Chapter Four

Conditional Heteroscedastic Models

Tingguo Zheng (zhengtg@gmail.com)
Department of Statistics and Data Science, School of Economics, Xiamen University

(For Undergraduate Program, 2025-2026, Spring Semester)

- *Reference: "Analysis of Financial Time Series (3ed)" by Ruey S. Tsay, 2010. (Chapter 3, Chapter 6, Chapter 7)*

- **Outline:**
  - Motivations
  - ARCH model
  - GARCH model
  - Asymmetric GARCH models
  - Value at risk

1 Motivation

1.1 Why do we need volatility models?

Wold's decomposition theorem states that any covariance stationary $\{y_t\}$ may be written as

$$y_t = \mu_t + u_t,$$

where

$$\mu_t = c_0 + c_1 t, \quad u_t = \theta(L)\varepsilon_t, \quad \theta(L) = \sum_{i=0}^{\infty} \theta_i L^i, \quad \sum_{i=0}^{\infty} \theta_i^2 < \infty, \quad \theta_0 = 1,$$

$$E[\varepsilon_t] = 0, \quad E[\varepsilon_t \varepsilon_\tau] = \begin{cases} \sigma^2 < \infty, & \text{if } t = \tau; \\ 0, & \text{otherwise.} \end{cases}$$

The innovation sequence $\{\varepsilon_t\}$ needs not to be independent.

Now suppose that y_t is a linear stationary process with *IID* innovations,

$$y_t = \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i}.$$

The unconditional mean and variance are

$$E[y_t] = 0, \quad E[y_t^2] = \sigma^2 \sum_{i=0}^{\infty} \theta_i^2.$$

The *conditional mean* is time varying $E[y_t|\mathcal{F}_{t-1}] = \sum_{i=1}^{\infty} \theta_i \varepsilon_{t-i}$.

The *conditional variance* of y_t is constant at

$$E[(y_t - E[y_t|\mathcal{F}_{t-1}])^2|\mathcal{F}_{t-1}] = \sigma^2. \qquad Constant!$$

This model is unable to capture the conditional variance dynamics.

1.2 Martingale difference sequence

Theorem 1 (**Law of Iterated Expectations, LIE).** *In general, for information sets I_t and J_t such that $I_t \subset J_t$ (J_t is the bigger info set). The Law of Iterated Expectations says*

$$E[Y|I_t] = E[E[Y|J_t]|I_t].$$

Definition 1 (Martingale difference sequence). Let $\{\varepsilon_t\}$ be a sequence of random variables with an associated information set I_t. The sequence $\{\varepsilon_t, I_t\}$ is called a *martingale difference sequence (MDS)* if (1) $I_{t-1} \subset I_t$, and (2) $E[\varepsilon_t | I_{t-1}] = 0$ (MDS property).

- The MDS sequence $\{\varepsilon_t\}$ is *uncorrelated* since

$$E[\varepsilon_t \varepsilon_{t-j}] = E[E(\varepsilon_t \mid I_{t-1})\varepsilon_{t-j}] = 0.$$

- If $\{y_t, I_t\}$ is a martingale, a MDS $\{\varepsilon_t, I_t\}$ may be constructed by defining

$$\varepsilon_t = y_t - E[y_t | I_{t-1}].$$

1.3 Random variance models

The econometric models to account for the very specific nature of financial series are generally written as

$$y_t = \mu_t(\theta) + \varepsilon_t, \tag{1}$$

$$\varepsilon_t = \sigma_t(\theta) z_t, \quad z_t \sim i.i.d.(0, 1) \tag{2}$$

where $\sigma_t > 0$,

$$\mu_t(\theta) = E[y_t|\mathcal{F}_{t-1}] := E_{t-1}[y_t],$$

$$\sigma_t^2(\theta) = E[(y_t - \mu_t(\theta))^2|\mathcal{F}_{t-1}] := E_{t-1}[\varepsilon_t^2].$$

The dynamics of the conditional mean $\mu_t(\theta)$ may be an ARMA(p, q) process or could consist of seasonality features.

(1) Model properties

- The process $\{\varepsilon_t\}$ is conditionally serially uncorrelated. Given that

$$E_{t-1}[\varepsilon_t] = E_{t-1}[y_t - \mu_t] = 0, \quad \text{M.D.S.}$$

we have that with the Law of Iterated Expectations:

$$E_{t-h}[\varepsilon_t] = E_{t-h}[E_{t-1}(\varepsilon_t)] = E_{t-h}[0] = 0.$$

This orthogonality property implies

$$\begin{aligned}
\mathrm{Cov}_{t-h}[\varepsilon_t, \varepsilon_{t+k}] &= E_{t-h}[\varepsilon_t \varepsilon_{t+k}] - E_{t-h}[\varepsilon_t] E_{t-h}[\varepsilon_{t+k}] \\
&= E_{t-h}[\varepsilon_t \varepsilon_{t+k}] = E_{t-h}[E_{t+k-1}(\varepsilon_t \varepsilon_{t+k})] \\
&= E_{t-h}\{\varepsilon_t E_{t+k-1}[\varepsilon_{t+k}]\} = 0.
\end{aligned}$$

- ε_t^2 is an unbiased estimator of $\sigma_t^2(\theta)$.
  - Suppose $z_t \sim NID(0,1)$ and independent of $\sigma_t^2(\theta)$ (this is self-evident since $\sigma_t^2(\theta)$ can be seen as a function of past information), we have

$$E_{t-1}[\varepsilon_t^2] = E_{t-1}[\sigma_t^2]E_{t-1}[z_t^2] = E_{t-1}[\sigma_t^2].$$

    because $z_t^2 \mid \mathcal{F}_{t-1} \sim \chi^2(1)$.
  - Moreover, the absolute return $|\varepsilon_t|$ is a biased estimator for σ_t, since if $\varepsilon_t \mid \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$, then $E_{t-1}|\varepsilon_t| = \sigma_t\sqrt{2/\pi}$.
    See *Half-normal distribution* (`https://en.wikipedia.org/wiki/Half-normal_distribution`)

- If the conditional distribution of z_t is time invariant with a finite fourth moment, the fourth moment of ε_t is

$$E[\varepsilon_t^4] = E[z_t^4]E[\sigma_t^4] \geq E[z_t^4]\{E[\sigma_t^2]\}^2 = E[z_t^4]\{E[\varepsilon_t^2]\}^2$$

$$E[\varepsilon_t^4] \geq E[z_t^4]\{E[\varepsilon_t^2]\}^2$$

by Jensen's inequality.[1]

  - If $z_t \sim NID(0,1)$, then $E[z_t^4] = 3$, the unconditional distribution for ε_t is therefore leptokurtic (尖峰态)

$$E[\varepsilon_t^4] \geq 3E[\varepsilon_t^2]^2 \quad \Rightarrow \quad \kappa = E[\varepsilon_t^4]/E[\varepsilon_t^2]^2 \geq 3.$$

[1] Jensen's inequality: $E[g(x)] \leq g(E[x])$ if $g(x)$ is concave (凹); $E[g(x)] \geq g(E[x])$ if $g(x)$ is convex (凸).

- The kurtosis can be expressed as a function of the variability of the conditional variance. In fact, if $\varepsilon_t \mid \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$, then

$$E_{t-1}[\varepsilon_t^4] = 3E_{t-1}[\sigma_t^4] = 3E_{t-1}\left\{[E_{t-1}(\varepsilon_t^2)]^2\right\}$$

$$E[\varepsilon_t^4] = 3E\left\{[E_{t-1}(\varepsilon_t^2)]^2\right\} \geq 3\left\{E[E_{t-1}(\varepsilon_t^2)]\right\}^2 = 3[E(\varepsilon_t^2)]^2$$

and we have

$$E[\varepsilon_t^4] - 3[E(\varepsilon_t^2)]^2 = 3E\left\{[E_{t-1}(\varepsilon_t^2)]^2\right\} - 3\left\{E[E_{t-1}(\varepsilon_t^2)]\right\}^2$$

$$E[\varepsilon_t^4] = 3[E(\varepsilon_t^2)]^2 + 3E\left\{[E_{t-1}(\varepsilon_t^2)]^2\right\} - 3\left\{E[E_{t-1}(\varepsilon_t^2)]\right\}^2$$

Therefore, the kurtosis

$$\kappa = \frac{E[\varepsilon_t^4]}{[E(\varepsilon_t^2)]^2} = 3 + 3\frac{E\left\{[E_{t-1}(\varepsilon_t^2)]^2\right\} - \left\{E[E_{t-1}(\varepsilon_t^2)]\right\}^2}{[E(\varepsilon_t^2)]^2}$$

$$= 3 + 3\frac{\mathrm{Var}\{E_{t-1}[\varepsilon_t^2]\}}{[E(\varepsilon_t^2)]^2} = 3 + 3\frac{\mathrm{Var}[\sigma_t^2]}{[E(\varepsilon_t^2)]^2}.$$

(2) Types of models

- *Conditionally heteroscedastic* (or GARCH-type) process.
  - σ_t^2 is generated by the past of ε_t.
  - The volatility is here a deterministic function of past of ε_t.
  - Different choice of this deterministic function produces different processes of this GARCH class.
  - The standard GARCH models are characterized by a volatility specified as a linear function of the past values of ε_t^2,

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim iid(0,1), \quad \sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

They will be studies in detail in our course.

