Volatility Measurement, Modeling, and Forecasting

Ka-fu WONG

University of Hong Kong

February 3, 2020

Contents

$$y_t = \epsilon_t$$

$$\epsilon_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$$

1. Leptokurtic distribution of stock returns

Normal distribution has $skewness = 0$ and $kurtosis = 3$.

A distribution with positive excess kurtosis is called leptokurtic. In terms of shape, a leptokurtic distribution has fatter tails.

Stock returns are well documented to be leptokurtic.

Table I, summary statistics of daily returns of Dow Jones Industrial Average, from Brock et al. (1992)[1]

Panel A: Daily Returns

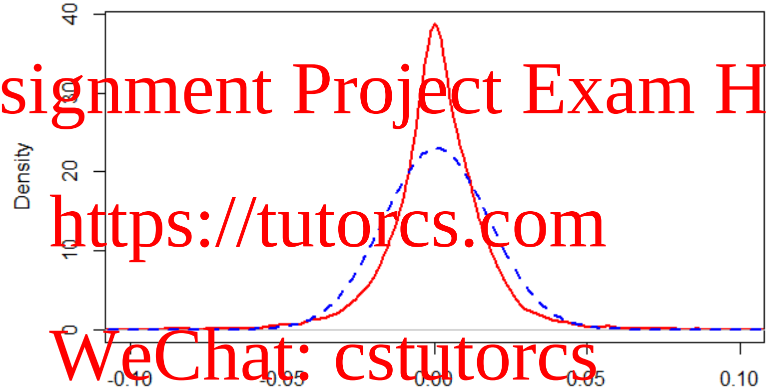
	Full Sample	97–14	15–38	39–62	62–86
N	25036	5255	7163	6442	6155
Mean	0.00017	0.00012	0.00014	0.00020	0.00020
Std.	0.0108	0.0099	0.0147	0.0075	0.0088
Skew	−0.1047**	−0.4804**	0.0193	−0.7614**	0.2707**
Kurtosis	16.00**	8.86**	12.75**	13.60**	11.57**
$\rho(1)$	0.033**	0.013	0.009	0.117**	0.079**
$\rho(2)$	−0.026**	−0.020	−0.029*	−0.068**	−0.001
$\rho(3)$	0.012*	0.041**	−0.006	0.036**	0.009
$\rho(4)$	0.046**	0.085**	0.055**	0.028*	−0.012
$\rho(5)$	0.022**	0.042**	[illegible]	0.014	−0.011
Bartlett std.	0.006	0.014	0.012	0.012	0.013

[1] Brock, W., Lakonishok, J., & LeBaron, B. (1992): "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns," *The Journal of Finance*, 47(5), 1731-1764.

Summary statistics of daily returns of Hang Seng Index

	HSI (1987-2012)	Dow Jones (1897-1986)
N	6462	25036
Mean	0.00049	0.00017
Std.	0.0174	0.0108
Skew	-1.2416	-0.1047
Kurtosis	36.04	16.00
$\rho(1)$	0.013	0.033
$\rho(2)$	-0.021	-0.026
$\rho(3)$	0.050	0.012
$\rho(4)$	-0.025	0.046
$\rho(4)$	-0.032	0.022
Bartlett std.	0.012	0.006

Density plot of daily returns of Hang Seng Index



Note: Red solid line is the empirical density of daily returns. Blue dotted line is the normal density with the mean and sd equal to the daily return's

Can we build models that maintain the normality assumption and yet account for the leptokurtic shape?

The answer is models of ARCH or GARCH family.

- ARCH: *A*uto*R*egressive *C*onditional *H*eteroskedasticity
- GARCH: *G*eneralized *A*uto*R*egressive *C*onditional *H*eteroskedasticity

ARCH was first developed by Engle (1982).

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4): 987-1007.

There are since over 150 different ARCH models developed.

Bollerslev, T. (2010): "Glossary to ARCH (GARCH)," in *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle* (eds. Tim Bollerslev, Jeffrey R. Russell -10- and Mark W. Watson), Chapter 8, pp.137-163. Oxford, U.K.: Oxford University Press, 2010.

2. Importance of volatility

Good volatility forecasts are crucial for the implementation and evaluation of asset and derivative pricing theories as well as trading and hedging strategies.

2.1. **Asset allocation.**

Two assets:

- a risky and a riskless (i.e., volatility = 0)
- Risky asset generally has a higher expected return than the riskless assets.

We would like to invest in a portfolio consisting of the two assets.

- Given the expected return on the risky and riskless assets, when the risky asset has a very *high volatility*, the portfolio will consist of the riskless asset only.
- Given the expected return on the risky and riskless assets, when the risky asset has a very *low volatility*, the portfolio will consist of more risky assets.

2.2. **Carry trade.**

Suppose interest rate is low in Japan, and high in US.

- Borrow Japanese Yen at the low Japanese interest rate, and lend US dollar at the slightly higher US interest rate.
- Borrow Japanese Yen at the low Japanese interest rate, and lend HK dollar at the slightly higher HK interest rate.

Profit from such trade is low risk when

- the volatility of exchange rate is *low*, and / or
- the Yen is expected to *depreciate*.

Suppose interest rate is low in US or HK dollar deposits, high in Renminbi deposit.

- Borrow HK dollar at the low Hong Kong interest rate, and lend Renminbi at the slightly higher Renminbi interest rate.

Profit from such trade is low risk when

- the volatility of exchange rate is *low*, and / or
- the Renminbi is expected to *appreciate*.

