Lecture 6
Autoregressive Conditional Heteroskedastic Models

CHEN Ying
FE5209 Financial Econometrics

Review on ARIMA

Population		Sample/Data	
Stochastic Process		Time Series	
White Noise	Theoretical behaviors: mean, var, ACF, PACF, roots	1. Empirical features: stationarity? How to estimate mean, var, ACF, PACF	-> ADF test -> visual inspection -> model selection
AR(p)		2. Model estimation: OLS, MLE, Yule-Walker	
MA(q)		3. Diagnostic checking: Q-test, normality	-> residuals
ARMA(p,q)		4. Forecast: 1-step and h-step	->var of forecast errors increase
I(d) (e.g. Random walk is I(1) process)			
ARIMA(p,d,q)			
Markov switching process			

Outline

- ARCH/GARCH Process.
- Estimation of the GARCH Models.
- Variants of the GARCH Model.
- Realized volatility

Readings
SDA chapter 18
FTS chapter 3 and 5
SFM chapter 12

Estimating conditional means and variances

We care volatility because expected asset returns are related to volatility (risk-return), we want to understand how to generate volatility and predict future volatility to estimate current asset prices.

Consider regression modeling with a ***constant*** conditional variance,

$$Var\left(Y_t \middle| X_{1,t}, \dots, X_{p,t}\right) = \sigma_\epsilon^2.$$

The general form for the regression of Y_t on $X_{1,t}, \dots, X_{p,t}$ is

$$Y_t = f\left(X_{1,t}, \dots, X_{p,t}\right) + \epsilon_t$$

where ϵ_t is independent of $X_{1,t}, \dots, X_{p,t}$ and has expectation of 0 and a constant conditional variance σ_ϵ^2.

Implied volatility

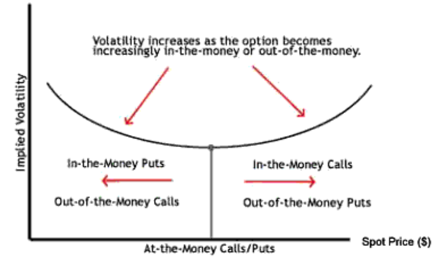
Implied volatility of the underlying asset can be inferred from observed option prices. Example: Consider a three-month European call option on a non-dividend paying stock with the following characteristics: c = \$1.875, S=\$21, X=\$20, r=10%.

Given c, S, X, r, and T, we can infer σ from the Black-Scholes option pricing formula. It is in fact equal to 23.5%.

$$C(S,t) = N(d_1)S - N(d_2)Xe^{-r(T-t)}$$

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right]$$

$$d_2 = d_1 - \sigma\sqrt{T-t}$$



where $N(\cdot)$ is the cumulative distribution function of the standard normal distribution, $T-t$ is the time to maturity, S is the spot price of the underlying asset, X is the strike price, r is the risk free rate and σ is the volatility of returns of the underlying asset.

Volatility smile: for a given expiration, options whose strike price differs substantially from the underlying asset's price have higher implied volatilities than that of at the money. These options are said to be either deep in-the-money or out-of-the-money.

The market prices of different options on the same asset are often used to calculate a *volatility term structure*. This is the relationship between the implied volatilities of the options and their maturities.

Estimating volatility from historical data

While traders use implied volatilities extensively, risk management is largely based on historical volatilities.

n number of observations
S_i stock price at the end of interval i
τ length of time interval

Let the return during the i-th interval be $r_i = \ln\left(\frac{S_i}{S_{i-1}}\right)$. The estimates of the standard deviation of u_i is

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{r_i} - \bar{r})^2}$$

where $\bar{r}$ is the mean of r_i. s is an estimate of $\sigma\sqrt{\tau}$ and σ can be estimated using s as follows: $\tilde{\sigma} = \frac{s}{\sqrt{\tau}}$.

i	S_i	S_i / S_{i-1}	$u_i = \ln(S_i / S_{i-1})$	μ	$(u_i - \mu)^2$
1	20.00				
2	20.10	1.005	0.499%	0.477%	0.000%
3	19.90	0.990	-1.000%	0.477%	0.022%
4	20.00	1.005	0.501%	0.477%	0.000%
5	20.50	1.025	2.469%	0.477%	0.040%
6	20.25	0.988	-1.227%	0.477%	0.029%
7	20.90	1.032	3.159%	0.477%	0.072%
8	20.90	1.000	0.000%	0.477%	0.002%
9	20.90	1.000	0.000%	0.477%	0.002%
10	20.75	0.993	-0.720%	0.477%	0.014%
11	20.75	1.000	0.000%	0.477%	0.002%
12	21.00	1.012	1.198%	0.477%	0.005%
13	21.10	1.005	0.475%	0.477%	0.000%
14	20.90	0.991	-0.952%	0.477%	0.020%
15	20.90	1.000	0.000%	0.477%	0.002%
16	21.25	1.017	1.661%	0.477%	0.014%
17	21.40	1.007	0.703%	0.477%	0.001%
18	21.40	1.000	0.000%	0.477%	0.002%
19	21.25	0.993	-0.703%	0.477%	0.014%
20	21.75	1.024	2.326%	0.477%	0.034%
21	22.00	1.011	1.143%	0.477%	0.004%
				$\Sigma (u_i - \mu)^2$	0.281%
				$(\Sigma (u_i - \mu)^2)/19$	0.015%
				$s=((\Sigma (u_i - \mu)^2)/19)$^0.5	**1.216%**

So, daily volatility equals 1.216%. What is the annual volatility? Assume that one year consists of 252 trading days.

$$1.216 \times \sqrt{252} = 19.3\%$$

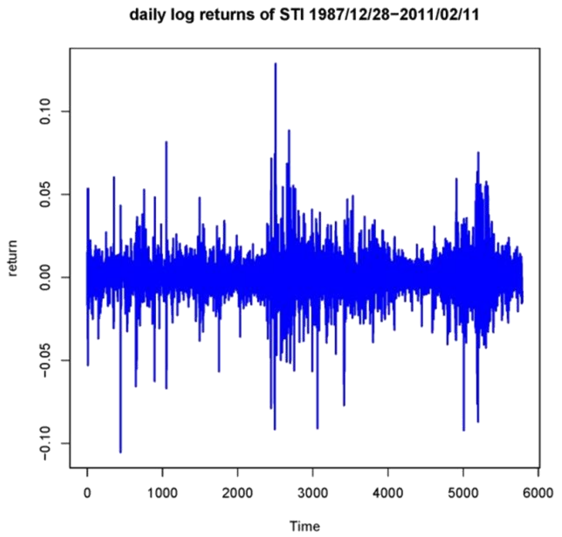
Conditional heteroscedasticity

Most popular option pricing models, such as BlackScholes, assume that the volatility of the underlying asset is constant. This assumption is far from perfect. In practice, the volatility of an asset, like the asset's price, is a stochastic variable. Unlike the asset price, it is not directly observable.

A widely observed phenomenon in finance, the so-called ***volatility clustering***, refers to the tendency of large changes in asset prices to be followed by large changes and small changes to be followed by small changes.

--reported by Mandelbrot in 1963.

Conditional variance is serially correlated. This is called **conditional heteroscedasticity**.



Estimate time-dependent conditional variances

Let $\sigma^2(X_{1,t}, \dots, X_{p,t})$ be the conditional variance of Y_t given $X_{1,t}, \dots, X_{p,t}$.

$$Y_t = f\left(X_{1,t}, \dots, X_{p,t}\right) + \sigma(X_{1,t}, \dots, X_{p,t}) z_t$$

where z_t has conditional (given $X_{1,t}, \dots, X_{p,t}$) mean equal to 0 and conditional variance equal to 1.The volatility function $\sigma(X_{1,t}, \dots, X_{p,t})$ should be nonnegative since it is a standard deviation.

- **Rolling window technique** $\tilde{\sigma}_t^2 = \frac{1}{M} \sum_{i=1}^{M} \mathrm{r_{t-i}}^2$
- **Recursive**: $\tilde{\sigma}_t^2 = \frac{1}{t-1} \sum_{i=1}^{t-1} \mathrm{r_{t-i}}^2$
- **Weighting scheme**: The formula above gives equal weights to each past return. Sometimes risk managers like to give more weight to more recent data, so they calculate volatility as: $\tilde{\sigma}_t^2 = \sum_{i=1}^{t} \alpha_i \mathrm{r_{t-i}}^2$, where α_i are the weights.

The ***Autoregressive conditional heteroskedasticity*** **(ARCH)** model and its generalization, the ***generalized autoregressive conditional heteroskedasticity*** **(GARCH)** model, provide a convenient framework to model volatility clustering.

ARCH(1) Process

Consider an *ARCH*(*1*) model for $y_t = f\left(X_{1,t}, \ldots, X_{p,t}\right) + \epsilon_t$:

$$\epsilon_t = \sigma_t\, z_t, \quad z_t \sim IID\ N(0,1)$$
$$\sigma_t^2 = a_0 + a_1 \epsilon_{t-1}^2$$

where ϵ_t can be understood as demeaned returns, σ_t denotes ***volatility*** – conditional standard deviation.

It is also popular to use h_t, denoting **the *conditional variance*** of ϵ_t, in ARCH models:

$$\epsilon_t = \sqrt{h_t}\, z_t, \quad z_t \sim IID\ N(0,1)$$
$$h_t = a_0 + a_1 \epsilon_{t-1}^2$$

Let $\mathfrak{I}_t$ denote the filtration information until time t:

$$E(\epsilon_t^2 | \mathfrak{I}_{t-1}) = E(h_t z_t^2 | \mathfrak{I}_{t-1}) = h_t E(z_t^2 | \mathfrak{I}_{t-1}) = h_t E(z_t^2) = h_t$$

Thus, we can write $\epsilon_t | \mathfrak{I}_{t-1} \sim N(0, h_t)$.

- The Gaussian assumption for z_t is not critical. We can relax it and allow for heavy-tailed distributions, such as the Student's t-distribution, as is typically required in finance.
- The parameters in the conditional variance equation should satisfy $a_0 > 0$ and $a_1 > 0$.
- ARCH(1) captures the effect that a large value in ϵ_t leads to a larger variance (volatility) in the following period.

Properties of ARCH(1) Process

$\epsilon_t = \sqrt{h_t}\, z_t$ is a nonlinear function with the following properties:

1. The unconditional mean of ϵ_t is zero, since
$$E(\epsilon_t) = E\big(E(\epsilon_t|\mathfrak{I}_{t-1})\big) = E\left(E\big(\sqrt{h_t}\, z_t\big|\mathfrak{I}_{t-1}\big)\right) = \mathrm{E}\left(\sqrt{h_t}E(z_t)\right) = 0$$

2. The conditional variance of ϵ_t is
$$E(\epsilon_t^2|\mathfrak{I}_{t-1}) = E(h_t z_t^2|\mathfrak{I}_{t-1}) = h_t E(z_t^2|\mathfrak{I}_{t-1}) = h_t = a_0 + a_1\epsilon_{t-1}^2$$

3. The unconditional variance of ϵ_t is obtained as
$$Var(\epsilon_t) = E(\epsilon_t^2) = E[E(\epsilon_t^2|\mathfrak{I}_{t-1})] = E(a_0 + a_1\epsilon_{t-1}^2)$$
$$= a_0 + a_1 E(\epsilon_{t-1}^2) = \frac{a_0}{1-a_1}$$

   It implies $a_0 > 0$ and $0 < a_1 < 1$.

4. The kurtosis of ϵ_t, K_ϵ, (heavy tails)
$$K_\epsilon = \frac{E(\epsilon_t^4)}{[Var(\epsilon_t)]^2} = \frac{E(\epsilon_t^4)}{E(\epsilon_t^2)^2} = 3\,\frac{1-a_1^2}{1-3a_1^2} > 3, \qquad if(1-3a_1^2) > 0$$

$$\mathbf{E}(\epsilon_t^4) = E[E(\epsilon_t^4|\mathfrak{I}_{t-1})] = E[E(h_t^2 z_t^4|\mathfrak{I}_{t-1})] = E(z_t^4)E\left[\left(a_0 + a_1\epsilon_{t-1}^2\right)^2\right] = 3[a_0^2 + 2a_0a_1E(\epsilon_{t-1}^2) + a_1^2E(\epsilon_{t-1}^4)]$$

Properties (1) through (4) also hold for higher order ARCH models.