- *Stochastic volatility processes*
  - $\{\varepsilon_t\}$ is generated by $\{v_t\}$, where v_t is a strong white noise and is independent of ε_t.
  - In these models, volatility is a latent process.
  - The most popular model in this class assumes that the process $\log \sigma_t$ or $\log \sigma_t^2$ follows an AR(1) of the form

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim iid(0,1),$$

$$\log \sigma_t = \omega + \phi \log \sigma_{t-1} + v_t$$

$$\text{or} \quad \log \sigma_t^2 = \omega + \phi \log \sigma_{t-1}^2 + v_t,$$

where the noise v_t and ε_t are independent.

- *Regime switching models*
  - The conditional variance is specified as $\sigma_t = \sigma(\Delta_t)$, where Δ_t is a latent (unobservable) integer-valued process, independent of z_t.
  - The state of the variable Δ_t is here interpreted as a regime.
  - The process Δ_t is generally supposed to be a finite-state Markov chain,

$$P = \begin{pmatrix} p_{11} & \cdots & p_{M1} \\ \vdots & \ddots & \vdots \\ p_{1M} & \cdots & p_{MM} \end{pmatrix},$$

    where $p_{ij} = \Pr[\Delta_t = j \mid \Delta_{t-1} = i]$ denotes the transition probability from the state $\Delta_{t-1} = i$ to the state $\Delta_t = j$, and it satisfies that $\sum_{j=1}^{M} p_{ij} = 1$, $i = 1, \ldots, M$.

2 The ARCH model

2.1 ARCH(q) model

An ARCH(q) model, originally introduced by Engle (1982), is a linear function of past squared disturbances:

$$\varepsilon_t = \sigma_t z_t, \qquad z_t \sim i.i.d.(0,1), \tag{3}$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2. \tag{4}$$

- The conditional volatility is a moving average of squared innovations.
- The parameter constraints: $\omega > 0$, and $\alpha_i \geq 0$, $i = 1, \cdots, q$.
- In practice, z_t is often assumed to follow the standard normal or a standardized Student-t or a generalized error distribution.

- Defining $v_t = \varepsilon_t^2 - \sigma_t^2$, where $E_{t-1}[v_t] = 0$, we can rewrite (4) as an AR(q) model for the squared innovation ε_t^2:

$$\varepsilon_t^2 = \omega + \alpha(L)\varepsilon_t^2 + v_t,$$

  where $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_q L^q$.

- The process ε_t is covariance stationary if and only if the sum of the autoregressive parameters is less than one, i.e., $\sum_{i=1}^q \alpha_i < 1$.

- In this case, the unconditional variance of innovation is

$$\sigma^2 = E[\varepsilon_t^2] = \omega/(1 - \alpha_1 - \cdots - \alpha_q).$$

- Given the stationary condition that $\sum_{i=1}^q \alpha_i < 1$, the process of $\{\varepsilon_t\}$ is a white noise.

2.2 Kurtosis of an ARCH model

ARCH models are able to generate *excess kurtosis*. Indeed, even if the *standardized innovations* z_t is assumed to be normal, the *unconditional distribution* for ε_t has fatter tails than the normal distribution.

- For instance, consider an ARCH(1) model

$$\sigma_t^2(\theta) = \omega + \alpha_1 \varepsilon_{t-1}^2,$$

  where $\varepsilon_t = \sigma_t z_t$.

- The unconditional variance is given by $\sigma^2 = \omega/(1 - \alpha_1)$ so that we should have $0 < \alpha_1 < 1$ for σ^2 to be finite.

- Under normality, the *conditional fourth moment* is given by

$$E_{t-1}[\varepsilon_t^4] = E_{t-1}[\sigma_t^4 z_t^4] = 3E_{t-1}[\sigma_t^4] = 3\sigma_t^4 = 3(\omega + \alpha_1 \varepsilon_{t-1}^2)^2,$$

  whereas the *unconditional fourth moment* is

$$\begin{aligned} m_4 = E[\varepsilon_t^4] &= E[E_{t-1}[\sigma_t^4 z_t^4]] = 3E[(\omega + \alpha_1 \varepsilon_{t-1}^2)^2] \\ &= 3\left(\omega^2 + 2\frac{\omega^2 \alpha_1}{1-\alpha_1} + \alpha_1^2 m_4\right) = \frac{3\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}. \end{aligned}$$

- Since the fourth moment of ε_t is positive, we must have that $\alpha_1^2 < 1/3$.

- The *unconditional kurtosis* is

$$\kappa = \frac{m_4}{\sigma^4} = \frac{3\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \times \frac{(1-\alpha_1)^2}{\omega^2} = 3\,\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3,$$

for $\alpha_1 > 0$ and $3\alpha_1^2 < 1$.[2]

- So, the excess kurtosis is always positive and the tails of the distribution of ε_t are fatter than that of a normal distribution, even if the conditional distribution is normal.

[2]For $3\alpha_1^2 \geq 1$, the kurtosis becomes infinite.

2.3 Testing for ARCH effects

The squared series ε_t^2 is used to check for conditional heteroscedasticity, which is also known as the ARCH effects. Two tests are available.

- Ljung-Box statistics $Q(m)$ to the $\{\varepsilon_t^2\}$, see McLeod and Li (1983)
  - The null hypothesis is that the first m lags of ACF of $\{\varepsilon_t^2\}$ are zero.
- The Lagrange-Multiplier (LM) test proposed by Engle (1982)
  - This test is equivalent to the usual F statistic for testing $\alpha_i = 0$ $(i = 1, \ldots, m)$ in the linear regression

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2 + e_t, \quad t = m+1, \ldots, T,$$

where e_t denotes the error term.

– The null hypothesis is H_0: $\alpha_1 = \cdots = \alpha_m = 0$ and the alternative hypothesis is $H_a : \alpha_1 > 0$ or $\cdots$ or $\alpha_m > 0$.

– The test statistic is

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T-2m-1)} \quad \sim \quad F(m, T-2m-1),$$

where $SSR_0 = \sum_{t=m+1}^{T}(\varepsilon_t^2 - \bar{\omega})^2$, $\bar{\omega}$ is the sample mean of ε_t^2, and $SSR_1 = \sum_{t=m+1}^{T} \hat{e}_t^2$, $\hat{e}_t$ is the least-squares residual of the prior linear regression.

– The LM test can also be computed as

$$(T-2m-1) \times R^2 \quad \sim \quad \chi^2(m),$$

where R^2 is obtained from a regression.

2.4 Maximum Likelihood Estimation

Data series: Suppose a time series $\{\varepsilon_T, \cdots, \varepsilon_1, \varepsilon_0, \cdots, \varepsilon_{-m+1}\}$, where m is the number of observations for initializing the process. $m = \max(p, q)$ if GARCH model; $m = q$ if ARCH model.

Information set: Denote by $\mathcal{F}_t = \{y_t, \cdots, y_1, y_0, \cdots, y_{-m+1}\}$ is the available information until time t.

Likelihood function: Under the normality assumption, the likelihood function of a ARCH(q) or GARCH(p, q) model is

$$\begin{aligned} f(\varepsilon_1, \cdots, \varepsilon_T|\mathcal{F}_0) &= f(\varepsilon_T|\mathcal{F}_{T-1})f(\varepsilon_{T-1}|\mathcal{F}_{T-2})\cdots f(\varepsilon_2|\mathcal{F}_1)f(\varepsilon_1|\mathcal{F}_0) \\ &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2(\theta)}{2\sigma_t^2(\theta)}\right). \end{aligned}$$

MLE: The ML estimator is obtained by maximizing this expression or, equivalently, the *log-likelihood function*

$$L_T(\theta|\mathcal{F}_T) = \sum_{t=1}^{T} l_t(\theta) = \sum_{t=1}^{T} \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2} \right],$$

where $l_t(\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2}$, and σ_t^2 can be evaluated recursively.