Uncovered interest rate parity:
At equilibrium, the expected change in the spot exchange rate must equal to the different in exchange rate, i.e.,

Expected gain from interest rate differential	=	Expected Loss from depreciation
$i^*_{t,t+k} - i_{t,t+k}$		$E_t(S_{t+k}) - S_t$

Time	HKD		Foreign Currency		HKD
t	1	$\rightarrow$	S_t		
	$\downarrow$		$\downarrow$		
$t+k$	$(1+i)$		$S_t \times (1+i^*)$	$\rightarrow$	$S_t \times (1+i^*)/S_{t+k}$

2.3. **Impact of Volatility on Macroeconomy.**

- The variance of inflation may have *impact on* various macro and government *decisions*.
  - For example, uncertainty in the adjustment of unemployment benefits and social security.
- *High variance in inflation* may also imply welfare loss.

- Previous studies have tried to measure the time-varying variance of inflation.

3. Clustering of Volatility

It is a well-established fact, dating back to Mandelbrot (1963)[2] and Fama (1965)[3], that *financial returns* display pronounced *volatility clustering*.

- Days of high volatility tend to be followed by days of high volatility
- Days of low volatility tend to be followed by days of low volatility

[2] Mandelbrot, B. (1963): "The Variation of Certain Speculative Prices," *The Journal of Business*, 36(4), 394-419.

[3] Fama, E. (1965): "The behavior of stock market prices," *Journal of Business*, 38(1), 34-105.

We would like to build models that allow

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

and the *σ_{t-1}^2 should be close to σ_t^2*, in processes like

$$y_t = \epsilon_t$$

or

$$y_t = \text{a bunch of things} + \epsilon_t$$

or

$$(1 - \beta_1 L - \beta_2 L^2 - ... - \beta_p L^p) y_t = (1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q) \epsilon_t$$

Problem: If we model variance as such, y_t will be *non-stationary* because variance is no longer constant across t.

4. Models that Ensure Stationarity

Consider

$$y_t = \epsilon_t$$

Stationarity means that we must have

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

How to allow volatility to change with t and yet keep the unconditional variance of ϵ_t same across t?

The solution is to work with *conditional variance*!

The conditional variance of ϵ_t

$$Var(\epsilon_t|\Omega_{t-1}) = E(\epsilon_t^2|\Omega_{t-1})$$

is allowed to change over the conditioned information $\Omega_{t-1} \equiv \{\epsilon_{t-1}, \epsilon_{t-2}, ...\}$ that is available at time $t-1$.

With this specification of conditional variance of ϵ_t

$$Var(\epsilon_t|\Omega_{t-1}) = E(\epsilon_t^2|\Omega_{t-1}),$$

we can maintain stationarity because the unconditional variance of ϵ_t

$$Var(\epsilon_t) = E[E(\epsilon_t^2|\Omega_{t-1})]$$

is constant across t.

That is,

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

and

$$\epsilon_t|\Omega_{t-1} \sim WN(0, \sigma_t^2)$$

5. How to Ensure Volatility Clustering

Volatility clustering can be understood as *persistence in conditional variance*.

We have seen processes with *persistence* (or presistence in conditional mean) — ARMA models.

Recall, a shock to the $MA(1)$ models will have an impact on conditional mean for 1 period.

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

$$E(y_t) = E(\epsilon_t) + \theta_1 E(\epsilon_{t-1}) = 0$$

$$E(y_t|\Omega_{t-1}) = \theta_1 \epsilon_{t-1}$$
$$E(y_{t+1}|\Omega_t) = \theta_1 \epsilon_t$$
$$E(y_{t+2}|\Omega_{t+1}) = \theta_1 \epsilon_{t+1}$$

Conditional mean is changing with time but unconditional mean is constant!

Recall $AR(1)$ models are like $MA(\infty)$.

$$y_t = \epsilon_t + \rho_1 y_{t-1}$$

$$E(y_t) = 0$$

$$E(y_t|\Omega_{t-1}) = \rho_1 y_{t-1}$$
$$E(y_{t+1}|\Omega_t) = \rho_1 y_t$$
$$E(y_{t+2}|\Omega_{t+1}) = \rho_1 y_{t+1}$$

Conditional mean is changing with time but unconditional mean is constant!

Stare at the definition of conditional mean again.

$$\sigma_t^2 = E(\epsilon_t^2 | \Omega_{t-1})$$

Learning from the persistence in conditional mean (conditional expectation) and noting that *variance is just an expectation of squared innovations*, it is logical to introduce persistence in conditional variance (conditional expectation of squared innovations) by

$$E(\epsilon_t^2 | \Omega_{t-1}) = \gamma_1 \epsilon_{t-1}^2$$

Conditional mean (AR(1))	Conditional variance
$E(y_t \vert \Omega_{t-1}) = \rho_1 y_{t-1}$	$E(\epsilon_t^2 \vert \Omega_{t-1}) = \gamma_1 \epsilon_{t-1}^2$

$$E(\epsilon_t^2|\Omega_{t-1}) = \gamma_1 \epsilon_{t-1}^2$$

With such specification, a big shock (ϵ_t) today will have an *impact* on the volatility of next period, $\sigma^2_{t+1} = E(\epsilon^2_{t+1}|\Omega_t)$.

- If $\gamma_1 > 0$, a big shock (ϵ_t) today will lead to a *bigger* volatility of next period, $\sigma^2_{t+1} = E(\epsilon^2_{t+1}|\Omega_t)$.
- If $\gamma_1 < 0$, a big shock (ϵ_t) today will lead to a *negative* volatility of next period, i.e., $\sigma^2_{t+1} = E(\epsilon^2_{t+1}|\Omega_t) < 0$.

The discussion implies the *restriction*: $\gamma_1 > 0$.

Even if the *restriction* ($\gamma_1 > 0$) is satisfied, this model is not good because it implies *zero volatility* when $\epsilon_{t-1} = 0$. A modification is

$$E(\epsilon_t^2|\Omega_{t-1}) = \omega + \gamma_1 \epsilon_{t-1}^2$$

where $\omega > 0$, or

$$\epsilon_t^2 = \omega + \gamma_1 \epsilon_{t-1}^2 + v_t$$

where $v_t \equiv \epsilon_t^2 - E(\epsilon_t^2|\Omega_{t-1}) = \epsilon_t^2 - \sigma_t^2$ is white noise. That is, ϵ_t^2 follows $AR(1)$.