Example: A simulated ARCH(1) process

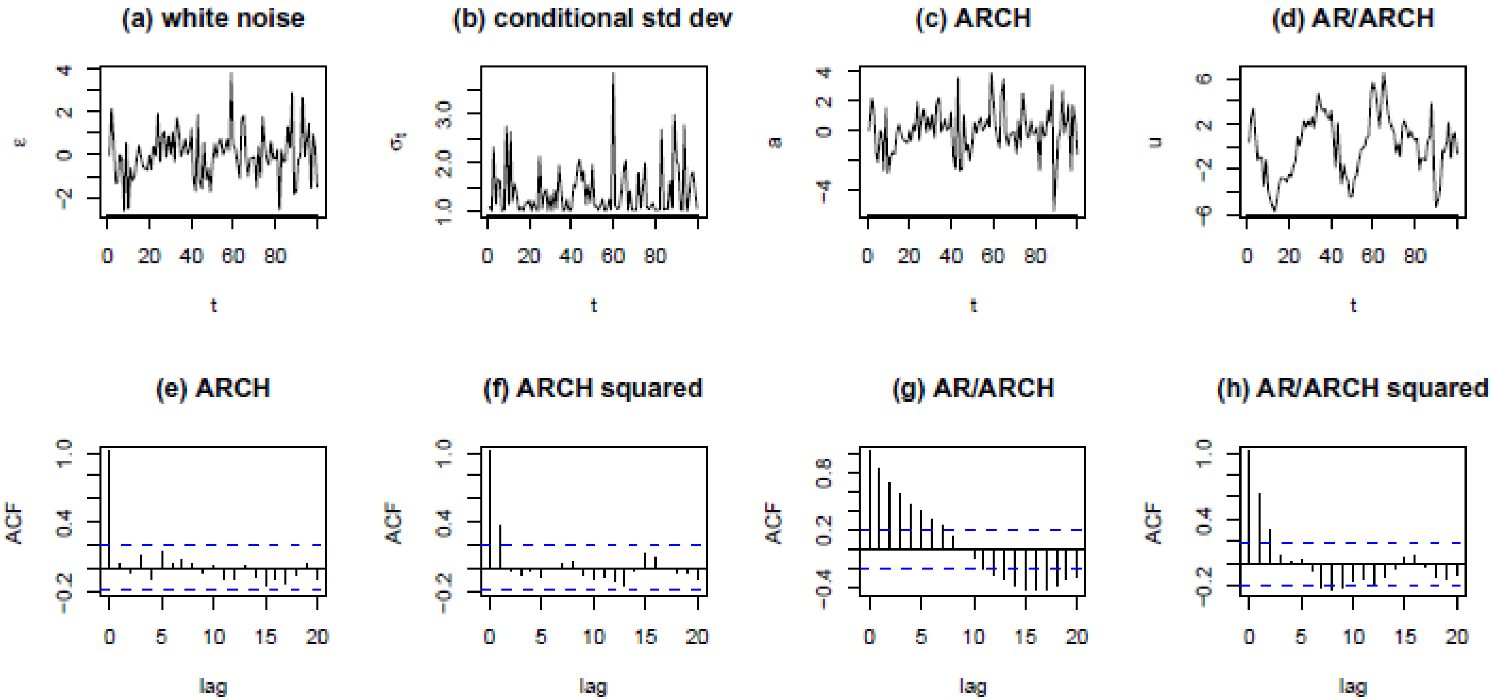


Fig. 18.2. *Simulation of 100 observations from an ARCH(1) process and an AR(1)/ARCH(1) process. The parameters are* $\omega = 1$, $\alpha_1 = 0.95$, $\mu = 0.1$, *and* $\phi = 0.8$.

(a) White noise: z_t (b) conditional std dev: $\sqrt{h_t}$ (σ_t)

(c) ARCH: $\epsilon_t = \sqrt{h_t} z_t$ (d) AR/ARCH: $y_t = f(y_{t-1}) + \epsilon_t$

R: 6_SimulatedARCH(1).R

ARCH(q) Process

Consider an *ARCH*(*q*) model for $y_t = f\left(X_{1,t}, \dots, X_{p,t}\right) + \epsilon_t$:

$$\epsilon_t = \sqrt{h_t}\, z_t, \qquad z_t \sim^{IID} N(0,1)$$

$$h_t = a_0 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2$$

The variable h_t is **the *conditional variance*** of ϵ_t .

Let $\mathfrak{I}_t$ denote the filtration information until time t. Then

$$E(\epsilon_t^2 | \mathfrak{I}_{t-1}) = E(h_t z_t^2 | \mathfrak{I}_{t-1}) = h_t E(z_t^2 | \mathfrak{I}_{t-1}) = h_t$$

Thus, we can write $\epsilon_t | \mathfrak{I}_{t-1} \sim N(0, h_t)$.

- The parameters in the conditional variance equation should satisfy $a_0 > 0$ and $a_i > 0$ for $i = 1, 2, \dots, q$.
- The Gaussian assumption for z_t is not critical. We can relax it and allow for more heavy-tailed distributions, such as the Student's t-distribution, as is typically required in finance.

Properties of ARCH(q) Process

The unconditional variance is $var(\epsilon_t) = \frac{a_0}{1-a_1-\cdots-a_q}$. Necessary and sufficient condition for weak stationarity of ARCH(q) process is $0 < \sum_j a_j < 1$.

ARCH models possess some drawbacks in practical applications:

- It might be difficult to determine the order of *q*, the number of lags of the squared residuals in the model.
- Due to the structure of the model, only the recent squared ϵ_{t-i} affect the current volatility, $\sqrt{h_t}$, *i.e.* the impact of a large shock lasts only for *q* periods. This may be unrealistic.
- Volatility clustering is ignored.

GARCH(1,1) Model

The *GARCH*(1,1) model is defined as

$$\epsilon_t = \sqrt{h_t} z_t$$

$$h_t = a_0 + a_1 \epsilon_{t-1}^2 + b_1 h_{t-1}$$

where $a_0 \geq 0, b_1 \geq 0, a_1 \geq 0.$

GARCH model allows the conditional variance to be modelled by past values of itself in addition to the past shocks.

We obtain for $t \geq 2$

$$h_t = h_0 \prod_{i=1}^{t} \left(b_1 + a_1 \epsilon_{t-i}^2 \right) + a_0 \left[1 + \sum_{k=1}^{t-1} \prod_{i=1}^{k} \left(b_1 + a_1 \epsilon_{t-i}^{\,2} \right) \right]$$

GARCH(1,1) is equivalent to a higher order ARCH process. To define the model, we need the definition of the probability measure for the starting value h_0 or the assumption that the system extends infinitely far into the past.

- The start of the system at time 0 requires that $h_0 > 0$ and finite with probability one, and h_0 and ϵ_0 are independent. The obtained model $\{h_t, \epsilon_t\}_{t=0}$ is the *conditional* model.

Properties of the GARCH(1,1) Model

ϵ_t is a nonlinear function with the following properties:

The conditional variance of ϵ_t is

$$E(\epsilon_t^2|\Im_{t-1}) = E(h_t z_t^2|\Im_{t-1}) = h_t E(z_t^2|\Im_{t-1}) = h_t = a_0 + a_1\epsilon_{t-1}^2 + b_1 h_{t-1}$$

The unconditional variance of ϵ_t is

$$Var(\epsilon_t) = E(\epsilon_t^2) = E[E(\epsilon_t^2|\Im_{t-1})] = E(a_0 + a_1\epsilon_{t-1}^2 + b_1\, h_{t-1})$$

$$= a_0 + a_1 E(\epsilon_{t-1}^2) + b_1 E(h_{t-1}) = a_0 + a_1 E(\epsilon_{t-1}^2) + b_1 E(\epsilon_{t-1}^2) = \frac{a_0}{1 - a_1 - b_1}$$

Weak stationarity: $a_0 > 0$ and $(a_1 + b_1) < 1$.

Volatility clustering.

Heavy tails: if $1 - 2a_1^2 - (a_1 + b_1)^2 > 0$, then the kurtosis of ϵ_t, K_ϵ,

$$K_\epsilon = \frac{E(\epsilon_t^4)}{[Var(\epsilon_t)]^2} = \frac{E(\epsilon_t^4)}{E(\epsilon_t^2)^2} = 3\,\frac{1 - (a_1 + b_1)^2}{1 - 2a_1^2 - (a_1 + b_1)^2} > 3$$

.

GARCH(p,q) Process

We consider the *GARCH*(p, q) process for the time series ϵ_t

$$\epsilon_t = \sqrt{h_t}\, z_t, z_t \sim^{IID} N(0,1)$$

$$h_t = a_0 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} b_j h_{t-j}$$

We assume that the parameters $(a_0, a_1, \dots, a_q, b_1, \dots, b_p)$ are restricted such that $h_t > 0$ for all t, which is ensured when $a_0 > 0$, $a_i \geq 0$ for $i = 1, 2, \dots, q$, and $b_j \geq 0$ for $j = 1, 2, \dots, p$. We also assume that the fourth moment of ϵ_t exists.

Weak stationarity: $a_0 > 0$ and $\sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j < 1$.

GARCH(1,1) and ARMA(1,1)

Let us consider the simple $GARCH(1,1)$ model:

$$\epsilon_t = \sqrt{h_t}\, z_t, z_t \sim N(0,1)$$

$$h_t = a_0 + a_1\epsilon_{t-1}^2 + b_1 h_{t-1}$$

The GARCH process ϵ_t^2 has the *ARMA*(1,1) representation

$$\epsilon_t^2 = a_0 + (a_1 + b_1)\epsilon_{t-1}^2 + u_t - b_1 u_{t-1},$$

where $u_t = h_t(z_t^2 - 1)$ is white noise*:*

- $E(u_t) = E(h_t)E(z_t^2 - 1) = 0$
- $var(u_t) = E(u_t^2) = E(h_t^2)E[(z_t^2 - 1)^2] = E[z_t^4 - 2z_t^2 + 1]E(a_0^2 + a_1^2\epsilon_{t-1}^4 + 2a_0a_1\epsilon_{t-1}^2) = const$
- $cov(u_t, u_{t+s}) = E(h_t(z_t^2 - 1)h_{t+s}(z_{t+s}^2 - 1))$
  $= E(h_t(z_t^2 - 1)h_{t+s})\, E(z_{t+s}^2 - 1) = 0$

Plug in the GARCH process, we have

$$\begin{aligned}\epsilon_t^2 &= h_t\, z_t^2 = u_t + h_t \\ &= a_0 + a_1\epsilon_{t-1}^2 + b_1 h_{t-1} + u_t \\ &= a_0 + (a_1 + b_1)\epsilon_{t-1}^2 + b_1(h_{t-1} - \epsilon_{t-1}^2) + u_t \\ &= a_0 + (a_1 + b_1)\epsilon_{t-1}^2 + b_1(h_{t-1} - h_{t-1}z_{t-1}^2) + u_t\end{aligned}$$

GARCH process and ARMA representation

Similarly, rewriting the equation in GARCH(p,q) for h_t , we obtain an *ARMA*(r, p) representation for ϵ_t^2:

$$\epsilon_t^2 = a_0 + \sum_{i=1}^{r} (a_i + b_i)\epsilon_{t-i}^2 + u_t - \sum_{j=1}^{p} b_j u_{t-j}$$

where $r = \max(p, q)$. $\{u_t\}$ is a martingale difference series. In compact representation, we can write the ARMA representation as

$$\Phi(L)\epsilon_t^2 = a_0 + b(L)u_t$$

where L is the lag operator, $\Phi(L) = 1 - \sum_{i=0}^{r} \Phi_i L^i$. $\Phi_i = a_i + b_i$, $r = \max(p, q)$, $b(L) = 1 - \sum_{i=1}^{p} b_i L^i$.

Sources of excess kurtosis

Sources of excess kurtosis:

Bai, Russel, and Tiao consider the *ARMA*(r, q) representation of the *GARCH*(p, q) process and analyse the relationship between

1. the excess kurtosis of ϵ_{t}, called the ***overall kurtosis*** and denoted by K_{ϵ} if it exists
2. the excess kurtosis of z_t, called the ***IID kurtosis*** and denoted by K_z,

and

3. the excess kurtosis of the normal GARCH process, called the ***GARCH kurtosis*** and denoted by $K_{\epsilon}^{(g)}$ if it exists.

Sources of excess kurtosis

Their results show that if ϵ_t follows the *GARCH*(p, q) process and satisfies the additional two assumptions (that u_t's are uncorrelated with zero mean and finite variance and that $\{\epsilon_t^2\}$ process is weakly stationary), the following holds:

$$K_\epsilon^{(g)} = \frac{6k_1}{1-2k_1}, where\ k_1 = \sum_i \Psi_i^2\ \ and\ \Psi(L) = b(L)/\Phi(L)$$

$$K_\epsilon = \frac{K_\epsilon^{(g)} + K_z + (5/6)K_\epsilon^{(g)}K_z}{1-(1/6)K_\epsilon^{(g)}K_z}$$

For a normal *GARCH*(1,1) model, the first expression reduces to

$$K_\epsilon = K_\epsilon^{(g)} = \frac{6a_1^2}{1-(a_1+b_1)^2 - 2a_1^2}$$

- The normal (Gaussian) *GARCH*(1,1) model is not capable of matching the large leptokurtosis typically found in the data.
- A non-normal *GARCH* model fits well the time-varying volatility relation and matches the sample kurtosis much better.

