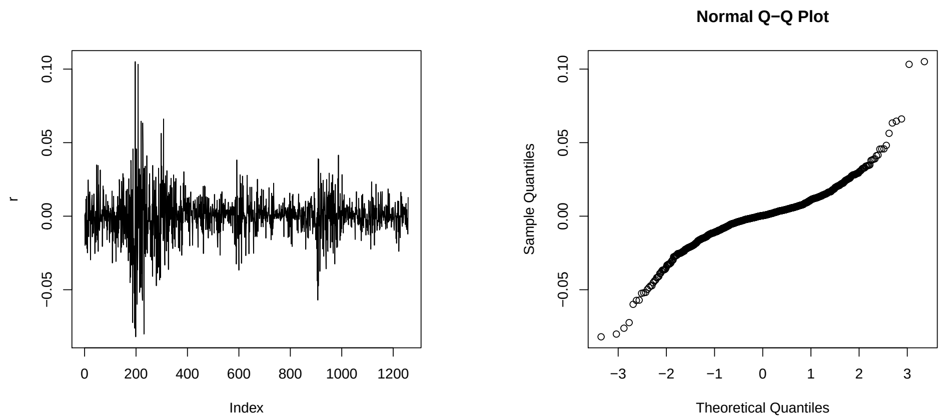
(d) Interval forecasts

(e) σ_t^2 can also help to explain the special patterns in financial data, such as the heavy tail, and **volatility clustering**.

Example 0.1 *For the SP500, the daily return r_t from 2008-2013, the oringinal return r_t and standardized residuals*

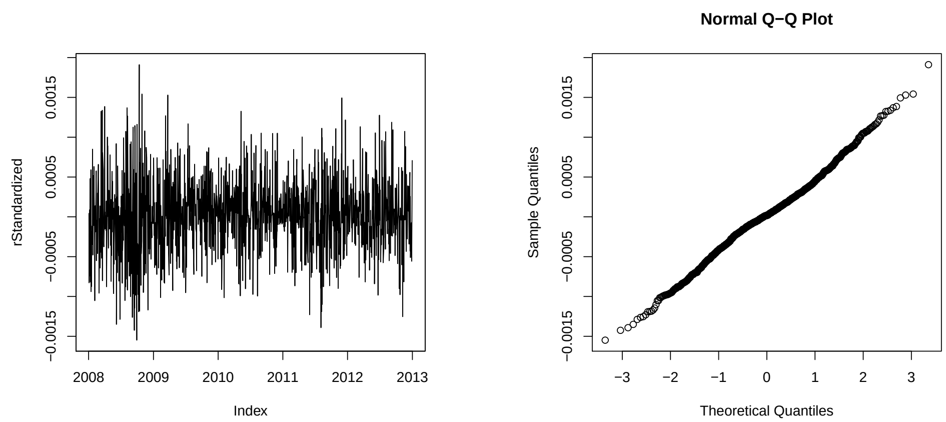
$$(r_t - \mu_t)/\sigma_t$$

are shown below (here $\mu_t = 0$ is assumed). It can be seen that after removing the conditional *standardization, the return follows normal.*

original return: 

Stylized facts: volatility clustering, *i.e. "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes.", or*

$$corr(|r_t|, |r_{t+1}|) > 0$$

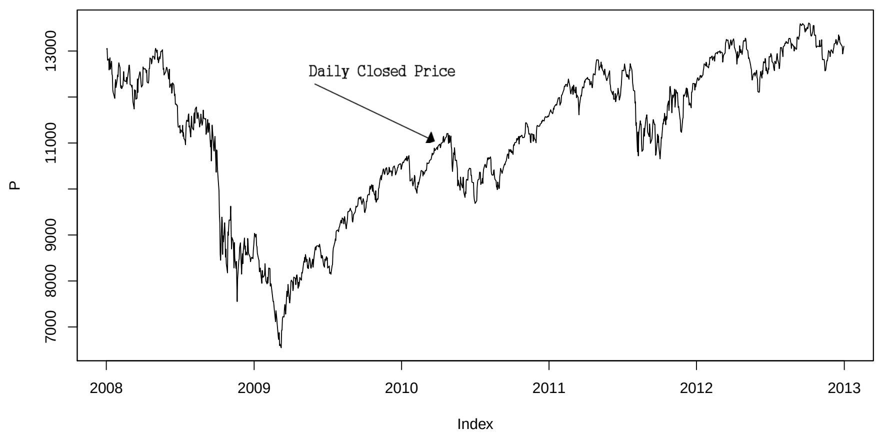
return with σ_t removed: 