This type of model is called *A*uto*R*egressive *C*onditional *H*eteroskedasticity of order 1, or $ARCH(1)$.

Note the similarity of

- the *AR(1) in levels/mean* and
- the *AR(1) in Conditional Heteroskedasticity*.

Conditional mean (AR(1))	Conditional variance
$E(y_t\|\Omega_{t-1}) = \rho_1 y_{t-1}$	$E(\epsilon_t^2\|\Omega_{t-1}) = \omega + \gamma_1 \epsilon_{t-1}^2$

Process of y_t	Process of ϵ_t^2
$y_t = \rho_1 y_{t-1} + u_t$	$\epsilon_t^2 = \omega + \gamma_1 \epsilon_{t-1}^2 + v_t$
$u_t \equiv y_t - E(y_t\|\Omega_{t-1})$	$v_t \equiv \epsilon_t^2 - E(\epsilon_t^2\|\Omega_{t-1})$

More general, we can have $ARCH(p)$ as

$$E(\epsilon_t^2|\Omega_{t-1}) = \omega + \gamma_1\epsilon_{t-1}^2 + ... + \gamma_p\epsilon_{t-p}^2$$

and ϵ_t^2 follows $AR(p)$.

$$\epsilon_t^2 = \omega + \gamma_1\epsilon_{t-1}^2 + ... + \gamma_p\epsilon_{t-p}^2 + v_t$$

A big shock to y_t in time t (ϵ_t) will cause an increase in the conditional variance, and hence a likely bigger changes to y_t — *of either directions* — in the following p periods.

6. Some properties of $ARCH(p)$

Consider $MA(1)$ with $ARCH(1)$.

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

$$\epsilon_t | \Omega_{t-1} \sim N(0, \sigma_t^2),$$

$$\sigma_t^2 \equiv E(\epsilon_t^2 | \Omega_{t-1}) = \omega + \gamma_1 \epsilon_{t-1}^2$$

(1) Mean of y_t: $E(y_t) = E(\epsilon_t) + \theta_1 E(\epsilon_{t-1}) = 0$
(2) Variance of y_t: $Var(y_t) = Var(\epsilon_t) + \theta_1^2 Var(\epsilon_{t-1}) = (1 + \theta_1^2)\sigma^2$
(3) Mean of ϵ_t: $E(\epsilon_t) = 0$

(4) Variance of ϵ_t:

$$\sigma^2 = E(\epsilon_t^2) = E\left[E(\epsilon_t^2|\Omega_{t-1})\right] = E\left[\omega + \gamma_1 \epsilon_{t-1}^2\right]$$

$$E(\epsilon_t^2) = \omega + \gamma_1 E(\epsilon_{t-1}^2)$$

$$\sigma^2 = \omega + \gamma_1 \sigma^2$$

$$\sigma^2 = \frac{\omega}{1-\gamma_1}$$

(5) *Implied restrictions*:

(a) $\omega > 0$
(b) $\gamma_1 > 0$
(c) $1-\gamma_1 > 0$ or $\gamma_1 < 1$

Consider $MA(q)$ with $ARCH(p)$.

$$y_t = \epsilon_t + \alpha_1\epsilon_{t-1} + \alpha_2\epsilon_{t-2} + ... + \alpha_q\epsilon_{t-q}$$

$$\epsilon_t|\Omega_{t-1} \sim N(0, \sigma_t^2);$$

$$\sigma_t^2 = E(\epsilon_t^2|\Omega_{t-1}) = \omega + \gamma_1\epsilon_{t-1}^2 + ... + \gamma_p\epsilon_{t-p}^2$$

(1) Mean of y_t: $E(y_t) = 0$
(2) Variance of y_t: $Var(y_t) = (1 + \alpha_1^2 + ... + \alpha_q^2)\sigma^2$
(3) Mean of ϵ_t: $E(\epsilon_t) = 0$

(4) Variance of ϵ_t:

$$\sigma^2 = E(\epsilon_t^2) = E\left[E(\epsilon_t^2|\Omega_{t-1})\right] = E\left[\omega + \gamma_1\epsilon_{t-1}^2 + ... + \gamma_p\epsilon_{t-p}^2\right]$$

$$E(\epsilon_t^2) = \omega + \gamma_1 E(\epsilon_{t-1}^2) + ... + \gamma_p E(\epsilon_{t-p}^2)$$

$$\sigma^2 = \omega + \gamma_1\sigma^2 + ... + \gamma_p\sigma^2$$

$$\sigma^2 = \frac{\omega}{1-(\gamma_1+\gamma_2+...+\gamma_p)} = \frac{\omega}{1-\sum_{i=1}^p \gamma_i}$$

(5) *Implied restrictions*:

(a) $\omega > 0$

(b) $\gamma_i > 0,\ i = 1, 2, ..., p$

(c) $1 - \sum_{i=1}^p \gamma_i > 0$ or $\sum_{i=1}^p \gamma_i < 1$

7. How to simulate ARCH(1)?

Suppose we are interested in generating T observations of ϵ_t that has the property of $ARCH(1)$.

$$\epsilon_t \sim N(0, \sigma_t^2), \quad \sigma_t^2 = \omega + \gamma_1 \epsilon_{t-1}^2$$

(1) Fixed the parameters of ω and γ_1. Compute the unconditional variance of ϵ_t.

$$\sigma^2 = \frac{\omega}{1 - \gamma_1}$$

(2) Generate $T + 1$ observations of *standard normal* random variables, $v_0, v_1, \ldots, v_T$