CAN: As shown by Weiss (1986), provided standardized innovations have finite fourth moments, the ML estimator is *consistent* and has the following asymptotic distribution

$$\sqrt{T}(\hat{\theta} - \theta) \Rightarrow \mathcal{N}(0, I(\theta)^{-1}).$$

In the sample analogue, a consistent estimate of the asymptotic covariance matrix can be estimated by

$$\frac{1}{T}\hat{I}(\hat{\theta})^{-1} = \frac{1}{T}\left(\sum_{t=1}^{T} \frac{\partial l_t(\hat{\theta})}{\partial \theta} \frac{\partial l_t(\hat{\theta})}{\partial \theta'}\right)^{-1}.$$

The maximum likelihood estimation of the parameters is often implemented by the Berndt, Hall, Hall and Hausman (1974) algorithm.

QMLE: When ε_t is non-Gaussian, using above ML estimator under the normal assumption is called the Quasi-ML estimator. The QMLE is also *consistent and asympotically normal (CAN)*.

2.5 Non-Gaussian Error Distributions

(1) Student's t Distribution – Bollerslev (1987)

The Student's t Distribution is the first attempt to capture the excess kurtosis of GARCH standardized residuals.

If a Student's t distribution with ν degrees of freedom is assumed, the probability density function (pdf) of ε_t takes the form

$$t(\varepsilon_t \mid \nu, \sigma_t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi(\nu-2)\sigma_t^2}} \left[1 + \frac{\varepsilon_t^2}{\sigma_t^2(\nu-2)}\right]^{-\frac{\nu+1}{2}}$$

where $\Gamma(\cdot)$ is the Gamma function and $\nu > 2$ (finite second moments) is the degree-of-freedom (or shape) parameter.

We also have $z_t = \varepsilon_t/\sigma_t$ and so $t(z_t \mid \nu) = t(\frac{\varepsilon_t}{\sigma_t} \mid \nu, \sigma_t)/\sigma_t$.

(2) Generalized Error Distribution – Nelson (1991).

If a random variable ε_t has a GED distribution with zero mean and variance σ_t^2, the pdf of the innovations is given by

$$g(\varepsilon_t|\nu) = \frac{\nu \exp\left[-\frac{1}{2}\left|\frac{\varepsilon_t}{\lambda\sigma_t}\right|^{\nu}\right]}{\sigma_t \lambda 2^{(1+\frac{1}{\nu})}\Gamma(\frac{1}{\nu})}, \quad \text{and} \quad \lambda \equiv [(2^{-2/\nu}\Gamma(1/\nu))/\Gamma(3/\nu)]^{1/2},$$

where ν $(0 < \nu \leq \infty)$ is the thickness-of-tail (or shape) parameter

- When $\nu = 2$, the GED becomes a standard normal distribution;
- When $\nu < 2$ and $\nu > 2$ the distribution has thicker and thinner tails than the normal respectively.

We also have $z_t = \varepsilon_t/\sigma_t$ and so $g(z_t \mid \nu) = g(\frac{\varepsilon_t}{\sigma_t} \mid \nu, \sigma_t)/\sigma_t$.

(3a) Skewed Student t Distribution – Bruce Hansen (1994)

The pdf of $z_t = \varepsilon_t/\sigma_t$ (ignoring the time t) is given by

$$g(z \mid \nu, \lambda) = bc\left(1 + \frac{\zeta^2}{\nu - 2}\right)^{-\frac{\nu+1}{2}},$$

where

$$\zeta = \begin{cases} (bz + a)/(1 - \lambda), & \text{if } z < -a/b; \\ (bz + a)/(1 + \lambda), & \text{if } z \geq -a/b. \end{cases}$$

The constant terms a and b are defined as

$$a = 4\lambda c \frac{\nu - 2}{\nu - 1}, \quad b^2 = 1 + 3\lambda^2 - a^2, \quad c = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu - 2)}\Gamma(\frac{\nu}{2})}.$$

Here, ν denotes the *degree of freedom parameter* and λ the *asymmetric parameter*. The density is defined for $2 < \nu < \infty$ and $-1 < \lambda < 1$.

Furthermore, it encompasses a large set of conventional densities. For instance,

- if $\lambda = 0$, Hansen's distribution is reduced to the traditional Student t distribution, without asymmetric.

- if $\nu = \infty$, the Student t distribution collapses to the normal density.

With the skewed t distribution, an empirical facts that returns is left-skewed can be found, i.e., $\lambda < 0$.

(3b) Skewed Student t Distribution – Lambert and Laurent (2001)

The resulting probability density function of this standardized skew-Student-t distribution is

$$g(z_t \mid \xi, \nu) = \begin{cases} \frac{2}{\xi + \frac{1}{\xi}} \varrho f[\xi(\varrho z_t + \bar{\omega}) \mid \nu], & \text{if } z < -\bar{\omega}/\varrho \\ \frac{2}{\xi + \frac{1}{\xi}} \varrho f[\frac{1}{\xi}(\varrho z_t + \bar{\omega}) \mid \nu], & \text{otherwise.} \end{cases}$$

where $f(\cdot)$ is the probability density function (pdf) of the standardized Student-t distribution, ξ is the skewness parameter, $\nu > 2$ is the degrees of freedom, and the parameters ϱ and $\bar{\omega}$ are given below:

$$\bar{\omega} = \frac{\Gamma[(\nu-1)/2]\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\nu/2)} \left(\xi - \frac{1}{\xi}\right), \qquad \varrho^2 = (\xi^2 + \frac{1}{\xi^2} - 1) - \bar{\omega}^2.$$

2.6 Building an ARCH Model

Among volatility models, specifying an ARCH model is relatively easy.

- $\surd$Modeling the mean effect and testing for ARCH effects.
  - standard ARIMA modelling of the mean effect;
    - No autocorrelations in residuals $\{\hat{\varepsilon}_t\}$.
    - Normal is OK.
  - test dependency of the squared residuals. H_o : no ARCH effects, v.s. H_a : ARCH effects.

- Order determination of the ARCH part of the model. Use PACF of the squared residuals or the information criteria.

- Estimation: Conditional MLE or Quasi MLE;

- Model checking:
  - ACF, PACF of the standardized residuals and the squared standardized residuals;
  - Skewness and kurtosis of the standardized residuals;
  - Q-stat of standardized residuals and squared standardized residuals.
- Software: Many packages available, e.g., OxMetrics, Eviews, Matlab, and R. We use OxMetrics Ox@GARCH.

Model checking

Let $\hat{\varepsilon}_t$ be the residual of the mean equation of an ARMA(p, q) model,

$$\hat{\varepsilon}_t = y_t - \sum_{i=1}^{p} \hat{\phi}_i y_{t-i} - \sum_{j=1}^{q} \hat{\theta}_j \hat{\varepsilon}_{t-j},$$

Let $\hat{\sigma}_t^2 = \hat{\omega} + \sum_{j=1}^{q} \hat{\alpha}_j^q \hat{\varepsilon}_{t-j}^2$ be the estimated conditional variance. Define the standardized residual as

$$\hat{e}_t = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_t}.$$

The ACF of $\hat{e}_t$ can be used to check the adequacy of the mean equation of y_t, i.e. the ARMA model, and the ACF of the squared residuals $\hat{e}_t^2$ can be used to check the adequacy of the volatility equation.

2.7 Model selection with information criterion

We use two types of information criterion: the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). In AIC and BIC, we choose the model that has the minimum value of

$$AIC = -2\log(L) + 2m, \tag{5}$$

$$BIC = -2\log(L) + m\log T, \tag{6}$$

where

- L is the likelihood of the data with a certain model;
- T is the number of observations and
- m is the number of parameters in the model.

2.8 Forecasting Volatility

Forecasts of the ARCH model can be obtained recursively.