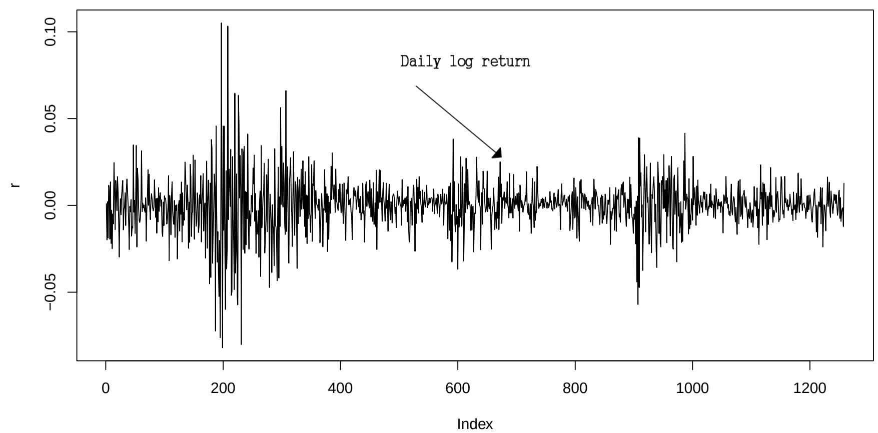
(F) Given I_{t-1}, if we can observe y_t repeatedly (actually we can; all the trading prices in a day), then we can estimate σ_t^2. This is what people do by using high frequency data. Such as the realized-volatility; see

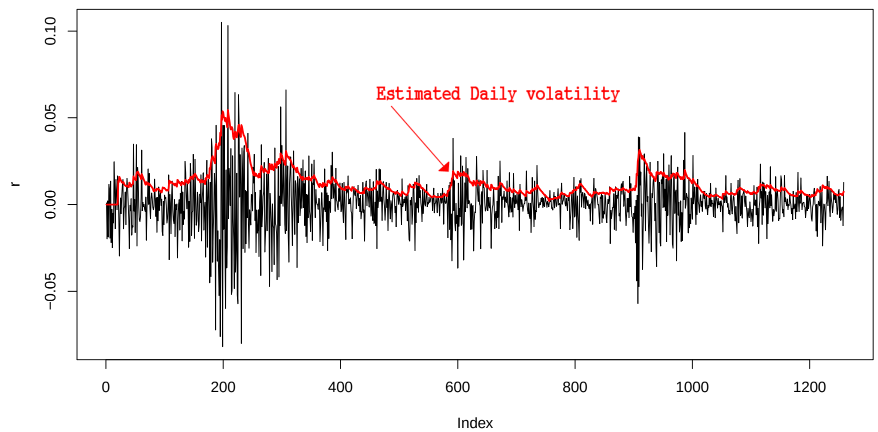
http://en.wikipedia.org/wiki/Realizedvariance

(G) A key problem: σ_t is not directly observable!! So its calculation is an important topic of research