(3) Generate ϵ_t recursively

- For $t = 0$, initialize $\sigma_t^2 = \sigma^2$, $\epsilon_t = v_t \sigma_t$
- For $t = 1, ..., T$, update $\sigma_t^2 = \omega + \gamma_1 \epsilon_{t-1}^2$, and $\epsilon_t = v_t \sigma_t$

For $t = 0$, initialize $\sigma_0^2 = \sigma^2$, $\epsilon_0 = v_0\sigma_0$

- For $t = 1$, $\sigma_1^2 = \omega + \gamma_1\epsilon_0^2$, and $\epsilon_1 = v_1\sigma_1$
- For $t = 2$, $\sigma_2^2 = \omega + \gamma_1\epsilon_1^2$, and $\epsilon_2 = v_2\sigma_2$
- Similarly, for $t = 3, \ldots T$, update $\sigma_t^2 = \omega + \gamma_1\epsilon_{t-1}^2$, and $\epsilon_t = v_t\sigma_t$

t	σ_t^2	v_t	ϵ_t
0	σ^2	v_0	$\epsilon_0 = v_0\sigma_0$
1	$\sigma_1^2 = \omega + \gamma_1\epsilon_0^2$	v_1	$\epsilon_1 = v_1\sigma_1$
2	$\sigma_2^2 = \omega + \gamma_1\epsilon_1^2$	v_2	$\epsilon_2 = v_2\sigma_2$
3	$\sigma_3^2 = \omega + \gamma_1\epsilon_2^2$	v_3	$\epsilon_3 = v_3\sigma_3$
...	...	...	...
...	...	...	...
T	$\sigma_T^2 = \omega + \gamma_1\epsilon_{T-1}^2$	v_T	$\epsilon_T = v_T\sigma_T$

8. How to estimate ARCH models?

$$y_t = \epsilon_t$$

$$\epsilon_t|\Omega_{t-1} \sim N(0, \sigma_t^2), \quad E(\epsilon_t^2|\Omega_{t-1}) = \omega + \gamma_1 \epsilon_{t-1}^2$$

We can think of it as a "regression" model in ϵ_t^2.

$$\begin{aligned} \epsilon_t^2 &= \sigma_t^2 + \nu_t \\ \epsilon_t^2 &= \omega + \gamma_1 \epsilon_{t-1}^2 + \nu_t \end{aligned}$$

where $\nu_t = \epsilon_t^2 - \sigma_t^2$ or $\sigma_t^2 = \epsilon_t^2 - \nu_t$.

Use *M*aximum *L*ikelihood (ML).

The model implies ϵ_t is normal distribution with mean zero and conditional variance σ_t^2, that is

$$p(\epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_0) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{\epsilon_t^2}{2\sigma_t^2}\right)$$

Let the parameter vector be $\theta = (\omega, \gamma_1)$. The likelihood function is

$$L(\theta) = p(\epsilon_1) \times p(\epsilon_2|\epsilon_1) \times p(\epsilon_3|\epsilon_2, \epsilon_1) \times ... \times p(\epsilon_T|\epsilon_{T-1}, \epsilon_{T-2}, ..., \epsilon_1)$$

The log-likelihood function would be

$$l(\theta) = \ln p(\epsilon_1) + \ln p(\epsilon_2|\epsilon_1) + \ln p(\epsilon_3|\epsilon_2, \epsilon_1) + ... + \ln p(\epsilon_T|\epsilon_{T-1}, \epsilon_{T-2}, ..., \epsilon_1)$$

Choose θ to maximize $L(\theta)$ or $l(\theta)$ *numerically*.

Note:

For ARCH(1), we need ϵ_{t-1} to estimate σ_t^2, in particular, ϵ_1 to estimate σ_2^2. Thus,

$$l(\theta) = \ln p(\epsilon_2|\epsilon_1) + \ln p(\epsilon_3|\epsilon_2, \epsilon_1) + ... + \ln p(\epsilon_T|\epsilon_{T-1}, \epsilon_{T-2}, ..., \epsilon_1)$$

For ARCH(2), we need ϵ_{t-1} and ϵ_{t-2} to estimate σ_t^2, in particular ϵ_2 and ϵ_1 to estimate σ_3^2. Thus,

$$l(\theta) = \ln p(\epsilon_3|\epsilon_2, \epsilon_1) + ... + \ln p(\epsilon_T|\epsilon_{T-1}, \epsilon_{T-2}, ..., \epsilon_1)$$

9. The Inflation Example of Engle (1982)

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4): 987–1007.

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003 was divided equally between *Robert F. Engle III* "for methods of analyzing economic time series with time-varying volatility (*ARCH*)" and Clive W.J. Granger "for methods of analyzing economic time series with common trends (cointegration)".

http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2003/engle-bio.html

p_t: log quarterly consumer price index
w_t: log quarterly manual wage rate
Sample period: 1958-II to 1977-II

$$\Delta p_t = \beta_1 \Delta p_{t-1} + \beta_2 \Delta p_{t-4} + \beta_3 \Delta p_{t-5} + \beta_4 (p_{t-1} - w_{t-1}) + \beta_5 + \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$$

$$\Delta p_t = \beta_1 \Delta p_{t-1} + \beta_2 \Delta p_{t-4} + \beta_3 \Delta p_{t-5} + \beta_4 (p_{t-1} - w_{t-1}) + \beta_5 + \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$$

	β_1	β_2	β_3	β_4	β_5	$\alpha_0(\times 10^{-6})$	α_1
Estimate	0.334	0.408	-0.404	-0.0559	0.0257	89	–
St. Err.	0.103	0.110	0.114	0.0136	0.00572		
t Stat.	3.25	3.72	3.55	4.12	4.49		

$$\Delta p_t = \beta_1 \Delta p_{t-1} + \beta_2 \Delta p_{t-4} + \beta_3 \Delta p_{t-5} + \beta_4 (p_{t-1} - w_{t-1}) + \beta_5 + \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$$

	β_1	β_2	β_3	β_4	β_5	$\alpha_0(\times 10^{-6})$	α_1
Estimate	0.162	0.264	-0.325	-0.0707	0.0328	14	0.955
St. Err.	0.108	0.0892	0.0987	0.0115	0.00491	8.5	0.298
t Stat.	1.50	2.96	3.29	6.17	6.67	1.56	3.2

Conclusion: Evidence of ARCH.