- The 1-step ahead forecast for σ_{t+1}^2 given I_t is

$$\sigma_t^2(1) = \omega + \alpha_1 \varepsilon_t^2 + \cdots + \alpha_q \hat{\varepsilon}_{t+1-q}^2.$$

- The 2-step ahead forecast is

$$\sigma_t^2(2) = \omega + \alpha_1 \sigma_t^2(1) + \alpha_2 \hat{\varepsilon}_t^2 + \cdots + \alpha_q \hat{\varepsilon}_{t+2-q}^2.$$

- The h-step ahead forecast for σ_{t+h}^2 is

$$\sigma_t^2(h) = \omega + \alpha_1 \hat{\sigma}_t^2(h-1) + \cdots + \alpha_p \hat{\sigma}_t^2(h-q) = \omega + \sum_{i=1}^{q} \alpha_i \hat{\sigma}_t^2(h-i),$$

with $\sigma_t^2(\ell) = \hat{\varepsilon}_{t+\ell}^2$ if $\ell \leq 0$.

Example 1. (Build a simple ARCH model for the monthly log returns of Intel stock)

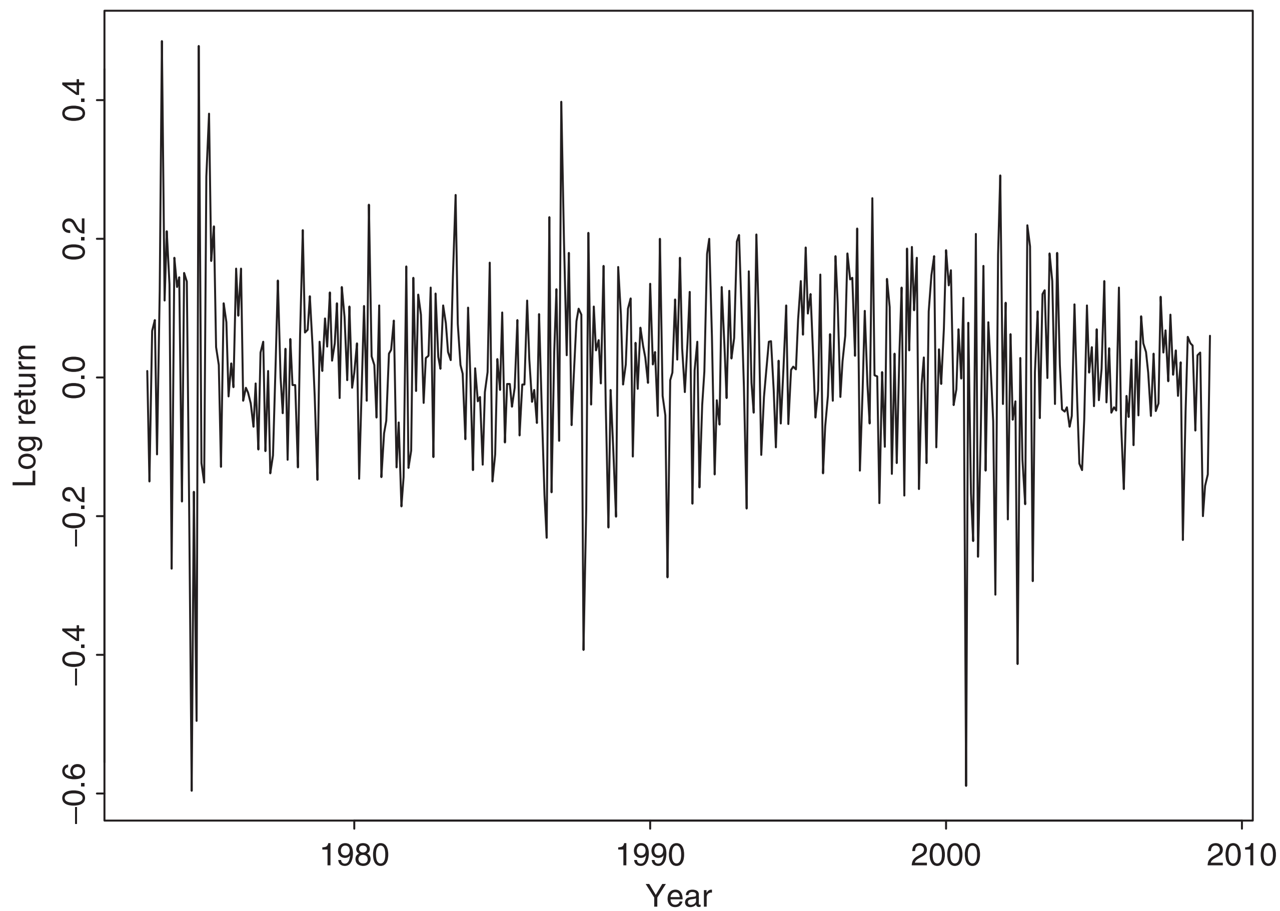

Figure 1 Time plot of monthly log returns of Intel stock from January 1973 to December 2008.

- The SACF and SPACF of the squared returns clearly show the existence of conditional heteroscedasticity.

- The SPACF in Figure 2(d) indicates that an ARCH(3) model might be appropriate.

- The model is specified as

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \alpha_3 \varepsilon_{t-3}^2$$

- Assuming that z_t are iid standard normal, we obtain the fitted model

$$r_t = \underset{(0.0057)}{0.0122} + \varepsilon_t, \quad \sigma_t^2 = \underset{(0.0010)}{0.0106} + \underset{(0.0757)}{0.2131}\, \varepsilon_{t-1}^2 + \underset{(0.0480)}{0.0770}\, \varepsilon_{t-2}^2 + \underset{(0.0688)}{0.0599}\, \varepsilon_{t-3}^2$$

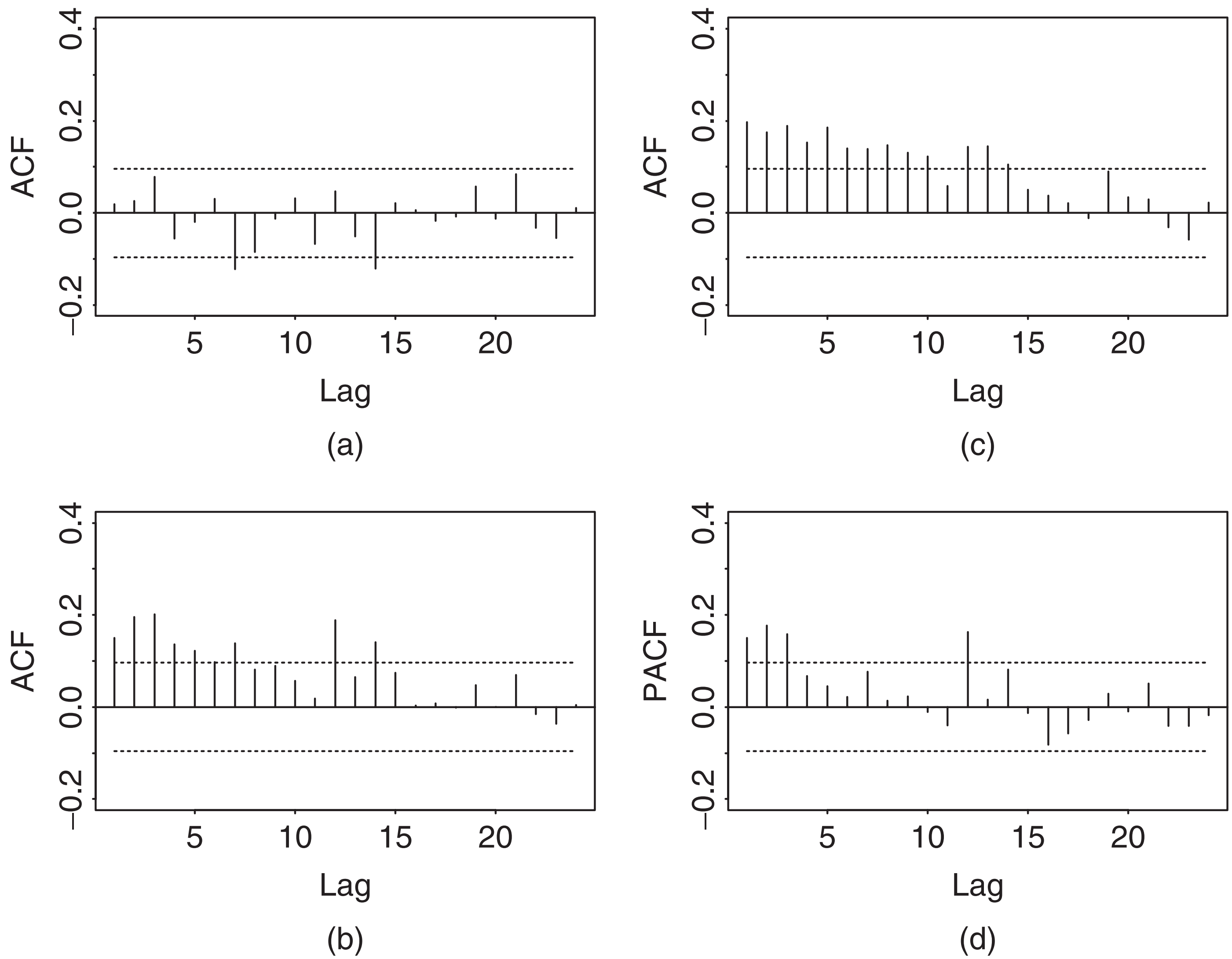

Figure 2 Sample ACF and PACF of various functions of monthly log stock returns of Intel Corporation from January 1973 to December 2008: (a) ACF of the log returns, (b) ACF of the squared log returns, (c) ACF of the absolute log returns, and (d) PACF of the squared log returns.