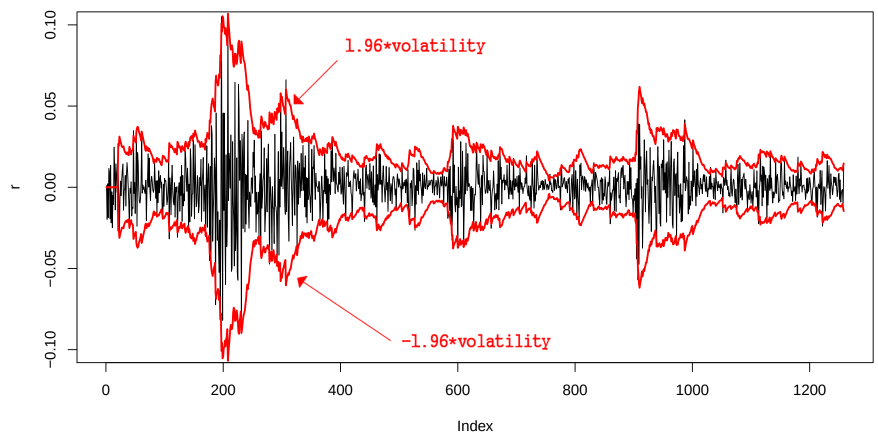
Example 0.2 *For DJI from 1/1/2008-1/4/2013; see code06A1.R.*







Estimated Daily volatility
r
Index

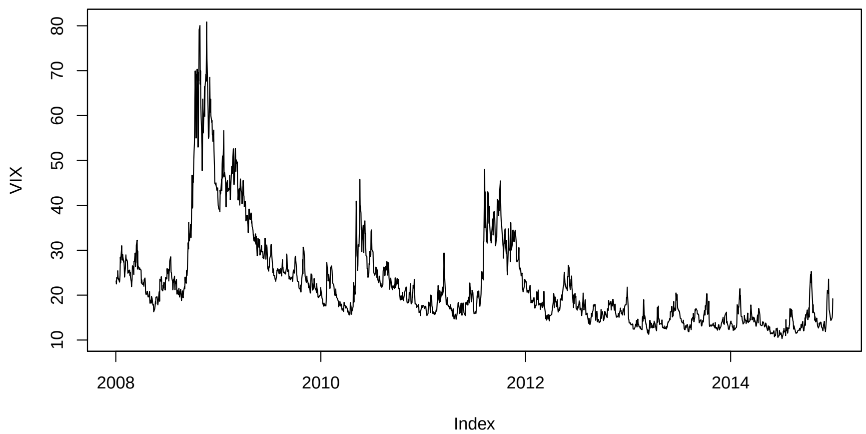


1.96*volatility
-1.96*volatility
r
Index

How to calculate volatility?

(A) Use high-frequency data: Realized volatility of daily returns in recent literature: use intraday log returns.

(B) Implied volatility of options data, e.g, VIX (http://en.wikipedia.org/wiki/VIX). See Figure below, the data is available in Yahoo https://sg.finance.yahoo.com/



(C) weighted average (Exponential smoothing, ARCH, GARCH)

Weighted average approach: exponential smoothing, ARCH, GARCH

(a) Suppose the returns u_i have mean 0, otherwise consider $u_i - E(u_i)$.

(b) We estimate the variance by

$$\sigma_t^2 = \sum_{i=1}^{m} w_i u_{t-i}^2 \quad \text{(for prediction)}$$

or

$$\sigma_t^2 = \sum_{i=0}^{m-1} w_i u_{t-i}^2 \quad \text{(for fitting)}$$

where $w_1 + ... + w_m = 1$. Usually, w_i decreases with i.

(c) m can be infinity providing $w_1 + w_2 + ... = 1$, a special case is

$$w_k = (1 - \lambda)\lambda^{k-1}, \quad k = 1, ...$$

$$\sigma_t^2 = (1 - \lambda)\sum_{i=1}^{\infty} \lambda^{i-1} u_{t-i}^2. \text{ i.e. } m = \infty$$

(d) This leads to a recurrence formula,

$$\sigma_t^2 = (1-\lambda)u_{t-1}^2 + \lambda\sigma_{t-1}^2$$

popular choices of λ are close to 1.

For example, RiskMetrics, a financial risk management company, tends to use a λ of 0.94.

See the first plot on page 6.

(e) Note that in the above model, the sum of coefficients of σ_{t-1}^2 and u_{t-1}^2 are respectively λ and $1-\lambda$, we can make it more general, leading to GARCH model

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta\sigma_{t-1}^2$$

with $\omega \geq 0, \alpha \geq 0, \beta \geq 0$, but $\alpha + \beta \leq 1$.