10. $GARCH(p,q)$

$$y_t = \epsilon_t$$

$$\epsilon_t | \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$$E(\epsilon_t^2 | \Omega_{t-1}) = \omega + \gamma_1 \epsilon_{t-1}^2 + ... + \gamma_1 \epsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + ... + \beta_q \sigma_{t-q}^2$$

$$\omega > 0, \gamma_i \geq 0, \beta_i \geq 0, \sum \gamma_i + \sum \beta_i < 1$$

Using backward substitutions, we can show that $GARCH(p,q)$ can be written as an infinite-order ARCH process with some restriction in the coefficients.
(Analogy: An $ARMA(p,q)$ process can be written as $MA(\infty)$ process.)

That is, GARCH can be viewed as a *parsimonious* way to *approximate* a high order *ARCH* process!

$$y_t = \epsilon_t$$

(1) Mean of ϵ_t: $E(\epsilon_t) = 0$

(2) Variance of ϵ_t:

$$E(\epsilon_t^2) = E\left[E(\epsilon_t^2|\Omega_{t-1})\right]$$

$$= E\left[\omega + \gamma_1\epsilon_{t-1}^2 + ... + \gamma_p\epsilon_{t-p}^2 + \beta_1\sigma_{t-1}^2 + ... + \beta_q\sigma_{t-q}^2\right]$$

$$E(\epsilon_t^2) = \omega + \gamma_1 E(\epsilon_{t-1}^2) + ... + \gamma_p E(\epsilon_{t-p}^2)$$

$$+ \beta_1 E(\sigma_{t-1}^2) + ... + \beta_q E(\sigma_{t-q}^2)$$

$$\sigma^2 = \omega + \gamma_1\sigma^2 + ...\gamma_p\sigma^2 + \beta_1\sigma^2 + ... + \beta_q\sigma^2$$

$$\sigma^2 = \frac{\omega}{1 - (\gamma_1 + ... + \gamma_p + \beta_1 + ... + \beta_q)}$$

$$\sigma^2 = \frac{\omega}{1 - (\sum_{i=1}^{p}\gamma_i + \sum_{i=1}^{q}\beta_i)}$$

$$y_t = \epsilon_t$$

11. Important Properties of GARCH(p,q)

(1) Unconditional variance is fixed but conditional variance is time-varying

$$E(\epsilon_t) = 0$$

$$\sigma^2 = \frac{\omega}{1 - (\sum_{i=1}^{p} \gamma_i + \sum_{i=1}^{q} \beta_i)}$$

$$E(\epsilon_t^2|\Omega_{t-1}) = \omega + \gamma_1 \epsilon_{t-1}^2 + ... + \gamma_1 \epsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + ... + \beta_q \sigma_{t-q}^2$$

(2) Unconditional distribution of conditionally Gaussian GARCH is symmetric and leptokurtic.

- Real-world financial asset returns, are often found to symmetrically distributed and have a *fatter tail than Gaussian distribution*.
- Ordinary Gaussian distribution does not provide a good approximation of the asset returns, but the *Gaussian distribution with GARCH* does.

(3) Conditional prediction error variance varies with conditional information set. $E(\epsilon^2_{t+h}|\Omega_t)$ is complicated but can be computed. And, we can show the conditional prediction error variance *approaches* the unconditional prediction error variance.

$$\lim_{h\to\infty} E(\epsilon^2_{t+h}|\Omega_t) = E(\epsilon^2_{t+h})$$

(4) ϵ_t follows GARCH implies ϵ_t^2 follows an ARMA. Take $GARCH(1,1)$, for illustration. First, undestand that

$$E(\epsilon_t^2|\Omega_{t-1}) \equiv \sigma_t \Leftrightarrow \epsilon_t^2 \equiv \sigma_t + \nu_t$$

where $\nu_t \equiv \epsilon_t^2 - E(\epsilon_t^2|\Omega_{t-1}) = \epsilon_t^2 - \sigma_t^2$ is white noise.

$$\sigma_t^2 = \omega + \gamma_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$\epsilon_t^2 - \nu_t = \omega + \gamma_1 \epsilon_{t-1}^2 + \beta_1 \left(\epsilon_{t-1}^2 - \nu_{t-1}\right)$$

$$\epsilon_t^2 = \omega + (\gamma_1 + \beta_1)\epsilon_{t-1}^2 - \beta_1 \nu_{t-1} + \nu_t$$

That is, $GARCH(1,1)$ can be written as $ARMA(1,1)$ of ϵ_t^2. (Recall, $ARCH(1)$ can be written as $AR(1)$ of ϵ_t^2.)

12. Extension of ARCH and GARCH Models

12.1. Threshold GARCH (TGARCH or GJR-GARCH).

$$r_t = \text{a bunch of things} + \epsilon_t$$

$$\sigma_t^2 = \omega + \gamma_1 \epsilon_{t-1}^2 + \alpha D_{t-1} \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$D_{t-1} = \begin{cases} 1 & \epsilon_{t-1} < 0 \\ 0 & \epsilon_{t-1} \geq 0 \end{cases}$$

GJR=Glosten-Jagannathan-Runkle

Glosten, L.R., Jagannathan R. & Runkle, D.E. (1993): "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *The Journal of Finance*, 48(5), 1779-1801.

$$\sigma_t^2 = \omega + \gamma_1 \epsilon_{t-1}^2 + \alpha D_{t-1} \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$D_{t-1} = \begin{cases} 1 & \epsilon_{t-1} < 0 \\ 0 & \epsilon_{t-1} \geq 0 \end{cases}$$

Suppose "a bunch of things" is empty. Then, ϵ_t is the same as return.