ARIMA(p_A, d, q_A)/GARCH(p_G, q_G) Models

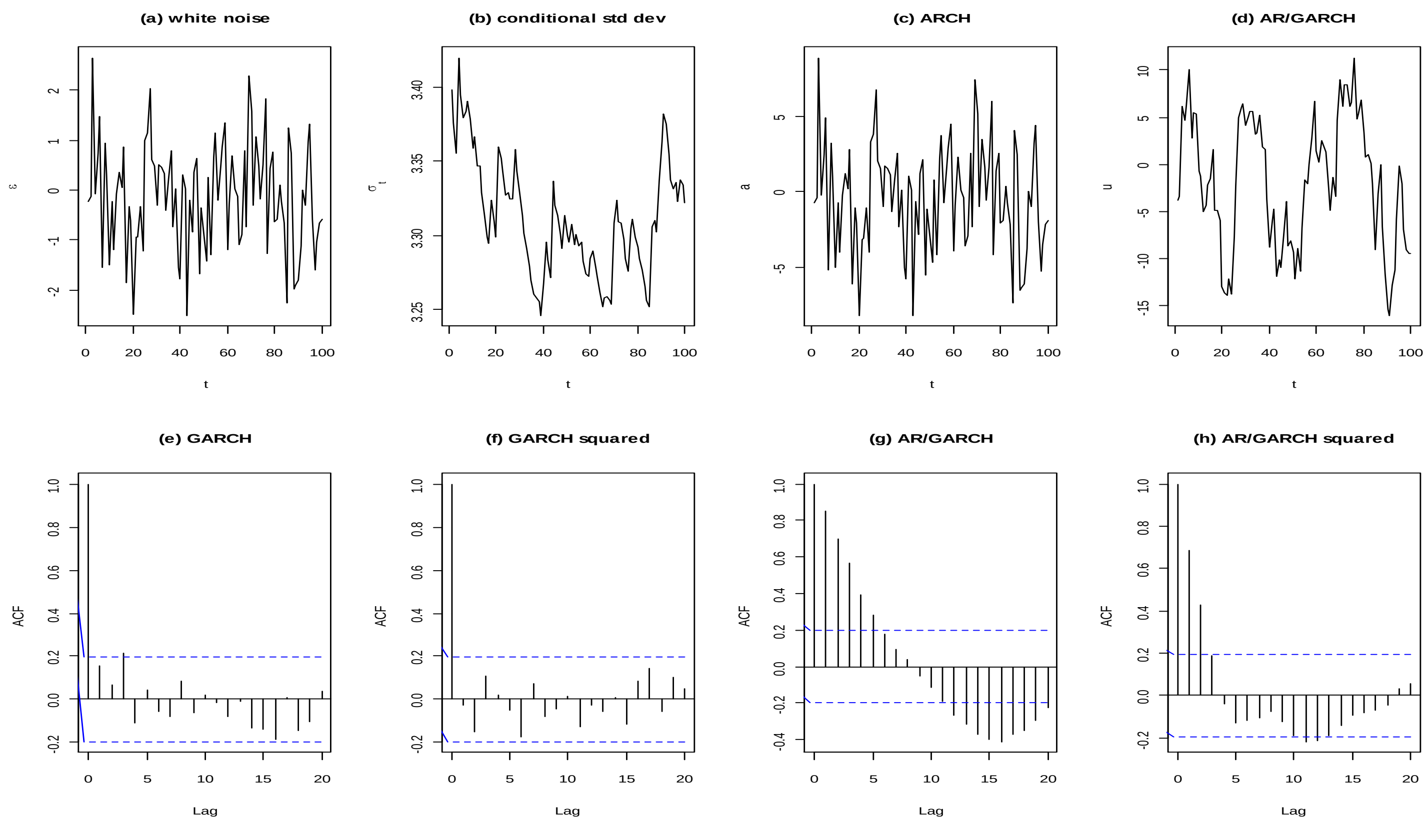


Fig. 18.3. *Simulation of GARCH(1,1) and AR(1)/GARCH(1,1) processes. The parameters are* $\omega = 1$, $\alpha_1 = 0.08$, $\beta_1 = 0.9$, *and* $\phi = 0.8$.

R: 6_SimulatedGARCH(1,1).R

(a) White noise: z_t (b) conditional std dev: $\sqrt{h_t}$ (σ_t)

(c) GARCH: $\epsilon_t = \sqrt{h_t} z_t$ (d) AR/GARCH: $y_t = f(y_{t-1}) + \epsilon_t$

R: 6_SimulatedARCH(1).R

Estimation of the GARCH Models

The conditional variance is an unobserved variable, which must itself be explicitly estimated along with the parameters of the model.

Engle suggested two possible methods for estimating the parameters in model, namely the ***least squares estimator* (LSE)** and **the *maximum likelihood estimator* (MLE)**. The LSE is given as

$$\hat{\boldsymbol{\theta}} = \left(\sum_{t=2}^{T} \tilde{\boldsymbol{\epsilon}}_{t-1}\tilde{\boldsymbol{\epsilon}}'_{t-1}\right)^{-1} \sum_{t=2}^{T} \tilde{\boldsymbol{\epsilon}}_{t-1}\tilde{\boldsymbol{\epsilon}}_{t}^{2}$$

where $\boldsymbol{\theta} = (a_0, a_1, \dots, a_p)'$ and $\tilde{\boldsymbol{\epsilon}}_{t} = (1, \epsilon_t^2, \dots, \epsilon_{t-p+1}^2)$.

The conditional likelihood, l_t , of ϵ_t is

$$l_t = \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{1}{2h_t}\epsilon_t^2\right)$$

By iterating the conditional argument, we obtain

$$\begin{aligned} f(\epsilon_T, \dots, \epsilon_1 | \epsilon_0) &= f(\epsilon_T | \epsilon_{T-1}, \dots, \epsilon_0) \dots f(\epsilon_2 | \epsilon_1, \epsilon_0) f(\epsilon_1 | \epsilon_0) \\ &= \prod_{t=1}^{T} f(\epsilon_t | \Im_{t-1}) \end{aligned}$$

Maximum Likelihood Estimation

The joint likelihood of the entire sample of *T* observations is

$$L = \prod_{t=1}^{T} l_t$$

and for the log likelihood we obtain

$$\log f(\epsilon_t, \dots, \epsilon_1 | \epsilon_0) = \sum_{t=1}^{T} \log f(\epsilon_t | \Im_{t-1})$$

$$= -\frac{T}{2} \log 2\pi + \sum_{t=1}^{T} -\frac{1}{2} \log h_t - \frac{1}{2} \sum_{t=1}^{T} \frac{\epsilon_t^2}{h_t}$$

The conditional log-likelihood function can thus be written as

$$L(\theta) = -\frac{1}{2} \sum_{t=1}^{T} \left(\log 2\pi + \log h_t(\theta) + \frac{\epsilon_t^2(\theta)}{h_t(\theta)} \right)$$

where $\theta = (a_0, a_1, \dots, a_q, b_1, \dots, b_p)$. The value of θ which maximizes $L(\theta)$ is referred to as maximum likelihood estimates or MLEs.

Newton-Raphson

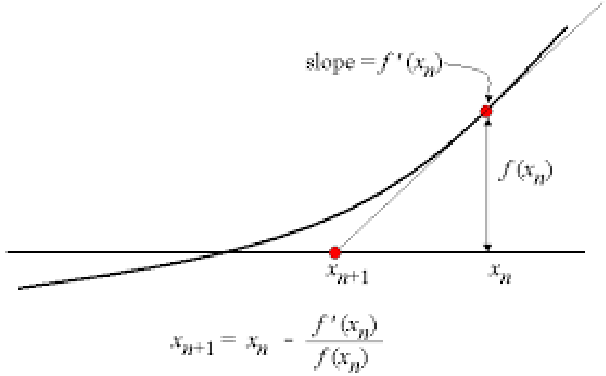
The MLE of a GARCH model is obtained by numerically maximizing the log-likelihood function using iterative optimization methods like Gauss-Newton or Newton-Raphson.

Starting values for the parameters in θ and initialization of the two series ϵ_t^2 and h_t need to be specified for iterative ML optimization.

The usual default solution to parameter initializations is to set parameter values in the mean equal to those estimated using a “first pass" LS estimation, and the parameters (except a_0) in the conditional-variance equation to zero.

Differences in parameter initialization of the packages may yield different results.

When the innovation process is not normal or the conditional distribution is not perfectly known, one may still use Gaussian MLE methods due to the property of asymptotic parameter efficiency. Such estimates are known as **pseudo-** or **quasi-MLE** (**PMLE** or **QMLE**).



Example: AR(1)/GARCH(1,1) fit to BMW returns

Consider the BMW daily log returns. An AR(1)/GARCH(1,1) model was fit to these returns using R's garchFit function in the fGarch package. Although garchFit allows the white noise to have a non-Gaussian distribution, in this example we specified Gaussian white noise (the default).

```
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)
 [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
```

The fitted model is $(r_t - 4.0092e-04) = 0.098596(r_{t-1} - 4.0092e-04) + \epsilon_t$

$$\epsilon_t = \sqrt{h_t} z_t, \qquad z_t \sim N(0,1)$$

$$h_t = 8.9043e-06 + 0.10210\epsilon_{t-1}^2 + 0.85944 h_{t-1}$$

Data: 6_bmw.txt
R: 6_BMWReturnsFit.R

Example: AR(1)/GARCH(1,1) fit to BMW returns

```
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    normalized:  2.89

Information Criterion Statistics:
  AIC   BIC   SIC  HQIC
-5.78 -5.77 -5.78 -5.77
```

In the output, ϕ is denoted by $\mathrm{ar1}$. The mean is mu, and omega ω is the intercept of the GARCH process.

Note that $\widehat{\phi} = 0.0986$ and is statistically significant, implying that there is a small amount of positive autocorrelation. Both α_1 and β_1 are highly significant and $\hat{\beta}_1 = 0.859$, which implies rather persistent volatility clustering.

There are two additional information criteria reported, SIC (Schwarz's information criterion) and HQIC (Hannan-Quinn information criterion). These are less widely used compared to AIC and BIC and will not be discussed here.

Diagnostic checking

We need to check the adequacy of the fitted model.

- The part of the data unexplained by the model (i.e., the residuals $\hat{\epsilon}_t$ and studendized residuals $\hat{z}_t = \hat{\epsilon}_t / \sqrt{\hat{h}_t}$) should be small and not exhibit any systematic or predictable patterns.

Ljung and Box test for residuals and studendized residuals

$$Q = T(T+2)\sum_{k=1}^{m}(T-k)^{-1}\hat{\rho}_{\epsilon,k}^2$$

and McLeod and Li test the joint hypothesis on the ACF of the squared residuals and studendized residuals:

$$H_0 : \rho_{\epsilon^2,1} = \rho_{\epsilon^2,2} = \cdots = \rho_{\epsilon^2,m} = 0$$

by performing a *Q* test on the squared residuals:

$$Q_2 = T(T+2)\sum_{k=1}^{m}(T-k)^{-1}\hat{\rho}_{\epsilon^2,k}^2$$

Under the null hypothesis of no autocorrelation, Q_2 has a χ^2 distribution with $m - p - q$ degrees of freedom.

Example: BMW returns

The Ljung-Box tests with an R in the second column are applied to the residuals (here R = residuals, not the R software), while the Ljung-Box tests with R^2 are applied to the squared residuals.

```
        Log Likelihood:
         18159    normalized:  2.9547

        Standardised Residuals Tests:
                                         Statistic p-Value

 Jarque-Bera Test   R    Chi^2  13355     0
 Shapiro-Wilk Test  R    W      NA        NA
 Ljung-Box Test     R    Q(10)  21.933    0.015452
 Ljung-Box Test     R    Q(15)  26.501    0.033077
 Ljung-Box Test     R    Q(20)  36.79     0.012400
 Ljung-Box Test     R^2  Q(10)  5.8285    0.82946
 Ljung-Box Test     R^2  Q(15)  8.0907    0.9201
 Ljung-Box Test     R^2  Q(20)  10.733    0.95285
 LM Arch Test       R    TR^2   7.009     0.85701

Information Criterion Statistics:
    AIC     BIC     SIC    HQIC
-5.9071 -5.8994 -5.9071 -5.9044
```

Stationarity of ARMA-GARCH models

We have to capture the conditional mean of the data with an adequate model so that the residuals obtained from this model satisfy the assumptions for the white-noise sequence $\{\epsilon_t\}$ which enters the conditional variance.

For asset returns, the conditional mean is typically captured by an AR or ARMA model:

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{q} b_i \epsilon_{t-i} + \epsilon_t$$

$$\epsilon_t = \sqrt{h_t}\, z_t, h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_j h_{t-i}$$

If all the roots of AR polynomial lie outside the unit circle, the ARMA-GARCH process y_t is strictly stationary if ϵ_t is strictly stationary.