(f) More general case is the GARCH(p,q) model [Generalized Autoregressive Conditional Heteroskedasticity] to be discussed later.

(g) A special case of GARCH(1,1) is that

$$\alpha + \beta = 1, \quad \text{or} \quad \alpha = \lambda, \quad \beta = 1 - \lambda.$$

which is also called IGARCH(1,1) [Integrated Generalized Autoregressive Conditional Heteroskedasticity]

(h) However, if $\alpha + \beta > 1$, the time series will diverge to infinity

Econometric modeling of the conditional variance/volatility:

For return r_t, denote the conditional mean by μ_t

$$r_t = \mu_t + u_t;$$

and usually we model μ_t by ARMA(p,q) model

$$\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{i=1}^{q} \psi_i u_{t-i};$$

Later on, we shall omit the discussion of μ_t in most cases.

Volatility models are concerned with the modelling of

$$\begin{aligned} \sigma_t^2 &= \mathbf{cov}(r_t \Big| r_{t-1}, r_{t-2}, ...) \\ &= \mathbf{cov}(u_t \Big| r_{t-1}, r_{t-2}, ...) \end{aligned}$$

the conditional variance of a return.

Univariate volatility models

(1) Autoregressive conditional heteroscedastic (ARCH) model of Engle (1982),

(2) Generalized Autoregressive Conditional Heteroscedastic (GARCH) model of Bollerslev (1986),

(3) GARCH-M (GARCH-in-Mean) models,

(4) IGARCH models (used by RiskMetrics), Exponential GARCH (EGARCH), TGARCH, ...

(5) Asymmetric parametric ARCH (APARCH) models of Ding, Granger and Engle (1994)

(6) Stochastic volatility (SV) models

ARCH(m) model

$$\begin{aligned} r_t &= \mu_t + u_t, \\ u_t &= \sigma_t \epsilon_t; \quad \sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + ... + \alpha_m u_{t-m}^2 \end{aligned}$$

where $\{\epsilon_t\}$ is a sequence of iid random variables (and independent of $u_{t-1}, u_{t-2}, ...$) with

$$E(\epsilon_t) = 0, \quad \mathbf{cov}(\epsilon_t) = 1 \ \text{(if it exists)},$$

and, $\omega > 0$ and $\alpha_i \geq 0$ for $i > 0$,

$$\epsilon_t \text{ is independent of } \{u_s, s < t\}$$

$$\epsilon_t \quad \text{and} \quad \{\sigma_s, s \leq t\} \text{ are independent}$$

We call $u_t = r_t - \mu_t$ "<u>residuals</u>" and ϵ_t "<u>standardized residuals</u>".

Commonly used distribution of ϵ_t:

(i) Standard normal,

(ii) standardized Student-t,

Properties of ARCH models

Consider an ARCH(1) model

$$r_t = \mu_t + u_t, \quad u_t = \sigma_t \epsilon_t; \quad \sigma_t^2 = \omega + \alpha_1 u_{t-1}^2;$$

where $\omega > 0$ and $\alpha_1 \geq 0$.

(1) $E(u_t) = 0$ and $E(u_t | r_{t-1}, r_{t-2}, ...) = 0$

(2) u_t is called the residuals, but ϵ_t is called standardized residuals, or innovation

(3) the model can be written as

$$u_t^2 = \omega + \alpha_1 u_{t-1}^2 + \eta_t$$

where $\eta_t = \sigma_t^2(\epsilon_t^2 - 1)$.

(4) If $\alpha_1 < 1$, Eu_t^2 is constant, and u_t^2 is also stationary

(5) u_t is WN, but u_t^2 (or $u_t^2 - Eu_t^2$) is not White Noise.