- When the lagged return is positive (*good news yesterday*), $D = 0$, so the effect of the lagged squared return on the current conditional variance is simply γ_1.
- When the lagged return is negative (*negative news yesterday*), $D = 1$, so the effect of the lagged squared return on the current conditional variance is simply $\gamma_1 + \alpha$.
- Allowance for asymmetric response has proved useful for modeling "*leverage* effects" in stock returns, which occur when $\alpha < 0$.
  - α can be negative but cannot be too negative.
  - The restriction is $\gamma_1 + \alpha > 0$

Leverage effect on stock returns and volatility.
Consider a business that comprises of real estates. Assume that the business does not rent out the properties. It holds the properties only for capital gain. The value of the business is clearly just the value of the properties. If it were a listed company, with no debt, then the equity capitalization would be the value of the properties, and the volatility of the share price would be equal to the volatility of the price of properties.

For concreteness, suppose the properties are initially worth \$100 million (financed with \$100 millions of equity and \$0 millions of debt).
Consider a 10% (i.e., 10 million) increase in property values. The value of equity will increase by 10% (=10/100).
Consider another 10% (i.e., 11 million=10%$\times$ 110 millions) increase in property values. The value of equity will increase by 10% (=11/110).

Consider a 10% (i.e., 10 million) decrease in property values. The value of equity will decrease by 10% (=10/100).
Consider another 10% (i.e., 9 million=10%$\times$ 90 millions) increase in property values. The value of equity will decrease by 10% (=9/90).

That is, the impact of 10% change in property price is the same regardless of the increase or decrease in property price last period.

Now consider the same company financed with 50% debt (at zero interest) and 50% equity. The claims of debt holders is fixed in nominal dollars whereas the equity holders get the benefit/cost of a higher/lower property price.

Again, suppose the properties are initially worth \$100 million (financed with \$50 millions of equity and \$50 millions of debt).

Consider a 10% (i.e., 10 million) increase in property values. the value of equity becomes \$60 millions, while the value of debt remains at \$50 millions. Equity holders enjoy a 20% increase (=10/50) in share value, against 10%(=10/100) in the unlevered case. In moving from 0% gearing to 50% gearing, the volatility of equity value has doubled.

Consider another 10% (i.e., 11 million=10%$\times$ 110 millions) increase in property values. The value of equity will increase by 18.33% (=11/60).

Consider a 10% (i.e., 10 million) decrease in property values. The value of equity will decrease by 20% (=10/50).

Consider another 10% (i.e., 9 million=10%$\times$ 90 millions) decrease in property values. The value of equity will decrease by 22.5% (=9/40).

That is, the impact of 10% change in property price is depends on the increase or decrease in property price last period. 22.5% versus 18.33%!

That is, asymmetric volatility in returns, depending on whether the direction of change in returns last period.

12.2. **Exponential GARCH.**

$$\ln\left(\sigma_t^2\right) = \omega + \gamma\left|\frac{\epsilon_{t-1}}{\sigma_{t-1}}\right| + \alpha\frac{\epsilon_{t-1}}{\sigma_{t-1}} + \beta_1 \ln\left(\sigma_{t-1}^2\right)$$

Volatility is driven by both the size and sign of shocks (both positive and negative). Hence, the model allows for asymmetric response depending on the sign of news.

- When the shock is positive, the impact of $\frac{\epsilon_{t-1}}{\sigma_{t-1}}$ on $\ln\left(\sigma_t^2\right)$ is $\alpha + \gamma$.
- When the shock is negative, the impact of $\frac{\epsilon_{t-1}}{\sigma_{t-1}}$ on $\ln\left(\sigma_t^2\right)$ is $\alpha - \gamma$.
- There is no restrictions on the sign of the parameters!

12.3. **GARCH with exogenous variables.**

$$y_t = \text{a bunch of things} + \epsilon_t$$

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

Financial market volume, for example, often helps to explain market volatility.

12.4. **GARCH-in-Mean (i.e., GARCH-M).**

$$y_t = \delta_0 + \delta_1 x_t + \alpha\sigma_t^2 + \epsilon_t$$

$$\epsilon_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$$

$$E(\epsilon_t^2|\Omega_{t-1}) = \omega + \gamma\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

The conditional variance appears in the conditional mean equation.

- High risk, high return.

The conditional expected return is

$$E[y_t|\Omega_{t-1}] = \delta_0 + \delta_1 x_t + \alpha\sigma_t^2$$

13. Diagnosing GARCH Models

- Estimate the model without GARCH in the usual way.
- Look at the time series properties of the *squared residuals*. Correlogram, AIC, SIC etc.
- $ARMA(1,1)$ in the squared residuals implies $GARCH(1,1)$.

14. Estimation of GARCH Models

Usually use maximum likelihood with the assumption of normal distribution. Maximum likelihood estimation finds the parameter values that maximize the likelihood function.