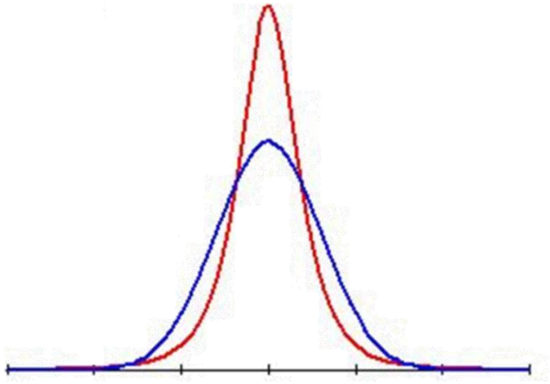
The parameters of an ARMA-GARCH model can be jointly estimated via MLE. Alternatively, a two-step procedure can be adopted:

1. Estimate the parameters of the conditional mean in the first equation
2. From residuals of the second equation, estimate the parameters of the GARCH model

Variants of the GARCH Model

Empirically relevant variants of the GARCH models are able to accommodate particular features of asset return series.

- ❑ Conditional non-normality of the error process can better explain the **leptokurtosis** of the return series. For example, daily percentage change in exchange rates based on 12 different rates for a 10-year period. The "true" distribution has heavier tails than the normal distribution: Extreme events are more common than predicted by the normal distribution.



No. of standard deviations	Real world (%)	Normal model (%)
>1 SD	25.04	31.73
>2 SD	5.27	4.55
>3 SD	1.34	0.27
>4 SD	0.29	0.01
>5 SD	0.08	0.00
>6 SD	0.03	0.00

- ❑ Asymmetric responses to negative and positive return innovations to model the **asymmetry** in the reaction of conditional volatility to the arrival of different news.
- ❑ **Long-memory**, i.e., variances generated by fractionally integrated processes.

GARCH Model with Student's *t*-distributed Innovations

The shortcoming of the GARCH model with Gaussian innovations is that the assumption of conditional normality for $\{\epsilon_t\}$ usually does not hold.

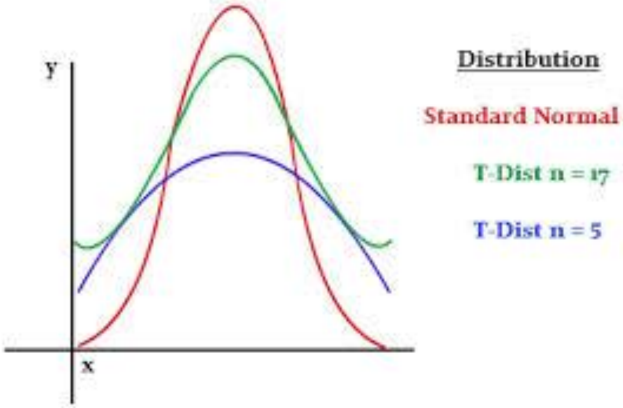
The error term or residual ϵ_t is conditionally normal if the ***standardized residual***

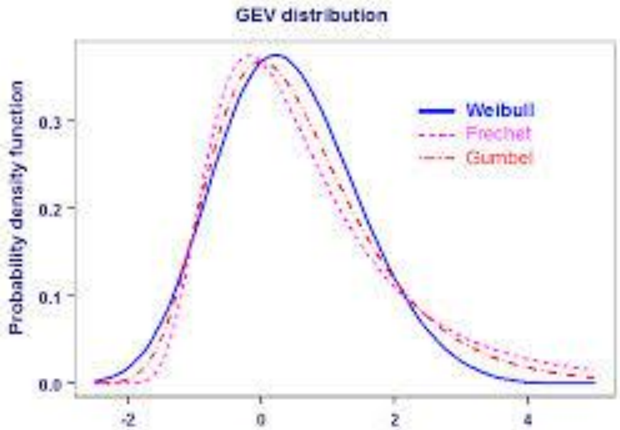
$$\hat{z}_t = \frac{\hat{\epsilon}_t}{\sqrt{h_t}}$$

is normally distributed.

The GARCH model is only able to capture partially the leptokurtosis in the unconditional distribution of asset returns.

We can estimate a GARCH model with the Student's *t*-distribution or generalized exponential distribution (GED).





GARCH Model with Student's *t*-distributed Innovations

The Student's *t*-distribution or the standardized $t(d)$ distribution has only one parameter, d, and its density is

$$f_{t(d)}(x;d) = \frac{\Gamma((d+1)/2)}{\Gamma(d/2)\sqrt{\pi(d-2)}}\left(\frac{1+x^2}{d-2}\right)^{-(1+d)/2}$$

It is a power function of the random variable x (allows the standardized $t(d)$ distribution to have fatter tails).

This distribution is symmetric around zero. The mean, variance, skewness, and excess kurtosis are $0, 1, 0$, and $6/(d-4)$, respectively.

When using the assumption that $\epsilon_t \sim t(d)$ in the GARCH model, the estimation can be done by quasi maximum likelihood estimation.

Example: ARMA(1)/GARCH(1,1) fit to BMW returns with Student-t distribution

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

Mean and Variance Equation:
 data ~ arma(1, 1) + garch(1, 1) [data = bmw]

Conditional Distribution: std

Coefficient(s):
         mu          ar1          ma1        omega       alpha1
 1.7358e-04  -2.9869e-01   3.6896e-01   6.0525e-06   9.2924e-02
      beta1        shape
   8.8688e-01 4.0461e+00

Std. Errors: based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      1.736e-04   1.855e-04    0.936  0.34929
ar1    -2.987e-01   1.370e-01   -2.180  0.02924 *
ma1     3.690e-01   1.345e-01    2.743  0.00608 **
omega   6.052e-06   1.344e-06    4.502 6.72e-06 ***
alpha1  9.292e-02   1.312e-02    7.080 1.44e-12 ***
beta1   8.869e-01   1.542e-02   57.529  < 2e-16 ***
shape   4.046e+00   2.315e-01   17.480  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The fitted model is:

$$(r_t - 1.736e-04)$$
$$= 0.2987(r_{t-1} - 1.736e-04) + \epsilon_t + 0.3690\epsilon_{t-1}$$
$$\epsilon_t = \sqrt{h_t} z_t, \qquad z_t \sim t(4.046)$$
$$h_t = 6.052e-06 + 0.09292\epsilon_{t-1}^2 + 0.8869 h_{t-1}$$

Example: AR(1)/GARCH(1,1) fit to BMW returns

The Jarque-Bera test of normality strongly rejects the null hypothesis that the innovation process is Gaussian.

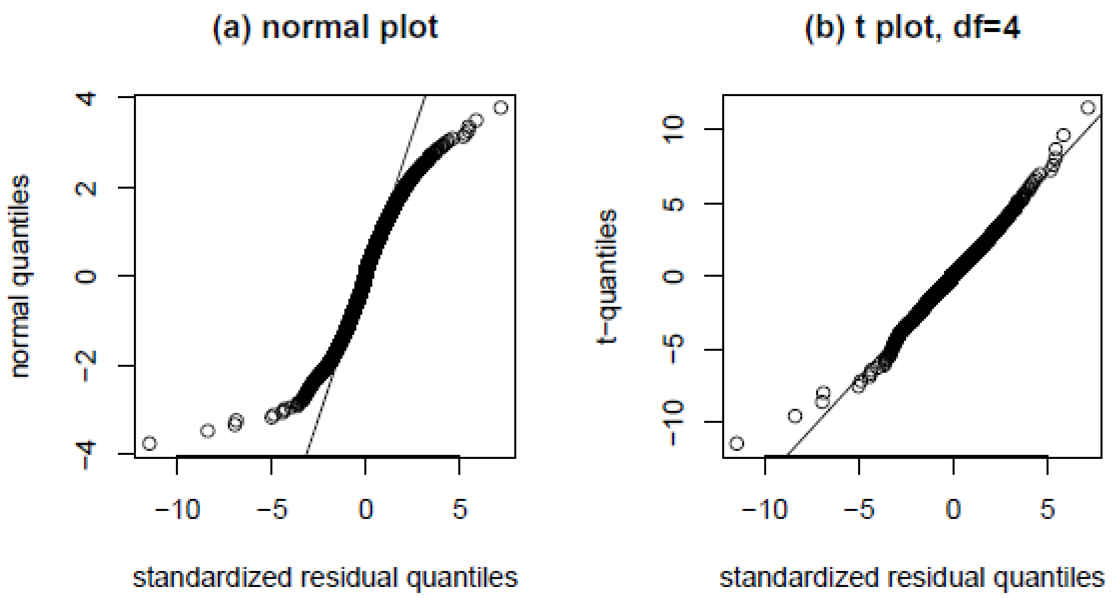


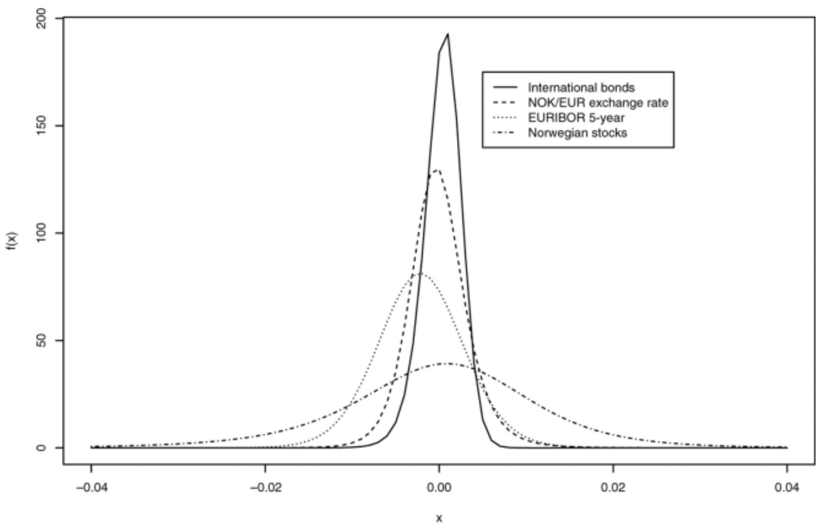
Fig. 18.4. *QQ plots of standardized residuals from an AR(1)/GARCH(1,1) fit to daily BMW log returns. The reference lines go through the first and third quartiles.*

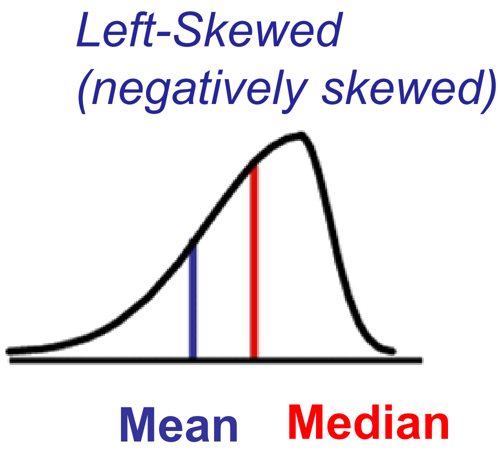
Data: 6_bmw.txt
R: 6_BMWReturnsFit.R

Asymmetry in returns (skewness)

Traders react more strongly to negative information than to positive information which leads to asymmetry in rates of return.

Rates of return are slightly negatively skewed.





To accommodate asymmetric distributions, we may use the generalized hyperbolic distribution that is able to accommodate both fat tails and asymmetry.

Leverage effect: Exponential GARCH model

Since future return volatility tends to respond asymmetrically with respect to negative or positive shocks, the symmetric GARCH model is not appropriate.

To overcome this limitation, a nonlinear ***exponential GARCH* (EGARCH)** specification was proposed by Nelson.

The conditional variance h_t is specified as

$$\log(h_t) = a_0 + \sum_{i=1}^{q} a_i g(z_{t-i}) + \sum_{i=1}^{p} b_i \log(h_{t-i})$$

where $\epsilon_t = \sqrt{h_t}\, z_t$ and $g(z_t) = \theta z_t + \gamma[|z_t| - E|z_t|]$ are the weighted innovations that model asymmetric effects between positive and negative asset returns.

The function $g(.)$ can be rewritten as

$$g(z_t) = \begin{cases} (\theta + \gamma) z_t - \gamma E|z_t| \; if \; z_t \geq 0 \\ (\theta - \gamma) z_t - \gamma E|z_t| \; if \; z_t < 0 \end{cases}$$

If $\gamma < 0$ a positive return shock or surprise will increase volatility less than a negative one of the same magnitude. This phenomenon is referred to as the ***leverage effect***.

Moreover, EGARCH relaxes the constraint of positive GARCH coefficients and guarantees the estimated volatility to be non-negative.

Exponential GARCH Model

For a standard Gaussian random variable z_t, we have $E(|z_t|) = \sqrt{2/\pi}$. For the standardized Student's t-distribution we have

$$E(|z_t|) = \frac{\sqrt{d}\Gamma[0.5(d-1)]}{\sqrt{d}\Gamma[0.5d]}$$

where d is the number of degrees of freedom.

In summary, the EGARCH model has two advantages over the symmetric GARCH specification:

- Function g enables the model to respond asymmetrically to positive and negative lagged values of ϵ_t.
- Use of the log-conditional variance in EGARCH specification relaxes the constraint of positive model coefficients.

APARCH Models

In some financial time series, large negative returns appear to increase volatility more than do positive returns of the same magnitude. This is called the ***leverage effect***.