- The estimates of α_2 and α_3 appear to be statistically nonsignificant at the 5% level. Therefore, the model can be simplified.

- Dropping the two nonsignificant parameters, we obtain the model

$$r_t = \underset{(0.0053)}{0.0126} + \varepsilon_t, \quad \sigma_t^2 = \underset{(0.0010)}{0.0111} + \underset{(0.0761)}{0.3560}\, \varepsilon_{t-1}^2.$$

- The Ljung – Box statistics of standardized residuals $\{\hat{z}_t\}$ give $Q(10) = 12.64$ with a p value of 0.24 and those of $\{\hat{z}_t^2\}$ give $Q(10) = 14.75$ with a p value of 0.14.

- Consequently, the ARCH(1) model is adequate for describing the conditional heteroscedasticity of the data at the 5% significance level.

3 The GARCH model

Due to the large persistence in volatility, the ARCH model often requires a large p to fit the data. In such cases, it is more parsimonious to use the GARCH (Generalized ARCH) model proposed by Bollerslev (1986).

Table 1 Information criteria

ARCH(q) model				GARCH(p, q) model				
q	BIC	AIC	HQ	$q \setminus p$	1	2	3	4
1	3.2312	3.2222	3.2258			AIC		
2	3.0894	3.0793	3.0829	1	2.8713	2.8638	2.8638	2.8643
3	3.0229	3.0112	3.0154	2	2.8677	**2.8637**	2.8642	2.8648
4	2.9706	2.9571	2.9619	3	2.8675	2.8641	2.8636	2.8638
5	2.9398	2.9247	2.9301	4	2.8667	2.8638	2.8641	2.8628
6	2.9298	2.9130	2.9189			BIC		
7	2.9190	2.9006	2.9072	1	2.8814	**2.8756**	2.8772	2.8794
8	2.9066	2.8865	2.8937	2	2.8795	2.8770	2.8793	2.8815
9	2.9030	2.8795	2.8879	3	2.8809	2.8792	2.8804	2.8822
10	2.9005	2.8753	2.8843	4	2.8818	2.8806	2.8826	2.8830
11	**2.9021**	2.8753	**2.8849**					
12	2.9042	**2.8757**	2.8859					

3.1 GARCH Model

Definition 2 (GARCH(p,q) process)**.** The process ε_t is called a (strong) GARCH(p,q) (with respect to the sequence z_t) if

$$\begin{cases} \varepsilon_t & = \sigma_t z_t, \quad z_t \sim i.i.d.(0,1) \\ \sigma_t^2 & = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \end{cases} \tag{7}$$

where the α_i and β_j are nonnegative constants and ω is a (strictly) positive constant.

Remark 1. The parameters $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, q$ and $\beta_i \geq 0$, $i = 1, \ldots, p$.

Remark 2. If $E[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$ and $\sigma_t^2 = E[\varepsilon_t^2 \mid \mathcal{F}_{t-1}] = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$, the $\{\varepsilon_t\}$ is a *semi-strong* GARCH process.

In our course, we mainly focus on the Strong-GARCH case.

In lag polynomial form, the conditional variance of a GARCH(p, q) process can be rewritten as:

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2, \tag{8}$$

where $\alpha(L) = \alpha_1 L + \cdots + \alpha_q L^q$, $\beta(L) = \beta_1 L + \cdots + \beta_p L^p$.

The GARCH(1,1) is the most popular model in the empirical literature:

$$\sigma_t^2 = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2. \tag{9}$$

The Yule-Walker equations for the squared process

In the GARCH(p, q) model the process $\{\varepsilon_t^2\}$ has an ARMA(m, p) representations,

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{m}(\alpha_j + \beta_j)\varepsilon_{t-j}^2 + \left(v_t - \sum_{i=1}^{p}\beta_i v_{t-i}\right),$$

where $m = \max(p, q)$, $v_t \equiv \varepsilon_t^2 - \sigma_t^2$ (m.d.s. and uncorrelated), $E_{t-1}[v_t] = 0$, and $v_t \in [-\sigma_t^2, \infty]$.

Remark 3. $\{v_t\}$ is a white noise with finite variance under the assumption that $\{\varepsilon_t\}$ has a finite fourth moments.

$$\begin{aligned} E_{t-1}(v_t^2) &= E_{t-1}[\sigma_t^4(z_t^2 - 1)^2] = \sigma_t^4 E_{t-1}[(z_t^2 - 1)^2] = 2\sigma_t^4 \\ &= 2[E_{t-1}(\varepsilon_t^2)]^2 \leq 2E_{t-1}(\varepsilon_t^4) \end{aligned}$$

Under this framework, we can apply the results of ARMA model.

- **Autocovariance function:** The sequence of autocovariances satisfy a linear difference equation of order $m = \max(p, q)$, for $j \geq q+1$

$$\begin{aligned}\gamma^2(j) = \mathrm{Cov}(\varepsilon_t^2, \varepsilon_{t-j}^2) &= (\alpha_1+\beta_1)\gamma^2(j-1) + \cdots + (\alpha_m+\beta_m)\gamma^2(j-m) \\ &= \sum_{k=1}^{m}(\alpha_k+\beta_k)\gamma^2(j-k).\end{aligned}$$

- **Autocorrelation function:** from the results of autocovariance function, the correlation coefficient ρ_j satisfies that for $j \geq q+1$

$$\rho_j = \mathrm{corr}(\varepsilon_t^2, \varepsilon_{t-j}^2) = (\alpha_1+\beta_1)\rho_{j-1} + \cdots + (\alpha_m+\beta_m)\rho_{j-m},$$

where $m = \max(p, q)$.

3.2 Stationarity

(1) Weak stationarity

Necessity: If the process $\{\varepsilon_t\}$ is covariance stationary, it should be satisfy that $\gamma_j = E(\varepsilon_t \varepsilon_{t-j}) = 0$ (already satisfied) and

$$\gamma_0 = \mathrm{Var}[\varepsilon_t] = E[\sigma_t^2] = E(\varepsilon_t^2) = \sigma^2.$$

Theorem 2 (*Covariance stationarity*)**.** *A process $\{\varepsilon_t\}$ which satisfies a GARCH(p, q) model with positive coefficient $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, q$, $\beta_j \geq 0$, $j = 1, \ldots, p$ is covariance stationary if and only if:*

$$\alpha(1) + \beta(1) = \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1.$$

- Supposing that $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$, the unconditional variance is give by

$$\sigma^2 = E[\varepsilon_t^2] = E[\sigma_t^2] = \omega/(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j).$$

- With this covariance stationary condition that $\alpha(1) + \beta(1) < 1$, the process $\{\varepsilon_t\}$ is a white noise.

(2) Strict stationarity

It should be noted that the condition $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ is a sufficient but non necessary conditions for strict stationarity.

In general for GARCH processes, the conditions for covariance stationarity are often more stringent than the conditions for strict stationarity.