15. Forecasting GARCH Models

- In financial applications, volatility forecasts are often of *direct interest*.

$$\sigma^2_{t+h,t} = E(\epsilon^2_{t+h} | \Omega_t)$$

- *Better* forecast confidence interval:

$$y_{t+h,t} \pm 1.96\sigma_h \quad \text{versus} \quad y_{t+h,t} \pm 1.96\sigma_{t+h,t}$$

16. Application: Stock Market Volatility

Objective: Model and forecast the volatility of daily returns on the Hang Seng Index (HSI)

Data: Daily returns on the Hang Seng Index (HSI) form January 2, 1987, through December 31, 2012. Excluding holidays, there are 6463 observations. (http://finance.yahoo.com/q?s=%5EHSI)

FIGURE 16.1. Daily Return of Hang Seng Index (1987 - 2012)

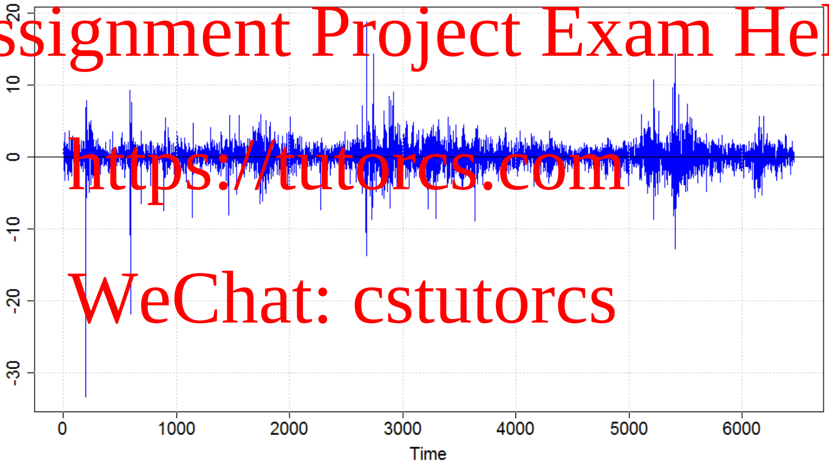


FIGURE 16.2. ACF of Daily Return

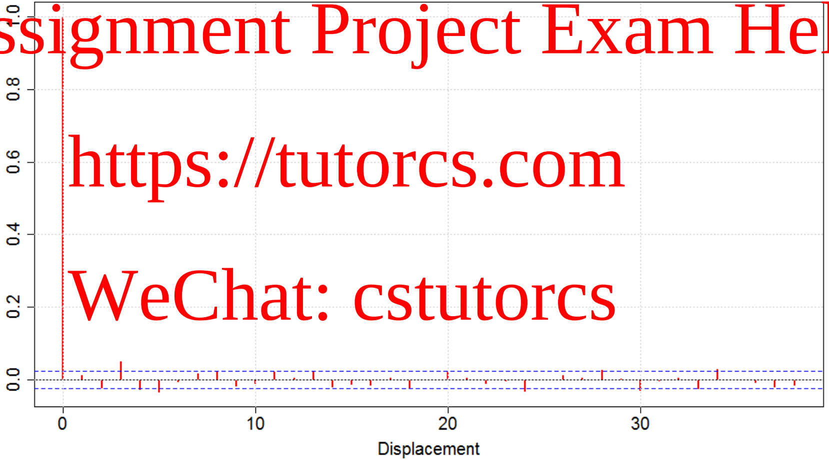


FIGURE 16.3. PACF of Daily Return

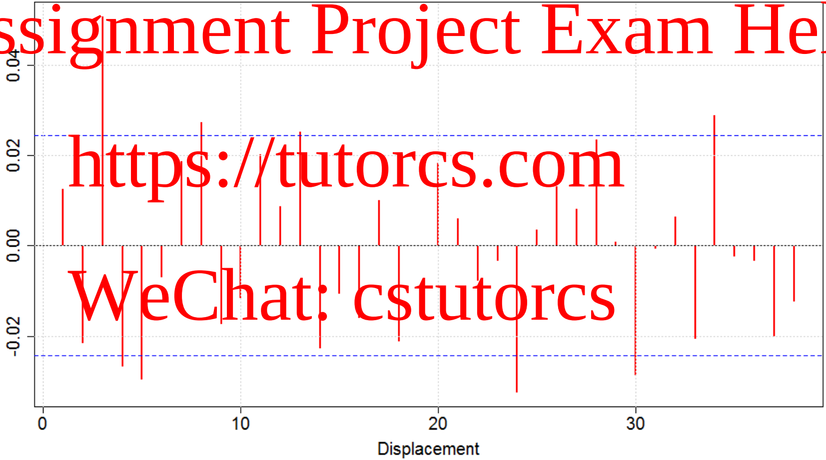


FIGURE 16.4. Histogram of Daily Return

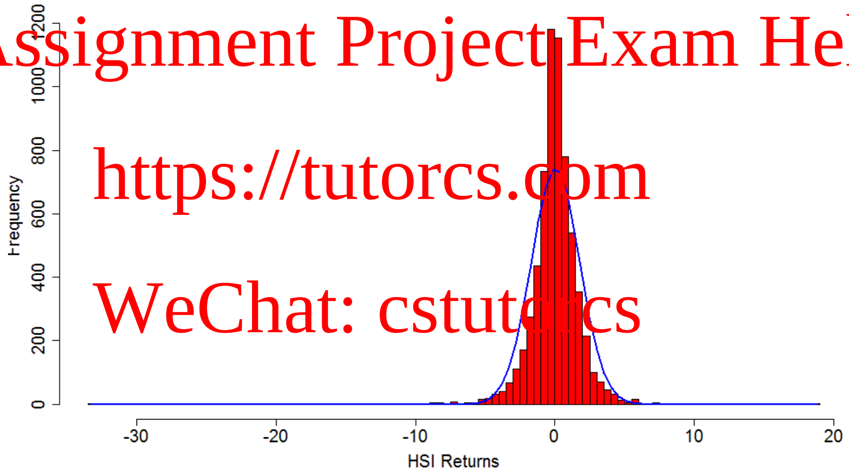