The APARCH(*p,q*) model for the conditional standard deviation is

$$h_t^{\delta} = \omega + \sum_{i=1}^{p} \alpha_i (|\epsilon_{t-1}| - \gamma_i \epsilon_{t-1})^{\delta} + \sum_{j=1}^{q} \beta_j h_{t-j}^{\delta}$$

The leverage effect is through the function g_{γ_i}, where $g_{\gamma}(x) = |x| - \gamma x$. When $\gamma > 0,\ g_{\gamma}(-x) > g_{\gamma}(x)$ for any $x > 0$, so there is a leverage effect. If $\gamma < 0$, then there is a leverage effect in the opposite direction to what is expected --- positive past values of ϵ_t increase volatility more than negative past values of the same magnitude.

Gamma function

Gamma>0, larger weights are given to negative than positive.
The larger the magnitude of gamma, the model is more asymmetric.

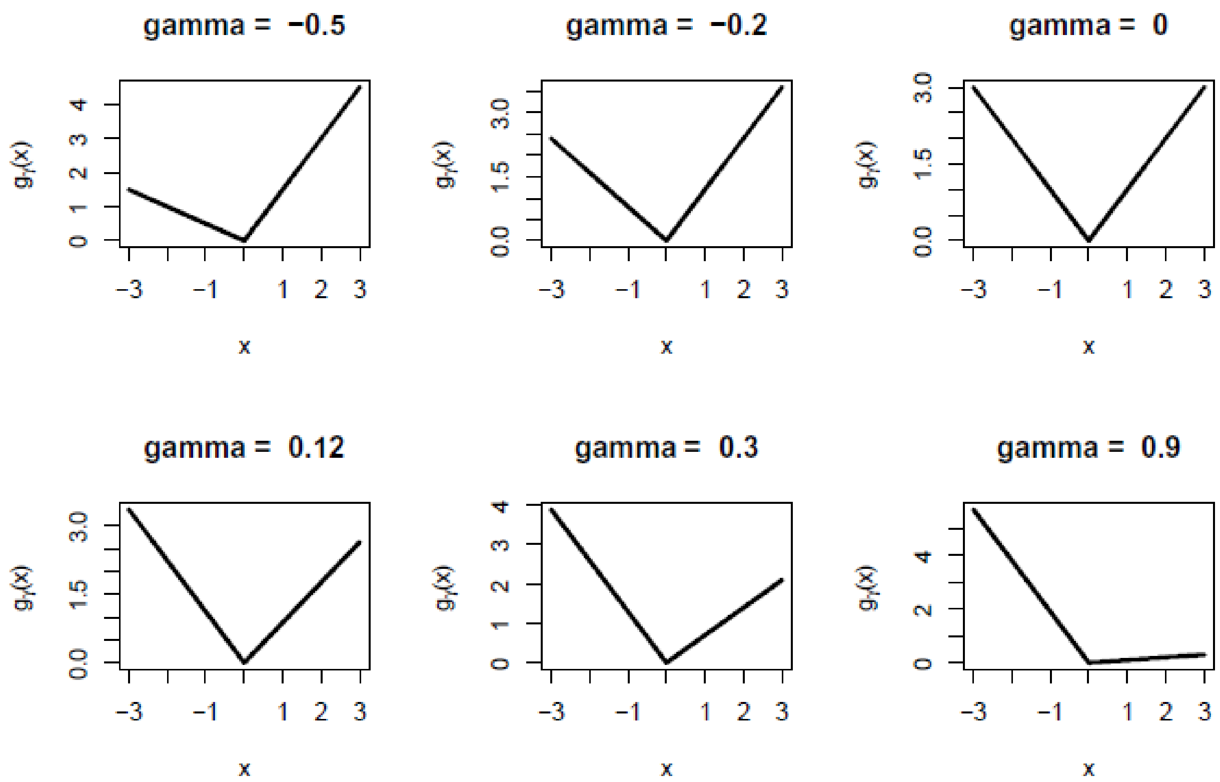


Fig. 18.7. *Plots of* $g_\gamma(x)$ *for various values of* γ.

Integrated GARCH Model

The estimation of ARCH processes on log-return data yields the similar pattern in the results:

For longer samples, the estimated parameters $a_1, \dots, a_q$ and $b_1, \dots, b_p$ of the model sum up to values that are typically close to one.

For shorter samples, the sum of the estimated coefficients, although not small, stays away from 1.

These two observed facts are known as the ***integrated GARCH (IGARCH) effect***.

Engle and Bollerslev introduced the integrated *GARCH*(*p, q*) (*IGARCH*(*p, q*)) process for which

$$a_0 > 0 \text{ and } \sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j = 1$$

The *IGARCH* model has a strictly stationary solution (h_t), and therefore $\{\epsilon_t\}$ is strictly stationary as well, but ϵ_t's do not have a finite second moment:

$$E(h) = a_0 + \sum_{i=1}^{q} a_i E(\epsilon^2) + \sum_{i=1}^{p} b_i E(h) = a + E(h)$$

Integrated GARCH Model

We can also rewrite the *GARCH*(1,1) model as

$$(1 - a_1 L - b_1 L)\epsilon^2 = a_0 + (1 - b_1 L)\,(\epsilon_t^2 - h_t)$$

If the polynomial $1 - a_1 L - b_1 L$ contains a unit root, we obtain the IGARCH model of Engle and Bollerslev.

For the GARCH case, integrated is not synonymous for nonstationarity.

The stationary *GARCH*(1,1) process is not persistent in variance if and only if $a_1 + b_1 < 1$. The stationary *IGARCH*(1,1) model with $a_1 + b_1 = 1$ is persistent in variance so that shocks to conditional variance never die out.

Forecasting with GARCH Models

GARCH models describe evolution of the conditional variance of ϵ_t, which can be linked with the evolution of y_t:

$$Var(y_t|y_{t-1}, y_{t-2}, \dots) = Var(\epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, \dots)$$

Consider *GARCH*(1,1) model and let t be the forecast origin. The 1-step ahead forecast is

$$h_t(1) = h_{t+1} = a_0 + a_1\epsilon_t^2 + b_1 h_t$$

and the 2-step ahead forecast is

$$h_t(2) = h_{t+2} = a_0 + a_1\epsilon_{t+1}^2 + b_1 h_{t+1}$$
$$= a_0 + a_1(a_0 + a_1\epsilon_t^2 + b_1 h_t) + b_1(a_0 + a_1\epsilon_t^2 + b_1 h_t)$$
$$= a_0(1 + (a_1 + b_1)) + (a_1 + b_1)(a_1\epsilon_t^2 + b_1 h_t)$$

This can be extended to general recursion form for the n-step ahead forecast

$$h_t(n) = a_0 + (a_1 + b_1)h_t(n-1), n > 1$$

Forecasting with GARCH Models

For the *GARCH*(1,1) model, the *n*−step ahead forecast can be written as

$$h_t(n) = h_{t+n} = \hat{a}_0\left(1 + \sum_{i=1}^{n-1}\left(\hat{a}_1 + \hat{b}_1\right)^i\right) + \left(\hat{a}_1 + \hat{b}_1\right)^{n-1}(\hat{a}_1\epsilon_t^2 + \hat{b}_1 h_t)$$

for any $n \geq 2$ where the quantities on the right−hand side are known.

For the *GARCH*(1,1) model, given that $(\hat{a}_1 + \hat{b}_1) < 1$, the *n*-step ahead forecast is

$$h_{t+n} = \hat{a}_0 \frac{1 - \left(\hat{a}_1 + \hat{b}_1\right)^{n-1}}{1 - \left(\hat{a}_1 + \hat{b}_1\right)} + \left(\hat{a}_1 + \hat{b}_1\right)^{n-1}\hat{h}_{t+1}, \qquad n \geq 2$$

As the forecast horizon grows, the long-term prediction will tend towards the unconditional volatility:

$$\hat{h} \longrightarrow \frac{\hat{a}_0}{1 - \left(\hat{a}_1 + \hat{b}_1\right)}, \qquad as\ n \longrightarrow \infty$$

Forecast evaluation

To measure forecast accuracy, we use usual summary statistics based directly on the deviation between forecasts and realizations such as the ***root mean squared error (RMSE)***, the ***mean absolute error (MAE)*** and the ***mean absolute percentage error (MAPE)***:

$$RMSE = \left[\frac{1}{n}\sum_{\tau=t+1}^{t+n}\left(\sqrt{\hat{h}_\tau} - \sqrt{h_\tau}\right)^2\right]^{1/2}$$

$$MAE = \frac{1}{n}\sum_{\tau=t+1}^{t+n}\left|\sqrt{\hat{h}_\tau} - \sqrt{h_\tau}\right|$$

$$MAPE = \frac{1}{n}\sum_{\tau=t+1}^{t+n}\left|\frac{\sqrt{\hat{h}_\tau} - \sqrt{h_\tau}}{\sqrt{h_\tau}}\right|$$

Does Anything Beat a GARCH(1,1)?

A voluminous literature has emerged for modeling the temporal dependencies in financial market volatility at the daily and lower frequencies using ARCH and stochastic volatility type models. Most of these studies find highly significant in-sample parameter estimates and pronounced intertemporal volatility persistence. Meanwhile, when judged by standard forecast evaluation criteria, based on the squared or absolute returns over daily or longer forecast horizons, standard volatility models provide seemingly poor forecasts.

Andersen, T. G., And T. Bollerslev (1998): “Answering the skeptics: Yes, standard volatility models do provide accurate forecasts,” International Economic Review, 39(4), 885–905.

- While $E(r_t^2|\Im_{t-1}) = h_t E(z_t^2) = h_t$ appears to justify the use of the squared returns over the relevant forecast horizon as a proxy for the ex-post volatility, the squared innovation may yield very noisy measurements due to the idiosyncratic error term, z .
- In empirically realistic situations the GARCH models actually produce accurate interdaily forecasts for the latent volatility factor if realized variance is used as benchmark.

Hansen, P. Reinhard, and Asger Lunde (2001): "A comparison of volatility models: Does anything beat a GARCH (1, 1)." Unpublished manuscript. Department of Economics, Brown University.

- Performed out-of-sample comparison of 330 different volatility models using daily exchange rate data (DM/$) and IBM stock prices. Interestingly, the best models do not provide a significantly better forecast than the GARCH(1,1) model, while an ARCH(1) model is clearly outperformed.

Illustration: Estimate Monthly Vola

Let r_t^m be the t-th month log return and $\{r_{t,i}\}_{i=1}^{T}$ be the daily log returns within the t-th month:

$$r_t^m = \sum_{i=1}^{T} r_{t,i} \, .$$

We have

$$var(r_t^m | F_{t-1}) = \sum_{i=1}^{T} var\left(r_{t,i} \middle| F_{t-1}\right) + 2 \sum_{i<j} cov[(r_{t,i}, r_{t,j}) | F_{t-1}]$$

where F_{t-1} denotes the information available at month $t-1$ (inclusive).

If $r_{t,i}$ is a white noise series, then $\sigma_m^2 = var\,(r_t^m | F_{t-1}) = T var\left(r_{t,1}\right) = T\sigma^2$ and

$$\widetilde{\sigma^2} = \sum_{i=1}^{T} \left(r_{t,i} - \bar{r_t}\right)^2 / (T-1), \qquad \bar{r_t} = \sum_{i=1}^{T} r_{t,i} / T$$

The estimated monthly volatility is then

$$\widetilde{\sigma_m^2} = \frac{T}{T-1} \sum_{i=1}^{T} \left(r_{t,i} - \bar{r_t}\right)^2 \approx \sum_{i=1}^{T} \left(r_{t,i} - \bar{r_t}\right)^2$$

If the sample mean $\bar{r_t}$ is zero, then $\widetilde{\sigma_m^2} \approx \sum_{i=1}^{T} r_{t,i}^2$

Apply the idea to intra-daily log returns we have Realized Variance $RV_t = \sum_{i=1}^{T} r_{t,i}^2$ with $r_{t,i}$ denoting the i-th intra-daily log return on day t.

Realized variance and high frequency data

Tick by tick data of the IBM share traded at the New York Stock Exchange (NYSE) in the year 2005. Data source: Trades and Quotes (TAQ).

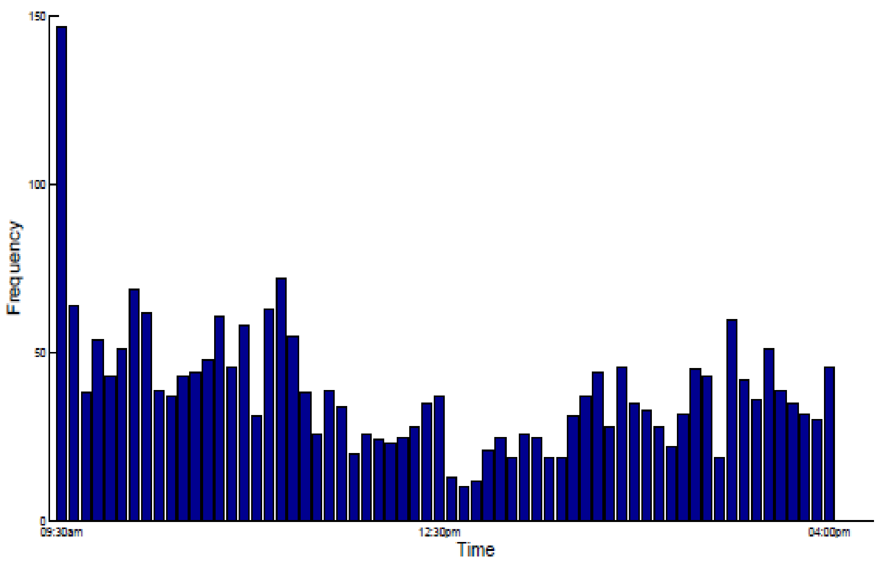
High-frequency (HF) financial data (Intradaily data, tick-by-tick data or transaction data): include transaction information recorded in markets at very high frequencies, e.g. minute, second, millisecond.