(6) assuming $\mathbf{cov}(u_t)$ is constant for all t[1], then

$$\mathbf{cov}(u_t) = \frac{\omega}{(1-\alpha_1)}, \quad \text{if } 0 < \alpha_1 < 1$$

[Proof:

$$E(u_t^2) = \omega + \alpha_1 E u_{t-1}^2$$

i.e.

$$\text{2nd moment } \mu_2 = \mathbf{cov}(u_t) = \omega + \alpha_1 \mathbf{cov}(u_t). \qquad]$$

(7) Under normality, i.e. $\epsilon_t \sim N(0,1)$, and assuming the moments are constants for all t[2],

$$\text{4th moment } \mu_4 = \frac{3\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

[1] actually this assumption is true, please try to prove it

[2] this assumption is also true

provided $0 < \alpha_1^2 < 1/3$. However, the 4th moment does not exist if $\alpha_1^2 > 1/3$.

[Proof: by (3),

$$
\begin{aligned}
u_t^4 &= \omega^2 + 2\omega\alpha_1 u_{t-1}^2 + \alpha_1^2 u_{t-1}^4 + \eta_t^2 \\
&\quad + 2(\omega + \alpha_1 u_{t-1}^2) * \eta_t.
\end{aligned}
$$

Taking expectation, we have by the assumption that ϵ_t is independent of σ_t

$$E\{(\omega + \alpha_1 u_{t-1}^2) * \eta_t\} = 0,$$

and the normality of ϵ_t,

$$E\eta_t^2 = E\sigma_t^4 E((\epsilon_t^2 - 1)^2) = 2E\sigma_t^4.$$

By the model $\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2$, we have

$$E\sigma_t^4 = \omega^2 + 2\omega * \alpha_1 \mu_2 + \alpha_1^2 \mu_4$$

Taking expectation to the first equation, we have

$$\mu_4 = 3(\omega^2 + 2\omega\alpha_1\mu_2) + 3\alpha_1^2\mu_4$$

By (6),

$$\begin{aligned}
\mu_4 &= 3(\omega^2 + 2\omega\alpha_1\frac{\omega}{1-\alpha_1}) + 3\alpha_1^2\mu_4 \\
&= 3\omega^2(1 + 2\frac{\alpha_1}{1-\alpha_1}) + 3\alpha_1^2\mu_4 = 3\omega^2\frac{1+\alpha_1}{1-\alpha_1} + 3\alpha_1^2\mu_4
\end{aligned}$$

Thus the result follows.]

(8) The kurtosis is

$$\frac{\mu_4}{\mu_2^2} = 3\frac{(1-\alpha_1^2)}{(1-3\alpha_1^2)} > 3$$

(9) This explains why the data has heavy tail if it indeed follows a ARCH model! even if the innovation is indeed normally distributed.

Estimation of ARCH using MLE

The log-likelihood function for normal distributed u_t is

$$\begin{aligned} L(\omega, \alpha_1, ..., \alpha_m) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=m+1}^{n}\{\log(\omega + \alpha_1 u_{t-1}^2 + ... + \alpha_m u_{t-m}^2) \\ &\qquad + \frac{u_t^2}{\omega + \alpha_1 u_{t-1}^2 + ... + \alpha_m u_{t-m}^2}\} \end{aligned}$$

The MLE is

$$(\hat{\omega}, ..., \hat{\alpha}_m) = arg \max_{\omega, \alpha_1, ..., \alpha_m} \{L(\omega, \alpha_1, ..., \alpha_m)\}$$

The selection of m can be done by AIC or BIC

$$AIC(m) = -2L(\omega, \alpha_1, ..., \alpha_m) + 2\frac{m+1}{n}$$

and

$$BIC(m) = -2L(\omega, \alpha_1, ..., \alpha_m) + \log(n)\frac{m+1}{n}$$

Prediction of σ^2_{t+h} based on $u_1, ..., u_t$

$$\hat{\sigma}^2_{t+1} = \omega + \alpha_1 u^2_{t+1-1} + ... + \alpha_m u^2_{t+1-m}$$