Example 2 (Strict stationarity for GARCH(1,1)). A GARCH(1,1) process

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \omega + \sigma_{t-1}^2(\alpha_1 z_{t-1}^2 + \beta_1) =: A_t \sigma_{t-1}^2 + B_t, \tag{10}$$

where $A_t = (\alpha_1 z_{t-1}^2 + \beta_1)$ and $B_t = \omega$. It can be easily seen that $\{A_t, B_t\}$ is an iid sequence. (10) is called a random recurrence equation. We have

$$\begin{aligned}
\sigma_t^2 &= A_t \sigma_{t-1}^2 + B_t \\
&= A_t A_{t-1} \sigma_{t-2}^2 + A_t B_{t-1} + B_t \\
&\vdots \\
&= A_t \cdots A_{t-k} \sigma_{t-k-1}^2 + \sum_{i=0}^{k} A_t \cdots A_{t-i+1} \underbrace{B_{t-i}}_{=\omega}.
\end{aligned} \tag{11}$$

Suppose $\gamma := E \log A_1 = \cdots = E \log A_t < 0$ (since z_t is iid), we have

$$\frac{1}{k} \underbrace{(\log A_t + \log A_{t-1} + \cdots + \log A_{t-k+1})}_{\text{random walk!}} \xrightarrow{k \to \infty} E \log A_1 \quad a.s.$$

For all ϵ, $\exists n_0(\epsilon)$ such that

$$\frac{1}{k}\log(A_t \cdots A_{t-k+1}) \leq \gamma/2 < 0, \quad \forall k \geq n_0(\epsilon)$$
$$\Rightarrow |A_t \cdots A_{t-k+1}| \leq e^{k\gamma/2}, \quad \forall k \geq n_0(\epsilon)$$

Hence (12) converges almost surely, i.e.,

$$\sigma_t^2 \to \omega \sum_{i=0}^{\infty} A_t \cdots A_{t-i+1} \tag{12}$$
$$\leq \omega \sum_{i=k}^{\infty} e^{k\gamma/2} + \omega \sum_{i=0}^{k-1} A_t \cdots A_{t-i+1}$$
$$= \frac{\omega e^{k\gamma/2}}{1-e^{\gamma/2}} + \omega \sum_{i=0}^{k-1} A_t \cdots A_{t-i+1} < \infty.$$

Theorem 3 (Nelson, 1990). *If* $\gamma := E[\log(\beta_1 + \alpha_1 z_t^2)] < 0$, *then GARCH(1,1) process with* $\omega > 0$ *has a strictly stationary solution, i.e.,* $\sigma_t^2 < \infty$ *a.s.and* $\{\varepsilon_t, \sigma_t^2\}$ *is strictly stationary. Its marginal distribution is given by*

$$\omega \sum_{i=0}^{\infty} (\alpha_1 z_{-1}^2 + \beta_1) \cdots (\alpha_1 z_{-i}^2 + \beta_1).$$

Remark 4. Under the strictly stationary condition, the representation for σ_t^2 at each time t is

$$\sigma_t^2 \equiv \omega \sum_{i=0}^{\infty} A_t \cdots A_{t-i+1}$$

and thus

$$\varepsilon_t = \left(\omega \sum_{i=0}^{\infty} A_t \cdots A_{t-i+1} \right)^{\frac{1}{2}} z_t$$

always exists and has the same distribution.

A special strictly stationary condition when $\alpha_1 + \beta_1 = 1$

Since the following result (using the Jensen's Inequality)

$$E[\ln(\beta_1 + \alpha_1 z_t^2)] \le \ln[E(\beta_1 + \alpha_1 z_t^2)] = \ln(\alpha_1 + \beta_1),$$

when $\alpha_1 + \beta_1 = 1$, the model is strictly stationary. Hence, $E[\ln(\beta_1 + \alpha_1 z_t^2)] < 0$ is a weaker requirement than $\alpha_1 + \beta_1 < 1$.

Example 3. For the ARCH(1) model with $\alpha_1 = 1$, $\beta_1 = 0$, $z_t \sim N(0,1)$, we have

$$E[\ln(z_t^2)] \le \ln[E(z_t^2)] = \ln(1) = 0.$$

It's strictly but not covariance stationary. The ARCH(q) is covariance stationary if and only if the sum of the positive parameters is less than one.

3.3 Forecasting volatitlity

The GARCH(p, q) model can be rewritten as

$$\sigma_t^2 = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = \omega + \sum_{j=1}^{m} (\beta_j \sigma_{t-j}^2 + \alpha_j \varepsilon_{t-j}^2)$$

where $m = \max(p, q)$. If $p < q$, $\beta_j = 0$ for $j = p + 1, \ldots, m$; otherwise if $q < p$, $\alpha_j = 0$ for $j = q + 1, \ldots, m$.

Forecasting with a GARCH(p, q) (Engle and Bollerslev, 1986):

$$\sigma_{t+h}^2 =$$

$$\begin{cases} \omega + \sum_{i=1}^{h-1} [\alpha_i \varepsilon_{t+h-i}^2 + \beta_i \sigma_{t+h-i}^2] + \sum_{i=h}^{m} [\alpha_i \varepsilon_{t+h-i}^2 + \beta_i \sigma_{t+h-i}^2], & \text{if } h \leq m; \\ \omega + \sum_{i=1}^{m} [\alpha_i \varepsilon_{t+h-i}^2 + \beta_i \sigma_{t+h-i}^2], & \text{if } h > m. \end{cases}$$

Given the estimated parameters and using the result that $\sigma_t^2(\ell) = E_t[\sigma_{t+\ell}^2] = E_t[\varepsilon_{t+\ell}^2]$ for $\ell > 0$, thus we get

$$\sigma_t^2(h) = E_t[\sigma_{t+h}^2] =$$
$$\begin{cases} \hat{\omega} + \sum_{i=1}^{h-1}[(\hat{\alpha}_i + \hat{\beta}_i)\sigma_t^2(h-i)] + \sum_{i=h}^{m}[\hat{\alpha}_i\hat{\varepsilon}_{t+h-i}^2 + \hat{\beta}_i\hat{\sigma}_{t+h-i}^2], & \text{if } h \le m; \\ \hat{\omega} + \sum_{i=1}^{m}[(\hat{\alpha}_i + \hat{\beta}_i)\sigma_t^2(h-i)], & \text{if } h > m. \end{cases}$$

In particular for a GARCH(1,1) and $h > 2$:

$$
\begin{aligned}
\sigma_t^2(h) &= \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1)\sigma_t^2(h-1) \\
&= \sum_{i=0}^{h-2}(\hat{\alpha}_1 + \hat{\beta}_1)^i \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1)^{h-1}\sigma_t^2(1) \\
&= \hat{\omega}\frac{1-(\hat{\alpha}_1 + \hat{\beta}_1)^{h-1}}{1-(\hat{\alpha}_1 + \hat{\beta}_1)} + (\hat{\alpha}_1 + \hat{\beta}_1)^{h-1}\sigma_t^2(1) \\
&= \hat{\sigma}^2[1-(\hat{\alpha}_1 + \hat{\beta}_1)^{h-1}] + (\hat{\alpha}_1 + \hat{\beta}_1)^{h-1}\sigma_t^2(1) \\
&= \hat{\sigma}^2 + (\hat{\alpha}_1 + \hat{\beta}_1)^{h-1}[\sigma_t^2(1) - \hat{\sigma}^2],
\end{aligned}
$$

where $\hat{\sigma}^2 = \hat{\omega}/(1-\hat{\alpha_1}-\hat{\beta}_1)$ is the unconditional variance. When the process is covariance stationary, it follows that

$$\sigma_t^2(h) \to \hat{\sigma}^2 \quad \text{as } h \to \infty. \quad \text{or} \quad \lim_{h\to\infty}\sigma_t^2(h) = \hat{\sigma}^2.$$

3.4 Integrated GARCH model

- When $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$, the unconditional variance σ^2 is not definite anymore. (Depending on time and σ^2 does not exist)

- The process ε_t is not covariance stationary anymore.

- However, it remains strictly stationary (the solution of $\sigma_t^2 < \infty$) as shown by Nelson (1990).

Such a case corresponds to the Integrated GARCH (or IGARCH) described by Engle and Bollerslev (1986).

Definition 3. The IGARCH(p,q) process is defined by

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim iidN(0,1), \quad \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-1}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

where $\omega \geq 0$, $\alpha_i \geq 0$ and $\beta_j \geq 0$ for all i and j, and the polynomial $1 - \alpha(z) - \beta(z) = 0$ has at least one unit root.

So we have the Integrated GARCH(p, q) model when (necessary condition)

$$\alpha(1) + \beta(1) = 1.$$

IGARCH(1,1) process

To illustrate, consider the IGARCH(1,1) which is characterised by $\alpha_1 + \beta_1 = 1$, i.e.,

$$\sigma_t^2 = \omega + (1 - \beta_1)\varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

with $0 < \beta_1 \leq 1$.

- We then see that

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 + (1 - \beta_1)(\varepsilon_{t-1}^2 - \sigma_{t-1}^2),$$

  we find that it is a unit root process of σ_t^2.