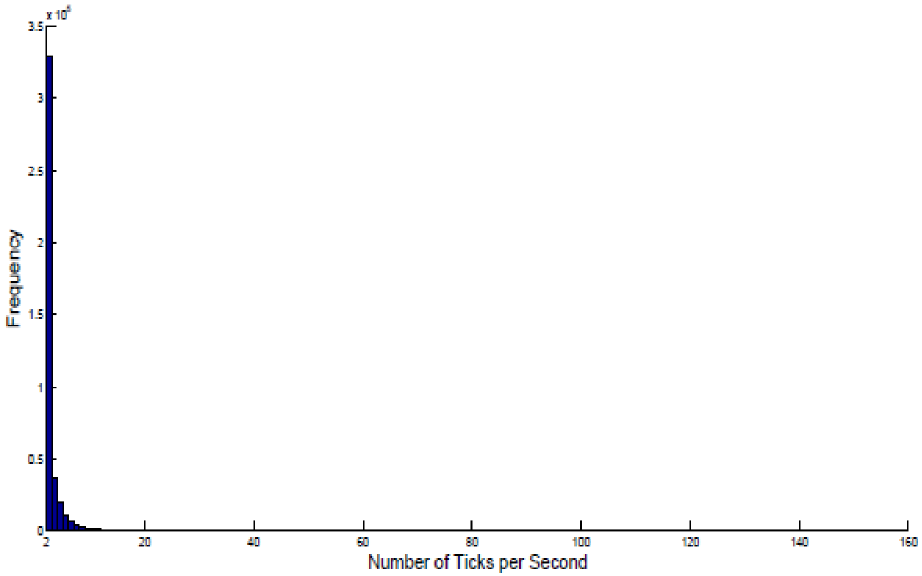
Basic features:

- ❑ Irregular time intervals
- ❑ Discrete values, e.g. price in multiples of tick size
- ❑ Large sample size
- ❑ Multi-dimensional variables, e.g. price, volume, quotes, etc.
- ❑ Daily periodic/Diurnal Pattern
- ❑ Multiple transactions within a single time unit
- ❑ Leptokurtic or heavy tails

Date	h	min	sec	price	volume
20050103	9	40	05	99.04	100
20050103	9	40	10	99.03	300
20050103	9	40	11	99.04	300
20050103	9	40	14	99.01	700
20050103	9	40	15	99.03	700
20050103	9	40	15	99.03	1000
20050103	9	40	17	99.03	100
20050103	9	40	27	99.04	200
20050103	9	40	28	99.03	1000
20050103	9	40	31	99.03	2000
20050103	9	40	35	99.02	400
20050103	9	40	36	99.03	600
20050103	9	40	37	99.02	200
20050103	9	40	40	99.02	100
20050103	9	40	40	99.02	300
20050103	9	40	56	99.02	200
20050103	9	40	57	99.02	100
20050103	9	40	59	99.02	1000

High frequency data





Left: Intraday Trading Intensity for IBM on 11th of March 1998. The height of each block represents the number of trades in 6 minutes time.

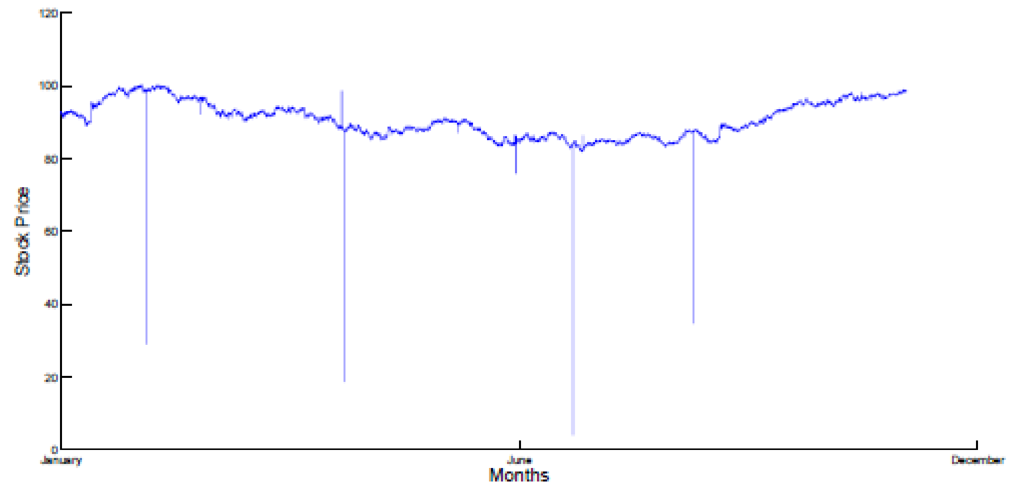
Right: Multiple transactions within one second: The histogram depicts the occurrence of seconds with 2,3,4 or more transactions per second. It is based on the data of GE in 2005. The maximum number of transaction per second is 153 transactions.

Data cleaning

Data cleaning:

- ❑ Detect errors in the raw data
- ❑ Synchronize the unequally spaced data by interpolation

(Fig: Raw HF stock price of IBM in 2004)



Market Microstructure:

- ❑ Asynchronous trading can introduce:
  - ❑ lag-1 cross-correlation between stock returns,
  - ❑ lag-1 serial correlation in a portfolio return,
  - ❑ in some situations (mean is nonzero) negative serial correlations of the return series of a single stock.
- ❑ Bid-ask bounce
- ❑ Impact of changes in tick size, after-hour trading, etc.
- ❑ Impact of daily price limits (many foreign markets)

Asynchronous Trading

Suppose log returns r_t are IID (μ, σ^2). For each time index t, the probability of no trading $P = \pi$, that is, r_t is unobserved if there is no trade. The observed log return series r_t^o turns to be:

$$r_t^o = \begin{cases} 0 & with\ prob\ \pi \\ r_t & with\ prob\ (1-\pi)^2 \\ r_t + r_{t-1} & with\ prob\ (1-\pi)^2\pi \\ \vdots & \vdots \\ \sum_{i=0}^{k} r_{t-i} & with\ prob\ (1-\pi)^2\pi^k \\ \vdots & \vdots \end{cases}$$

One can use this relation to show:

$$\mathrm{E}(r_t^o) = \mu,$$
$$var(r_t^o) = \sigma^2 + (2\pi\mu^2)(1-\pi)$$
$$cov(r_t^o, r_{t-j}^o) = -\mu^2\pi^j \qquad j \geq 1.$$

Impact of Microstructure Noise

Let $p_{t,i}$ be the observed i-th log price on day t. It is contaminated by market microstructure noise $u_{t,i}$:

$$p_{t,i} = p^*_{t,i} + u_{t,i} \qquad i = 1,2,\cdots,n$$

where $p^*_{t,i}$ is the efficient log price following the Brownian motion (continuous-time stochastic process with $u_{t,i}$ IID normal distributed).

Observed intra-day return and the naive RV are equal to:

$$r^{(T)}_{t,i} = r^{*(T)}_{t,i} + \varepsilon^{(T)}_{t,i}, \qquad \varepsilon^{(T)}_{t,i} = u_{t,i} - u_{t,i-1}$$

$$RV_t = RV^*_t + 2\sum_{i=1}^{n} r^*_{t,i}\,\varepsilon_{t,i} + \sum_{j=1}^{n} \varepsilon^2_{t,i}$$

$$E(RV_t) = RV^*_t + 2n\mathrm{E}(u^2_{t,i})$$

When n is large, the naïve estimator with the contaminated returns will be dominated by the variance of noise!

Optimal frequency: Minimizing Mean Squared Error.

Consistent estimators under presence of Microstructure Noise?

Two Scaled Realized Variance (TSRV)

Zhang, Lan, Per A. Mykland, and Yacine Aït-Sahalia. "A tale of two time scales: Determining integrated volatility with noisy high-frequency data." Journal of the American Statistical Association 100.472 (2005): 1394-1411.

- Subsampling: Allocate the intra-day observations into Q non-overlapping subsamples, each with n_q observations
- Compute RV for each subsamples: $RV_t^{(q,n_q)}$ and averaging them $\left(\frac{1}{Q}\right)\sum_{q=1}^{Q} RV_t^{(q,n_q)}$ to remove the impact of microstructure noise.

RV based on returns at lower sampling frequency

$$RV_t^{(q,n_q)} = \left(\frac{1}{n_q}\right)\sum_{i=1}^{n_q} r_{t,i}^2,$$

RV based on returns at higher sampling frequency

$$RV_t = \left(\frac{1}{n}\right)\sum_{i=1}^{n} r_{t,i}^2,$$

TSRV

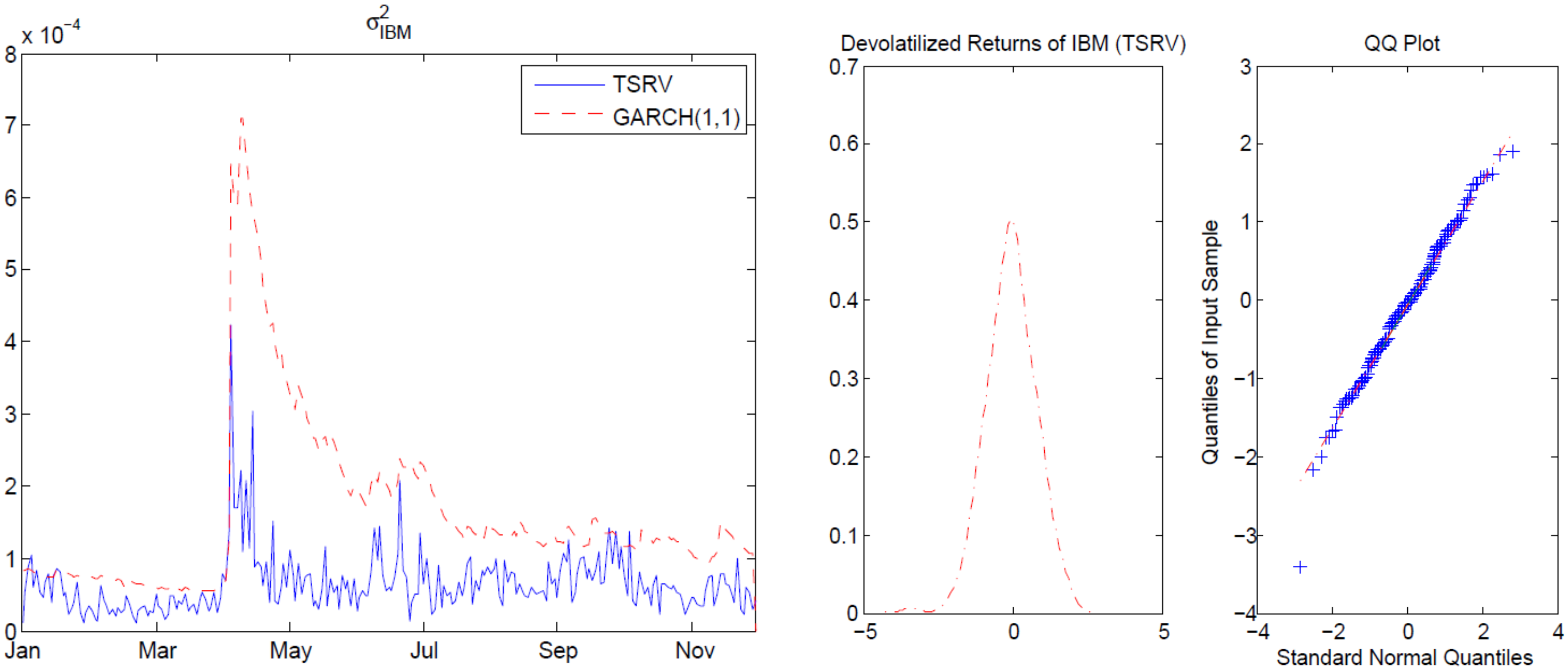
$$TSRV_t = \left(\frac{1}{Q}\right)\sum_{q=1}^{Q} RV_t^{(q,n_q)} - \bar{n}\ /nRV_t$$

Where $\bar{n} = \sum_{q=1}^{Q} n_q$

Realized Kernel*: Hansen, P. R., Lunde, A. (2006b). Realized variance and market microstructure noise (with discussion), Journal of Business and Economic Statistics 24:127–218.*

Pre-averaging*: Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., & Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. Stochastic processes and their applications, 119(7), 2249-2276.*

Example: Realized volatility and GARCH



(Daily) variance of IBM for year 2005.