Because ($u^2_{t+1} = \sigma^2_{t+1} + \sigma^2_{t+1}(\epsilon^2_{t+1} - 1)$) and

$$\begin{aligned} \sigma^2_{t+2} &= \omega + \alpha_1 u^2_{t+2-1} + \alpha_2 u^2_{t+2-2} + ... + \alpha_m u^2_{t-m+2} \\ &= \omega + \alpha_1 \sigma^2_{t+2-1} + \alpha_2 u^2_{t+2-2} + ... + \alpha_m u^2_{t+2-m} + \alpha_1 \sigma^2_{t+1}(\epsilon^2_{t+1} - 1) \\ &\approx \omega + \alpha_1 \sigma^2_{t+2-1} + \alpha_2 u^2_{t+2-2} + ... + \alpha_m u^2_{t+2-m} \end{aligned}$$

Thus

$$\hat{\sigma}^2_{t+2} = \omega + \alpha_1 \hat{\sigma}^2_{t+2-1} + \alpha_2 u^2_{t+2-2} + ... + \alpha_m u^2_{t+2-m}$$

$$\hat{\sigma}^2_{t+3} = \omega + \alpha_1 \hat{\sigma}^2_{t+3-1} + \alpha_2 \hat{\sigma}^2_{t+3-2} + ... + \alpha_m u^2_{t+3-m}$$

R

```
library(fGarch)

[object] = garchFit(~ arma(p,q) + garch(m, s),

        data = [mydata],

        cond.dist = c("norm", "snorm", "ged", "sged",

                "std", "sstd"), shape=??, trace = TRUE)

summary([object])

predict([object], n.ahead = ??)    # please note that this

                                    # is to predict the mean in the older version

# in above, shape=??  only for the degree-of-freedom of

   t-distribution
```

[object]@residuals: (raw, unstandardized) residual values u_t.

[object]@fitted: the fitted values.

[object]@h.t: the conditional variances (h.t = sigma.t^2).

[object]@sigma.t: the conditional standard deviation.

[object]mGARCHt@fit$se.coef: estimated parameters in the model

```
plot([object])

# plot has a number of plots for model check

# another useful QQ plot (against t-distribution)

qqplot(qt((1:n)/(n+1), df = ??), [data])

# or qqplot(qt(ppoints(n), df = ??)  [data] )

# where n is the length of [data]
```

Example 0.3 *We fit ARCH(5) to the daily return of DJI, $r_t * 100$ (for easy of exposition). The fitted ARCH(5) model is*

$$r_t * 100 = 0.003659 + u_t; \quad u_t = \sigma_t \epsilon_t; \quad \epsilon_t \sim N(0;1)$$

and

$$\begin{aligned} \sigma_t^2 &= 0.388298 + 0.042448 u_{t-1}^2 + 0.203262 u_{t-2}^2 \\ &\quad + 0.197531 u_{t-3}^2 + 0.223231 u_{t-4}^2 + 0.188031 u_{t-5}^2 \end{aligned}$$

Model checking statistics indicate that there are some higher order dependence in the volatility. By checking the QQ-plot, it seems that the innovation ϵ_t does not follow normal.

```
Call:
 garchFit(formula = ~garch(5, 0), data = r * 100, delta = 2, cond.dist = "norm",
    include.delta = F, trace = F)

Mean and Variance Equation:
 data ~ garch(5, 0)
<environment: 0x093c4874>
 [data = r * 100]

Conditional Distribution:
 norm

Coefficient(s):
      mu     omega    alpha1    alpha2    alpha3    alpha4    alpha5
0.003659  0.388298  0.042447  0.203263  0.197531  0.223230  0.188031

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.003659    0.026337    0.139   0.8895
omega   0.388298    0.042772    9.078  < 2e-16 ***
alpha1  0.042447    0.022453    1.890   0.0587 .
alpha2  0.203263    0.039950    5.088 3.62e-07 ***
alpha3  0.197531    0.039836    4.959 7.10e-07 ***
alpha4  0.223230    0.043942    5.080 3.77e-07 ***
alpha5  0.188031    0.038788    4.848 1.25e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Log Likelihood:
 -1995.987    normalized:  -1.586635

Description:
 Sat Mar 21 13:12:03 2015 by user: xia


Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  88.1211   0
 Shapiro-Wilk Test  R    W      0.982711  4.179757e-11
 Ljung-Box Test     R    Q(10)  15.43769  0.1169006
 Ljung-Box Test     R    Q(15)  18.39082  0.2426765
 Ljung-Box Test     R    Q(20)  20.57512  0.4225087
 Ljung-Box Test     R^2  Q(10)  34.98461  0.0001256224
 Ljung-Box Test     R^2  Q(15)  43.07804  0.0001531184
 Ljung-Box Test     R^2  Q(20)  48.14984  0.0004052813
 LM Arch Test       R    TR^2   48.73513  2.326924e-06

Information Criterion Statistics:
     AIC      BIC      SIC     HQIC
3.184399 3.212985 3.184338 3.195142
```