Table 1. D'Agostino skewness test

H_0: Data shows no skewness versus H_1: Data have a skewness

skew	=	-1.2416
z	=	-21.3346
p-value	<	2.2e-16

TABLE 2. Anscombe-Glynn kurtosis test

H_0: Kurtosis equals to 3 versus H_1: Kurtosis is not equal to 3

kurt	=	36.0440
z	=	43.6738
p-value	<	2.2e-16

Table 3. Jarque-Bera Normality Test

H_0: Normal versus H_1: Not Normal

JB	$=$	295655.4
p-value	$<$	2.2e-16

FIGURE 16.5. Squared Daily Return

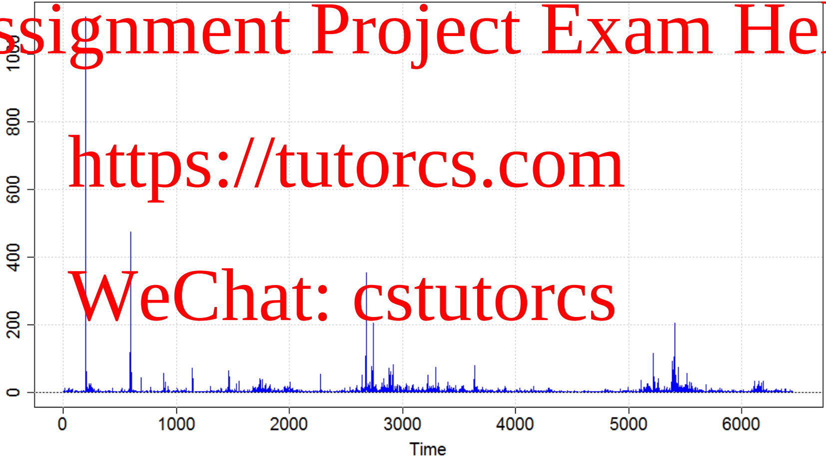


FIGURE 16.6. ACF of Squared Daily Return

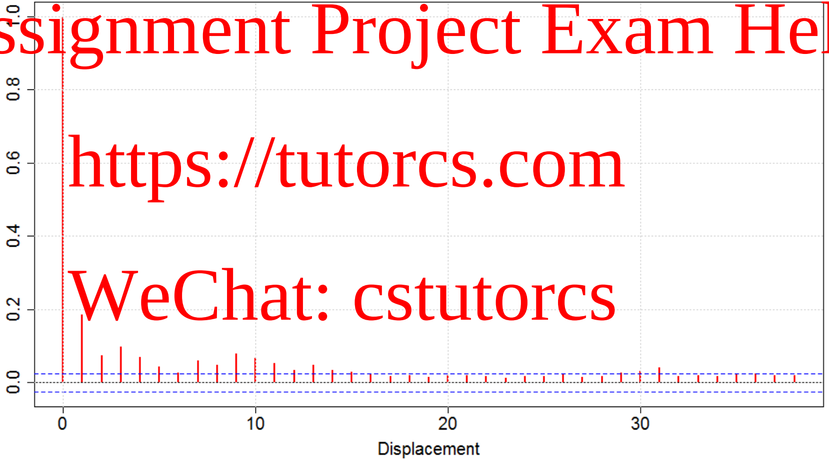


FIGURE 16.7. PACF of Squared Daily Return

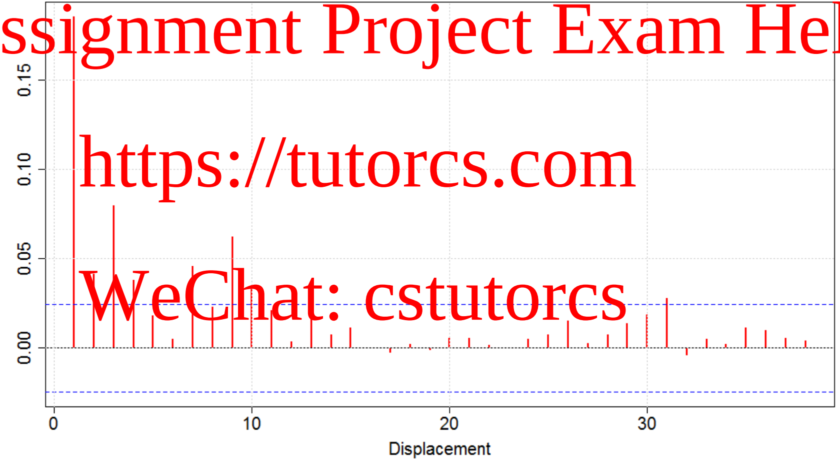
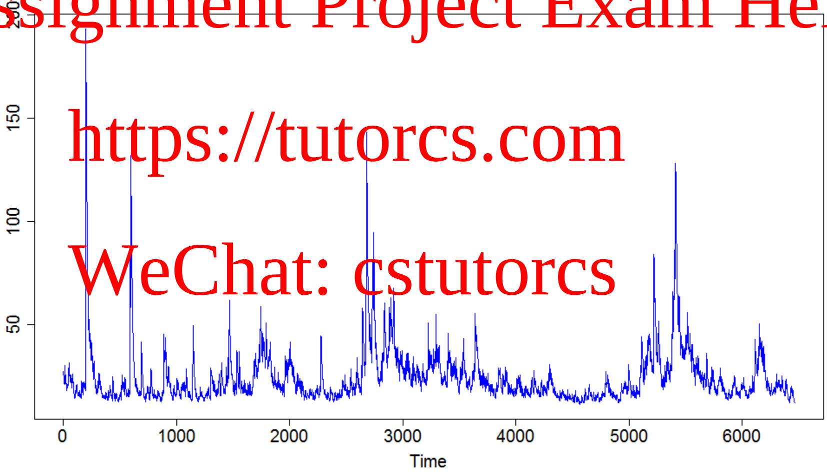


FIGURE 16.8. Estimated Volatility based on GARCH(1,1) of Daily Return



17. R package to estimate GARCH

Several R packages will allow us to estimate and forecast volatility of GARCH models.

- rugarch is often used
- fGarch is a competing alternative