R lab

Run the following code to load the data set Tbrate, which has three variables: the 91-day T-bill rate, the log of real GDP, and the inflation rate. In this lab you will use only the T-bill rate.

```
data(Tbrate,package="Ecdat")
library(tseries)
library(fGarch)
# r = the 91-day treasury bill rate
# y = the log of real GDP
# pi = the inflation rate
Tbill = Tbrate[,1]
Del.Tbill = diff(Tbill)
```

Problem 1 Plot both Tbill and Del.Tbill. Use both time series and ACF plots. Also, perform ADF and KPSS tests on both series. Which series do you think are stationary? Why? What types of heteroscedasticity can you see in the Del.Tbill series?

Data:6_Tbrate.txt

R lab

In the following code, the variable Tbill can be used if you believe that series is stationary. Otherwise, replace Tbill by Del.Tbill. This code will fit an ARMA/GARCH model to the series.

```
garch.model.Tbill = garchFit(formula= ~arma(1,0) + garch(1,0),Tbill)
summary(garch.model.Tbill)
garch.model.Tbill@fit$matcoef
```

Problem 2

a) Which ARMA/GARCH model is being fit? Write down the model using the same parameter names as in the R output.
b) What are the estimates of each of the parameters in the model?

R lab

Next, plot the residuals (ordinary or raw) and standardized residuals in various ways using the code below. The standardized residuals are best for checking the model, but the residuals are useful to see if there are GARCH effects in the series.

```
res = residuals(garch.model.Tbill)
res_std = res / garch.model.Tbill@sigma.t
par(mfrow=c(2,3))
plot(res)
acf(res)
acf(res^2)
plot(res_std)
acf(res_std)
acf(res_std^2)
```

R lab

Problem 3

a) Describe what is plotted by acf(res). What, if anything, does the plot tell you about the fit of the model?

b) Describe what is plotted by acf(res^2). What, if anything, does the plot tell you about the fit of the model?

c) Describe what is plotted by acf(res_std^2). What, if anything, does the plot tell you about the fit of the model?

d) What is contained in the variable garch.model.Tbill@sigma.t?

e) Is there anything noteworthy in the plot produced by the code plot(res_std)?

Problem 4 Now find an ARMA/GARCH model for the series del.log.-tbill, which we will define as diff(log(Tbill)). Do you see any advantages of working with the differences of the logarithms of the T-bill rate, rather than with the difference of Tbill as was done earlier?

R lab

On Black Monday, the return on the S&P 500 was -22.8%. Ouch! This exercise attempts to answer the question, “what was the conditional probability of a return this small or smaller on Black Monday?" “Conditional“ means given the information available the previous trading day. Run the following R code:

```
library(Ecdat)
library(fGarch)
data(SP500,package="Ecdat")
returnBlMon = SP500$r500[1805]
x = SP500$r500[(1804-2*253+1):1804]
plot(c(x,returnBlMon))
results = garchFit(~arma(1,0)+garch(1,1),data=x,cond.dist="std")
dfhat = as.numeric(results@fit$par[6])
forecast = predict(results,n.ahead=1)
```

R Lab

The S&P 500 returns are in the data set SP500 in the Ecdat package. The returns are the variable r500. (This is the only variable in this data set.) Black Monday is the 1805th return in this data set. This code fits an AR(1)/GARCH(1,1) model to the last two years of data before Black Monday, assuming 253 trading days/year. The conditional distribution of the white noise is the *t*-distribution (called “std" in garchFit). The code also plots the returns during these two years and on Black Monday.

From the plot you can see that Black Monday was highly unusual. The parameter estimates are in results@fit$par and the sixth parameter is the degrees of freedom of the *t*-distribution. The predict function is used to predict one-step ahead, that is, to predict the return on Black Monday; the input variable n.ahead specifies how many days ahead to forecast, so n.ahead=5 would forecast the next five days. The object forecast will contain meanForecast, which is the conditional expected return on Black Monday, meanError, which you should ignore, and standardDeviation, which is the conditional standard deviation of the return on Black Monday.

R lab

Problem 5 Use the information above to calculate the conditional probability of a return less than or equal to -0.228 on Black Monday.

Problem 6 Compute and plot the standardized residuals. Also plot the ACF of the standardized residuals and their squares. Include all three plots with your work. Do the standardized residuals indicate that the AR(1)/GARCH(1,1) model fits adequately?

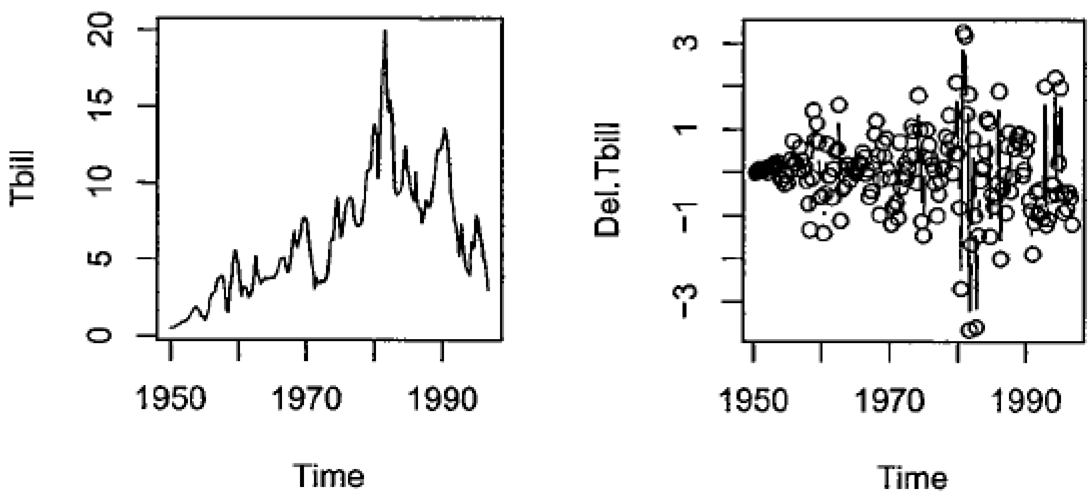
Problem 7 Would an AR(1)/ARCH(1) model provide an adequate fit? (Warning: If you apply the function summary to an fGarch object, the AIC value reported has been normalized by division by the sample size. You need to multiply by the sample size to get AIC.)

Problem 8 Does an AR(1) model with a Gaussian conditional distribution provide an adequate fit? Use the arima function to fit the AR(1) model. This function only allows a Gaussian conditional distribution.

R lab – Results & Discussions

Problem 1 Which series do you think are stationary? Why? What types of heteroskedasticity can you see in the Del.Thill series?

The plot of Tbill (below, left) looks non-stationary because it seems not to be mean-reverting. The plot of Del.Thill (below, right) appears stationary though there is clearly some volatility clustering and an increase in volatility with time (or with the T-bill rate).



R lab – Results & Discussions

Problem 1 Which series do you think are stationary? Why? What types of heteroskedasticity can you see in the Del.Thill series?

The ADF test accepts the null hypothesis of unit-root nonstationary for Tbill and the KPSS test rejects the null hypothesis of stationary for this series. Thus, the two tests agree with each other and with the plot of Tbill that the rate is non-stationary.

```
> adf.test(Tbill)
Augmented Dickey-Fuller Test data: Tbill
Dickey-Fuller = -1.925,  Lag order = 5, p-value 0.6075
alternative hypothesis: stationary
> kpss.test(Tbill)
KPSS Test for Level Stationarity data: Tbill
KPSS Level= 2.777, Truncation lag parameter 3, p-value 0.01
```

For Del.Tbill, the ADF tests reject the null hypothesis of unit-root non-stationarity and KPSS test accepts the null hypothesis of stationarity (the p-value is greater than 0.1). Thus, from both tests and the plot, we conclude that Del. Tbill is stationary.

R lab – Results & Discussions

Problem 2 (a) Which ARMA/GARCH model is being fit? Write down the model using the same parameter names as in the R output.
(b) What are the estimates of each of the parameters in the model?

(a) The model is AR(1) for the conditional mean and ARCH(1) for the conditional variance. The model for the conditional mean is $\mu + \phi(Y_{t-1} - \mu)$ and the model for the conditional variance is $\omega + \alpha_1 Y_{t-1}^2$.

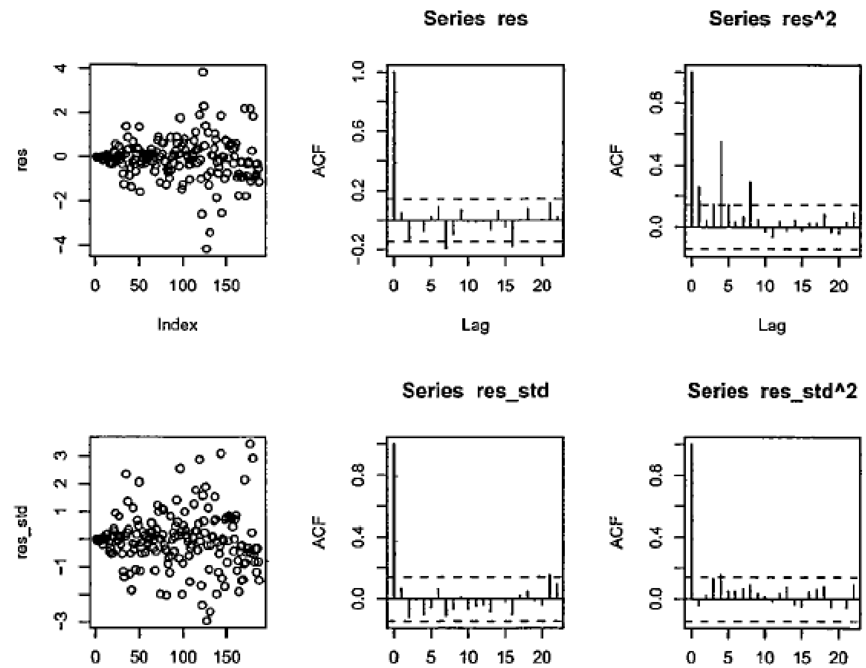
(b) The estimates arc in the table below. Here ar1 is ϕ.
Error Analysis:

	Estimate	Std. Error	t value	Pr(>ltl)	
mu	0.08350	0.05391	1.549	0.121395	
ar1	0.24163	0.07280	3.319	0.000902	***
omega	0.33816	0.06145	5.503	3.73e-08	***
alpha1	0.83483	0.24295	3.436	0.000590	***

R lab – Results & Discussions

Problem 3 (a,b,c) Describe what is plotted by acf(res)/acf(res^2)/ acf(res_std 2). What, if anything, does the plot tell you about the fit of the model?

(a) acf (res) plots the autocorrelation function of the residuals. The autocorrelations at low lags are not significantly different from 0, which indicates that the AR(1) model for the conditional mean is adequate.



(b) acf (res^2) plots the autocorrelation function of the squared residuals. Several are significant, which shows that there is volatility clustering. However, this does not indicate a problem with the model, since the ARCH(1) specification may be adequate. We need to look at the ACF of the standardized residuals to check.

(c) acf (res_std^2) plots the ACF of the squared standardized residuals. The standardized residuals are the raw residuals divided by the estimated conditional standard deviation. The ACF of the squared standardized residuals does not have significant autocorrelations at low lags and indicates that the ARCH(1) model for the conditional variance is adequate.

R lab – Results & Discussions

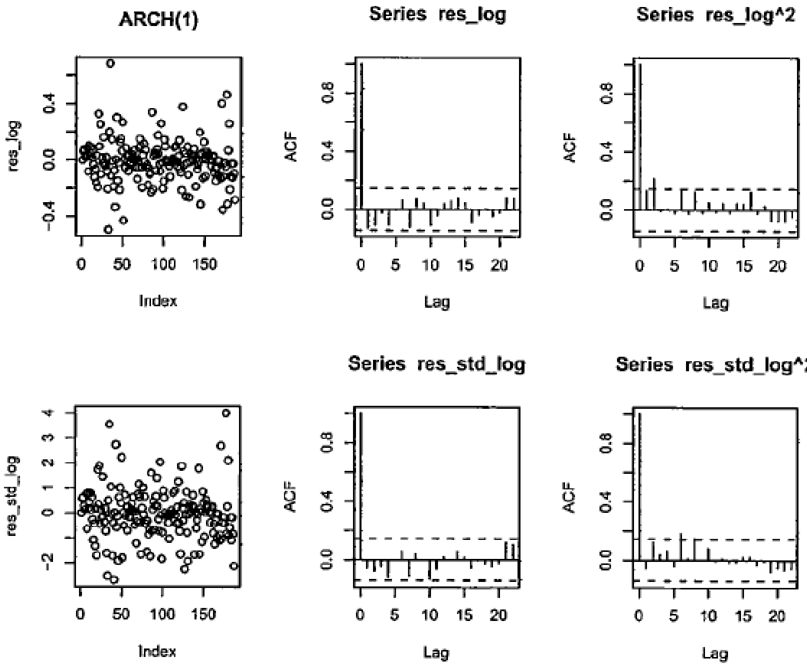
Problem 3 (d) What is contained in the variable garch.model.Tbill @sigma.t?
(e) Is there anything noteworthy in the plot produced by the code plot(res_std)?