Based on the test, it seems there is still information in the fitted residuals.

We then try ARCH(8)

```
Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      3.659e-03   2.583e-02    0.142 0.887353
omega   2.786e-01   4.067e-02    6.851 7.32e-12 ***
alpha1  1.000e-08   2.154e-02    0.000 1.000000
alpha2  1.361e-01   3.508e-02    3.878 0.000105 ***
alpha3  1.273e-01   3.580e-02    3.556 0.000377 ***
alpha4  1.452e-01   4.131e-02    3.516 0.000439 ***
alpha5  1.393e-01   3.438e-02    4.051 5.09e-05 ***
alpha6  1.067e-01   3.609e-02    2.957 0.003105 **
alpha7  1.447e-01   3.651e-02    3.962 7.42e-05 ***
alpha8  6.818e-02   2.747e-02    2.482 0.013072 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 -1958.38    normalized:  -1.556741
```

```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  58.83563  1.675327e-13
 Shapiro-Wilk Test  R    W      0.9874148 5.978969e-09
 Ljung-Box Test     R    Q(10)  14.11174  0.1679557
 Ljung-Box Test     R    Q(15)  17.40699  0.295122
 Ljung-Box Test     R    Q(20)  19.36417  0.4982802
 Ljung-Box Test     R^2  Q(10)  8.637683  0.5667988
 Ljung-Box Test     R^2  Q(15)  17.28372  0.3021885
 Ljung-Box Test     R^2  Q(20)  21.43408  0.3719906
 LM Arch Test       R    TR^2   12.92092  0.3748148

Information Criterion Statistics:
     AIC      BIC      SIC     HQIC
3.129379 3.170216 3.129254 3.144727
```

The standardized residuals show no auto-correlation.

The fitted model ARCH(8) with Student-t innovations is

$$r_t * 100 = 0.003659 + u_t; \quad u_t = \sigma_t \epsilon_t; \quad \epsilon_t \sim t(7)$$

and

$$\begin{aligned}\sigma_t^2 &= 0.25564955 + 0.0000u_{t-1}^2 + 0.1314u_{t-2}^2 \\ &+0.1365u_{t-3}^2 + 0.1054u_{t-4}^2 + 0.1774u_{t-5}^2 \\ &+0.1298u_{t-6}^2 + 0.14718u_{t-7}^2 + 0.07543u_{t-8}^2\end{aligned}$$

By checking the QQ-plot, it seems that the innovation ϵ_t is more close to t(7).

see code06A2.R.

For prediction of volatility:

(i) by calling [object]@residuals, the last 8 residuals $r_t - \mu_t$ *are -0.747866173, 0.446203960, -0.915879087, -0.396824391, -0.190223445, -0.143142976, -1.218988145, 1.271440116.*

(ii) $\sigma^2_{T+1} = 0.25564955 + 0.0000 * 1.271440116^2 + 0.1314 * (-1.218988145)^2 + 0.1365 * (-0.143142976)^2 + 0.1054 * (-0.190223445)^2 + 0.1774 * (-0.396824391)^2 + 0.1298 * (-0.915879087)^2 + +0.14718 * (0.446203960)^2 + 0.07543 * (-0.747866173)^2 =$

(iii) $\sigma^2_{T+2} = 0.25564955 + 0.0000 * \sigma^2_{T+1} + 0.1314 * 1.271440116^2 + ... + 0.14718(-0.143142976)^2 + 0.07543 * (-1.218988145)^2 =$