- So, the h-step ahead forecast for σ_{t+h}^2 is

$$\sigma_t^2(h) = (h-1)\omega + \sigma_t^2(1), \quad \text{for} \quad h \geq 1.$$

- As it has been shown by Lumsdaine (1996), the statistical properties of the estimator of $(\omega, \beta_1)'$ are asymptotically normal.

- When $\omega = 0$, we get the *EWMA (Exponentially Weighted Moving Average Volatility)* model with values of β_1 equal to 0.94, i.e.,

$$\sigma_t^2 = 0.06\varepsilon_{t-1}^2 + 0.94\sigma_{t-1}^2.$$

This model is popularly used in the area of RiskMetrics.

3.5 Fractionally Integrated GARCH model

Again, GARCH(p, q) model can be rewritten as ARMA(m, p) representations, where $m = \max(p, q)$

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{m}(\alpha_j + \beta_j)\varepsilon_{t-j}^2 + \left(v_t - \sum_{i=1}^{p}\beta_i v_{t-i}\right).$$

or

$$\phi(L)\varepsilon_t^2 = \omega + \beta(L)v_t,$$

with

$$\phi(L) = 1 - (\alpha_1 + \beta_1)L - \cdots - (\alpha_m + \beta_m)L^m$$

and

$$\beta(L) = 1 - \beta_1 L - \cdots - \beta_p L^p.$$

Baillie, Bollerslev and Mikkelsen (1996) introduce the FIGARCH(p, d, q) model by defining ε_t^2 as an "ARMA in squares" representation

$$\phi(L)(1-L)^d \varepsilon_t^2 = \omega + \beta(L) v_t$$

where $v_t = \varepsilon_t^2 - \sigma_t^2$,

- $\phi(L) = \phi(L) = 1 - \sum_{i=1}^{q} \phi_i L^i$ and $\beta(L) = 1 - \sum_{i=1}^{p} \beta_i L^i$ are finite order lag polynomials with the roots assumed to lie outside the unit circle

- $0 < d < 1$ is the fractional differencing parameter

- When $d = 0$, FIGARCH model reduces to GARCH model

- When $d = 1$ it becomes an IGARCH model.

The FIGARCH(p, d, q) model can be represented as

$$\beta(L)\sigma_t^2 = \omega + \left[\beta(L) - \phi(L)(1-L)^d\right]\varepsilon_t^2.$$

Finally, the conditional variance of ε_t, or infinite ARCH representation of FIGARCH process, is given by

$$\sigma_t^2 = \frac{\omega}{\beta(1)} + \left[1 - \frac{\phi(L)}{\beta(L)}(1-L)^d\right]\varepsilon_t^2 = \omega/\beta(1) + \lambda(L)\varepsilon_t^2,$$

where $\lambda(L) = \lambda_1 L + \lambda_2 L^2 + \cdots$.

For any $0 < d < 1$, the λ_j coefficients will be characterized by a slow hyperbolic decay implying persistent impulse response weights (Conrad and Karanasos 2006).

3.6 The GARCH-M Model

(1) The expected risk premium

To compensate an investor for bearing more risk, an investor should receive a higher expected return, resulting in a positive relationship between the mean and the risk of the portfolio. This issue is also called the *expected risk premium*.

- Letting $\mu_t = E_{t-1}(r_t)$ and $\sigma_t^2 = E_{t-1}(r_t^2)$.

- The fundamental relationship between risk and return is specified as

$$\mu_t = c + \delta\sigma_t^{2\lambda}$$

  where $\delta > 0$ to allow for a positive relationship.

- The compensation an investor wishes to receive from bearing higher risk is given by

$$\frac{d\mu_t}{d\sigma_t} = 2\lambda\delta\sigma_t^{2\lambda-1},$$

  - Case 1: $\lambda = 0.5$ (a linear relationship between μ_t and σ_t). Compensation for bearing more risk increases at the constant rate given by

$$\frac{d\mu_t}{d\sigma_t} = \delta.$$

  - Case 2: $\lambda = 1.0$ (a nonlinear relationship between μ_t and σ_t). Compensation for bearing more risk (increase in σ_t) increases at an increasing rate

$$\frac{d\mu_t}{d\sigma_t} = 2\lambda\sigma_t.$$

(2) GARCH-M model

Engle, Lilien, and Robbins (1987) extend the Engle's ARCH model to allow the conditional variance to be a determinant of the conditional mean of the process.

The GARCH-in-mean (GARCH-M) consists of the system:

$$y_t = c + \phi x_t + \delta \sigma_t^{2\lambda} + \varepsilon_t, \quad \varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$$
$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-1}^2$$

where y_t is a financial return.

The process specifying the conditional variance is a GARCH(p, q) process.

ML Estimation

Let $\theta = (\delta_0, \delta_1, \delta_2, \omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)$ be the parameters vector.

The MLE procedure is the maximization of the conditional log likelihood function under the NID assumption:

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} L_t(\theta) = -\frac{T}{2}\ln(2\pi) - \sum_{t=1}^{T}\left(\frac{1}{2}\ln(\sigma_t^2) + \frac{\varepsilon_t^2}{2\sigma_t^2}\right),$$

with $\varepsilon_t = y_t - \delta_0 - \delta_1 x_t - \delta_2 \sigma_t^{2\lambda}$.

Moreover the consistency of the parameters estimation requires that both the first two conditional moments are correctly specified and simultaneously estimated.

4 Asymmetric and Nonlinear Models

Drawbacks of GARCH

- Black (1976): Volatility tends to rise in response to "bad news" and to fall in response to "good news" *(leverage effect)*;
- The volatility in the GARCH process is *symmetric*, determined only by the magnitude of the previous return and shock, not by its sign.
- The parameters in GARCH are restricted to be positive to ensure positivity of σ_t^2. *(nonnegativity)* (The theoretical ACFs of squared residuals are nonnegative, but the real data do not support). When estimating, however, sometimes best fits are achieved for negative parameters.

4.1 **The EGARCH model** – Nelson (1991)

To allow for asymmetric effects between positive and negative asset returns, Nelson (1991) considered the weighted innovation

$$g(z_t) = \underbrace{\varphi z_t}_{\text{sign effect}} + \underbrace{\gamma(|z_t| - E|z_t|)}_{\text{magnitude effect}},$$

where φ and γ are real constants.

- Both z_t and $|z_t| - E|z_t|$ are zero-mean iid sequences with continuous distribution. Therefore, $E[g(z_t)] = 0$.
- The asymmetry of $g(z_t)$ can easily be seen by rewriting it as

$$g(z_t) = \begin{cases} (\varphi + \gamma)z_t - \gamma E(|z_t|), & \text{if } z_t \geq 0, \\ (\varphi - \gamma)z_t - \gamma E(|z_t|), & \text{if } z_t \leq 0. \end{cases}$$

- **Remark:**

(1) $E|z_t| = \sqrt{2/\pi}$ given that $z_t \sim NID(0,1)$.

(2) For z_t distributed as a standardized Student t distribution with ν degrees of freedom, we have

$$E|z_t| = \frac{2\sqrt{\nu-2}\Gamma((\nu+1)/2)}{(\nu-1)\Gamma(\nu/2)\sqrt{\pi}}.$$

(3) For a generalized error distribution (GED) with fat-tailedness parameter ν, analyzed by Nelson (1991), we have

$$E|z_t| = \lambda 2^{1/\nu}\Gamma(2/\nu)\Gamma(1/\nu),$$

where $\lambda = \left[2^{-2/\nu}\Gamma(1/\nu)\Gamma(3/\nu)\right]^{1/2}$. We will introduce this distribution later.

EGARCH model

An EGARCH(p, q) model (Exponential GARCH(p, q)) put forward by Nelson (1991) is given by

$$\varepsilon_t = \sigma_t z_t, \quad \ln(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i g(z_{t-i}) + \sum_{j=1}^{p} \beta_j \ln(\sigma_{t-j}^2), \tag{13}$$

with $\alpha_1 = 1$ and $g(z_t) = \varphi z_t + \gamma(|z_t| - E|z_t|)$, where *the parameters ω, β_i and α_i are not restricted to be non-negative.*

- It uses logged conditional variance to relax the positiveness constraint of model coefficients.

- The use of $g(z_t)$ enables the model to respond asymmetrically to positive and negative lagged values of ε_t.