(d) garch. model.Tbill@sigma. t contains the estimates of the conditional standard deviations.
(e) As can be seen in the lower left plot above, the variability of res_std increases with time (as does the Tbill rate itself- see the plot in Problem 1). This problem could perhaps be remedied by using a transformation of Tbill such as a log transformation.

R lab – Results & Discussions

Problem 4 Do you see any advantages of working with the differences of the logarithms of the T-bill rate, rather than with the difference of Tbill as was done earlier?

In the following plots, the fit of the AR(1) /ARCH(1) model to del.log.tbill shows an inadequacy of the ARCH(1) model for the conditional variance because there is autocorrelation in the squared standardized residuals.

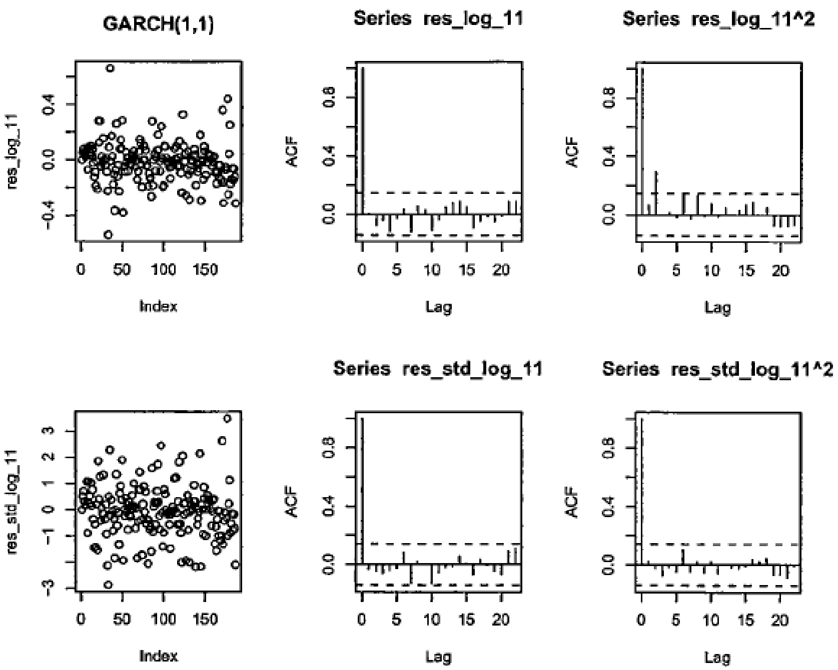


R lab – Results & Discussions

Problem 4 Do you see any advantages of working with the differences of the logarithms of the T-bill rate, rather than with the difference of Tbill as was done earlier?

However, if we use a GARCH(1,1) model for the conditional variance, then the residual plots look good, as can be seen in the next figure.

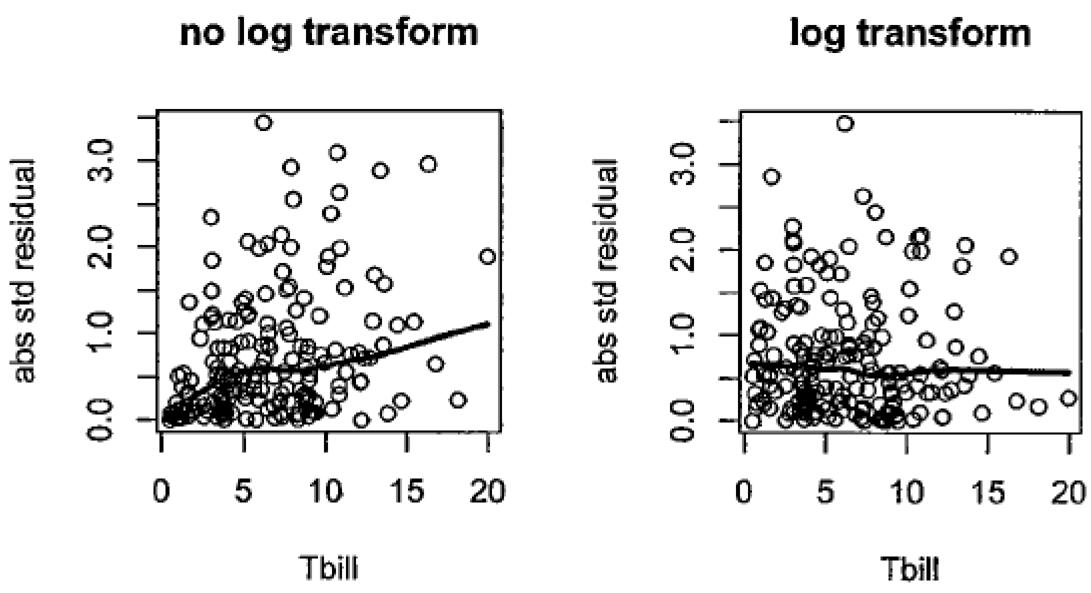
The advantage of using a log transformation prior to taking differences is that the log transformation removes a type of heteroscedasticity that GARCH models cannot model adequately, specifically that in the original Tbill series the conditional variance of the difference is proportional to the rate. This behaviour can be seen in the plots in Problem 1.



R lab – Results & Discussions

Problem 4 Do you see any advantages of working with the differences of the logarithms of the T-bill rate, rather than with the difference of Tbill as was done earlier?

The best way to detect heteroscedasticity is to plot absolute residuals. This is done in the following graph. The plots in this graph show that the standardized residuals from AR(1) / ARCH(1) fit to Del.Tbill still have this type of heteroscedasticity (see left plot), but the standardized residuals from AR(1)/GARCH(1,1) fit to del.log.tbill do not (right plot). The R function lowess was used to add smooth curves to the plots. Using lowess helps one see the patterns but is not essential.



R lab – Results & Discussions

Problem 4 Do you see any advantages of working with the differences of the logarithms of the T-bill rate, rather than with the difference of Tbill as was done earlier?

Here is the additional R code need to produce the plot.

```
garch.model.Tbill_log_11 = garchFit(formula= -arma(1,0)+ garch(1,1),del.log.tbill)
summary(garch.model.Tbill_log_11)
res_log_11 = residuals(garch.model.Tbill_log_11) res_std_log_11 = res_log I
garch.model.Tbill_log_11@sigma.t par(mfrow=c(1,2))
plot (Tbill [-1] ,abs(res_std) ,xlab="Tbill" ,ylab="abs std residual", main="no log
transform")
lines(lowess (Tbill [-1], abs (res_std)), col="red", h7d=3)
plot (Tbill [-1] , abs (res_std_log_11), xlab="Tbill", ylab="abs std residual",
main="log
transform")
lines(lowess(Tbill[-1] ,abs(res_std_log_11)),col="red",lwd=3)
```

R lab – Results & Discussions

Problem 5 Use the information above to calculate the conditional probability of a return less than or equal to -0.228 on Black Monday.

```
> probBlackMonday = pstd(returnB1Mon,mean=forecast$meanForecast,
sd=forecast$standardDeviation,nu=dfhat)
> round(probBlackMonday,7)
[1] 2.11e-05
```

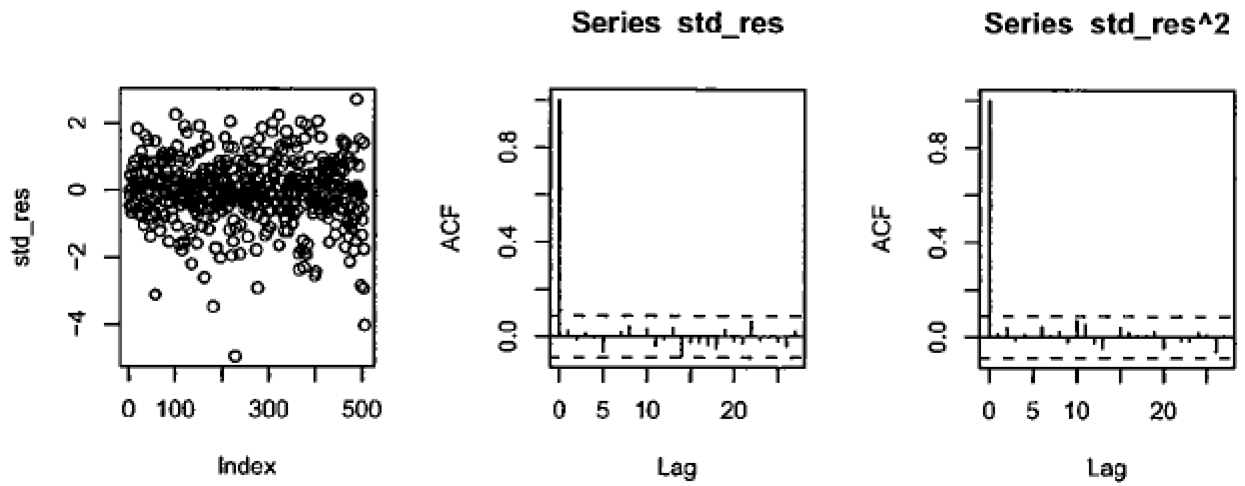
The conditional probability is 2.11e-05.

R lab – Results & Discussions

Problem 6 Do the standardized residuals indicate that the AR(1)/GARCH(1,1) model fits adequately?

The plots are produced by the R code below. The first ACF plot shows no correlation in the standardized residuals, which indicates that the AR(1) model is suitable for the conditional mean. Moreover, there is no autocorrelation in the squared standardized residuals, so the GARCH(1,1) model is suitable for the conditional variance.

```
std_res =results@residuals/results@sigma.t
par(mfrow=c(1,3))
plot(std_res)
acf(std_res)
acf(std_res^2)
```



R lab – Results & Discussions

Problem 7 Would an AR(1)/ARCH(1) model provide an adequate fit?

The AR(1)/GARCH(1,0) model does provide an adequate fit, though the AR(1)/GARCH(1,1) model's fit is slightly better.
The normalized AIC values are -6.518632 and -6.518026 for the AR(1)/GARCH(1,1) and AR(1)/GARCH(1,0) models respectively. Multiplying by the sample size, the AIC values are -3298.428 and-3298.121, so the AR(1)/GARCH(1,1) model is only slightly better than the AR(1)/GARCH(1,0) model.

```
> -6.518632 *length(x)
[1] -3298.428
> -6.518026 * length(x)
[1] -3298.121
```

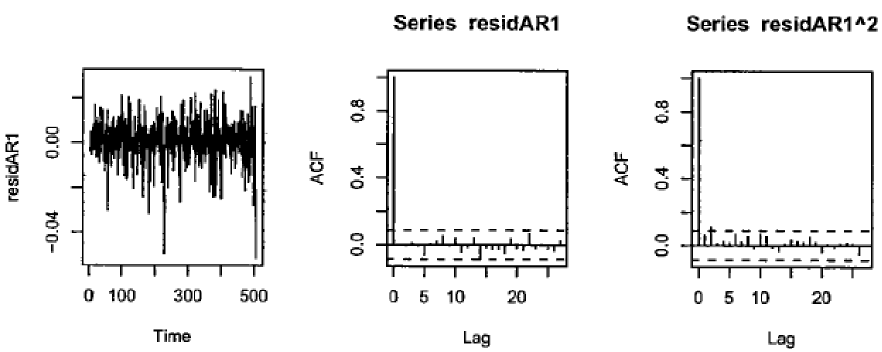
The standardized residual plots for the AR(1)/GARCH(1,0) do not show any serious problems. However, the lag-2 autocorrelation of the squared standardized residuals of the AR(1)/GARCHI(1,0) model hits the test bound, but this autocorrelation is much smaller with the AR(1)/GARCH(1,1) model.

R lab – Results & Discussions

Problem 8 Does an AR(1) model with a Gaussian conditional distribution pro vide an adequate fit?

The fit of the AR(1) model with a Gaussian conditional distribution is worse than the fit of the AR(1)/GARCH models.

```
fitAR1 = arima(x,order=c(1,0,0))
fitAR1
par(mfrow=c(1,3))
residAR1 = residuals(fitAR1)
plot(residAR1)
acf(residAR1)
acf(residAR1^2)
```



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

Coefficients:

	ar1	intercept
	0.1275	8e-04
s.e.	0.0455	5e-04

sigma-2 estimated as 9.45e-05: log likelihood= 1626.52, aic = -3247.03

The AIC is -3247.03 and is much larger than AIC for the AR(1)/GARCH models. Also, the squared residuals show some autocorrelation, especially at lag 2.

The poor fit of the AR(1) model could be due both to the lack of a GARCH effect in the model and to the conditional Gaussian assumption.