- **Stationarity:** In order to simply state the stationarity conditions, we write the EGARCH(p, q) model as:

$$\left(1 - \sum_{i=1}^{p} \beta_i L^i\right) \ln(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i L^i g(z_t).$$

  Since $\ln(\sigma_t^2)$ is written in ARMA(p, q) form, when $1 - \beta(z) = 1 - \sum_{i=1}^{p} \beta_i z^i$ and $\alpha(z) = \sum_{i=1}^{q} \alpha_i z^i$ have no common roots, conditions for *covariance stationarity* of $\ln(\sigma_t^2)$ are equivalent to all the roots of $1 - \beta(z) = 0$ lying outside the unit circle.

To better understand the EGARCH model, let us consider the simple model with order (1,1):

$$\varepsilon_t = \sigma_t z_t, \quad (1 - \beta L)\ln(\sigma_t^2) = \omega + \alpha g(z_{t-1}),$$

where the z_t are iid standard normal.

- In this case, $E(|z_t|) = \sqrt{2/\pi}$ and the model for $\ln(\sigma_t^2)$ becomes

$$(1 - \beta L)\ln(\sigma_t^2) = \begin{cases} \omega_* + (\varphi + \gamma) z_{t-1}, & \text{if } z_{t-1} \geq 0 \\ \omega_* + (\varphi - \gamma)|z_{t-1}|, & \text{if } z_{t-1} < 0. \end{cases}$$

where $\omega_* = \omega - \sqrt{\frac{2}{\pi}}\gamma$. This is a nonlinear function similar to that of the threshold autoregressive (TAR) model of Tong (1978, 1990).

- Furthermore, we have

$$\sigma_t^2 = \sigma_{t-1}^{2\beta} \exp(\omega_*) \begin{cases} \exp\left[(\varphi+\gamma)\dfrac{\varepsilon_{t-1}}{\sigma_{t-1}}\right], & \text{if } \varepsilon_{t-1} \geq 0 \\ \exp\left[(\varphi-\gamma)\dfrac{\varepsilon_{t-1}}{\sigma_{t-1}}\right], & \text{if } \varepsilon_{t-1} < 0. \end{cases}$$

  - The coefficients $(\varphi+\gamma)$ and $\varphi-\gamma$ show the asymmetry in response to positive and negative a $t-1$.
  - The model is, therefore, nonlinear if $\gamma \neq 0$.
  - Since negative shocks tend to have larger impacts, we expect φ to be negative.

4.2 The GJR-GARCH model

The GJR model (Glosten, Jagannathan, and Runkle, 1993) is defined by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p}[\alpha_i \varepsilon_{t-i}^2 + \delta_i \Pi_{t-i}^- \varepsilon_{t-i}^2] + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{14}$$

where Π_t^- is a dummy variable, defined as

$$\Pi_t^- = \begin{cases} 1, & \text{if } \varepsilon_t < 0; \\ 0, & \text{otherwise.} \end{cases}$$

The conditional volatility is positive when parameters satisfy $\omega > 0$, $\alpha_i \geq 0$, $\alpha_i + \delta_i \geq 0$ and $\beta_j \geq 0$, for $i = 1, \cdots, p$ and $j = 1, \cdots, q$.

The process is covariance stationary if and only if $\sum_{i=1}^{p}(\alpha_i + \delta_i/2) + \sum_{j=1}^{q} \beta_j < 1$ (Hentschel, 1995).

4.3 The APARCH Model

This model has been introduced by Ding, Granger, and Engle (1993). The APARCH (p, q) model can be expressed as: :

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^\delta,$$

where $\delta > 0$ and $-1 < \gamma_i < 1$ $(i = 1, \ldots, q)$.

This generalized version of ARCH model includes several other models as special cases.

- ARCH(q) model, just let $\delta = 2$ and $\gamma_i = 0$, $i = 1, \ldots, q$, $\beta_j = 0$, $j = 1, \ldots, p$.

- GARCH(p, q) model just let $\delta = 2$ and $\gamma_i = 0$, $i = 1, \ldots, q$.

- Taylor/Schwert's GARCH in standard deviation model just let $\delta = 1$ and $\gamma_i = 0$, $i = 1, \ldots, q$.

- GJR model just let $\delta = 2$.
  - When $\delta = 2$ and $0 \leq \gamma_i < 1$,

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (1 - \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} 4 \alpha_i \gamma_i S_{t-i}^- \varepsilon_{t-i}^2$$

where

$$S_{t-i}^- = \begin{cases} 1, & \text{if } \varepsilon_{t-i} < 0; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, this model becomes the GJR model.

– When $\delta = 2$ and $-1 < \gamma_i \leq 0$,

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (1+\gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 - \sum_{i=1}^{q} 4\alpha_i \gamma_i S_{t-i}^{+} \varepsilon_{t-i}^2$$

where

$$S_{t-i}^{+} = \begin{cases} 1, & \text{if } \varepsilon_{t-i} > 0; \\ 0, & \text{otherwise.} \end{cases}$$

This model allows positive shocks to have a stronger effect on volatility.

- Zakoian's TARCH model (Zakoian (1991)), let $\delta = 1$ and $\beta_j = 0$, $j = 1, \ldots, p$. We have

$$\begin{aligned}\sigma_t &= \omega + \sum_{i=1}^{q} \alpha_i(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}) \\ &= \omega + \sum_{i=1}^{q} \alpha_i(1-\gamma_i)\varepsilon_{t-i}^{+} - \sum_{i=1}^{q} \alpha_i(1+\gamma_i)\varepsilon_{t-i}^{-}\end{aligned}$$

where

$$\varepsilon_{t-i}^{+} = \begin{cases} \varepsilon_{t-i}, & \text{if } \varepsilon_{t-i} > 0; \\ 0, & \text{otherwise.} \end{cases} \quad \text{and} \quad \varepsilon_{t-i}^{-} = \begin{cases} \varepsilon_{t-i}, & \text{if } \varepsilon_{t-i} < 0; \\ 0, & \text{otherwise.} \end{cases}$$

If we let $\beta_j \neq 0$, $j = 1, \ldots, p$ then we get a more general class of TARCH models.

News Impact Curve

Pagan and Schwert (1990) and Engle and Ng (1993)

The *news impact curve* relates past return shocks (news) to current volatility, and measures how new information is incorporated into volatility estimates.

In the GARCH model, this curve is a quadratic function centered on $\varepsilon_{t-1} = 0$.

For asymmetric models, the curve is designed to increase differently in the two directions.

GARCH(1,1) Model

For the GARCH(1,1) model

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

The news impact curve is a quadratic function centered on $\varepsilon_{t-1} = 0$,

$$\sigma_t^2 = A + \alpha_1\varepsilon_{t-1}^2,$$

with $A = \omega + \beta_1\sigma^2$ and σ^2 the unconditional variance.

EGARCH(1,1) Model

For the EGARCH(1,1) model

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \varphi z_{t-1} + \gamma(|z_{t-1}| - E|z_{t-1}|)$$

where $z_t = \varepsilon_t/\sigma_t \sim NID(0,1)$. The news impact curve is

$$\sigma_t^2 = \begin{cases} A\exp\left(\frac{\varphi+\gamma}{\sigma}\varepsilon_{t-1}\right), & \text{for } \varepsilon_{t-1} > 0; \\ A\exp\left(\frac{\varphi-\gamma}{\sigma}\varepsilon_{t-1}\right), & \text{for } \varepsilon_{t-1} \le 0. \end{cases}$$

with $A = \sigma^{2\beta}\exp\left(\omega - \gamma\sqrt{2/\pi}\right)$, where $\varphi < 0$, and $\varphi + \gamma > 0$. It has a minimum at $\varepsilon_{t-1} = 0$.

The EGARCH allows good news and bad news to have different impact on volatility, while the standard GARCH does not.

GJR-GARCH(1,1) Model

For the GJR-GARCH(1,1) model

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

where $S_{t-1}^- = 1$ if $\varepsilon_{t-1} < 0$, otherwise $S_{t-1}^- = 0$. The news impact curve is

$$\sigma_t^2 = \begin{cases} A + \alpha\varepsilon_{t-1}^2, & \text{if } \varepsilon_{t-1} > 0; \\ A + (\alpha+\gamma)\varepsilon_{t-1}^2, & \text{if } \varepsilon_{t-1} \le 0. \end{cases}$$

with $A = \omega + \beta\sigma^2$, where $\omega > 0$, $0 \le \beta < 1$, $\sigma > 0$, $0 \le \alpha < 1$, and $\alpha + \beta < 1$. It has a minimum at $\varepsilon_{t-1} = 0$.

NIC Comparison

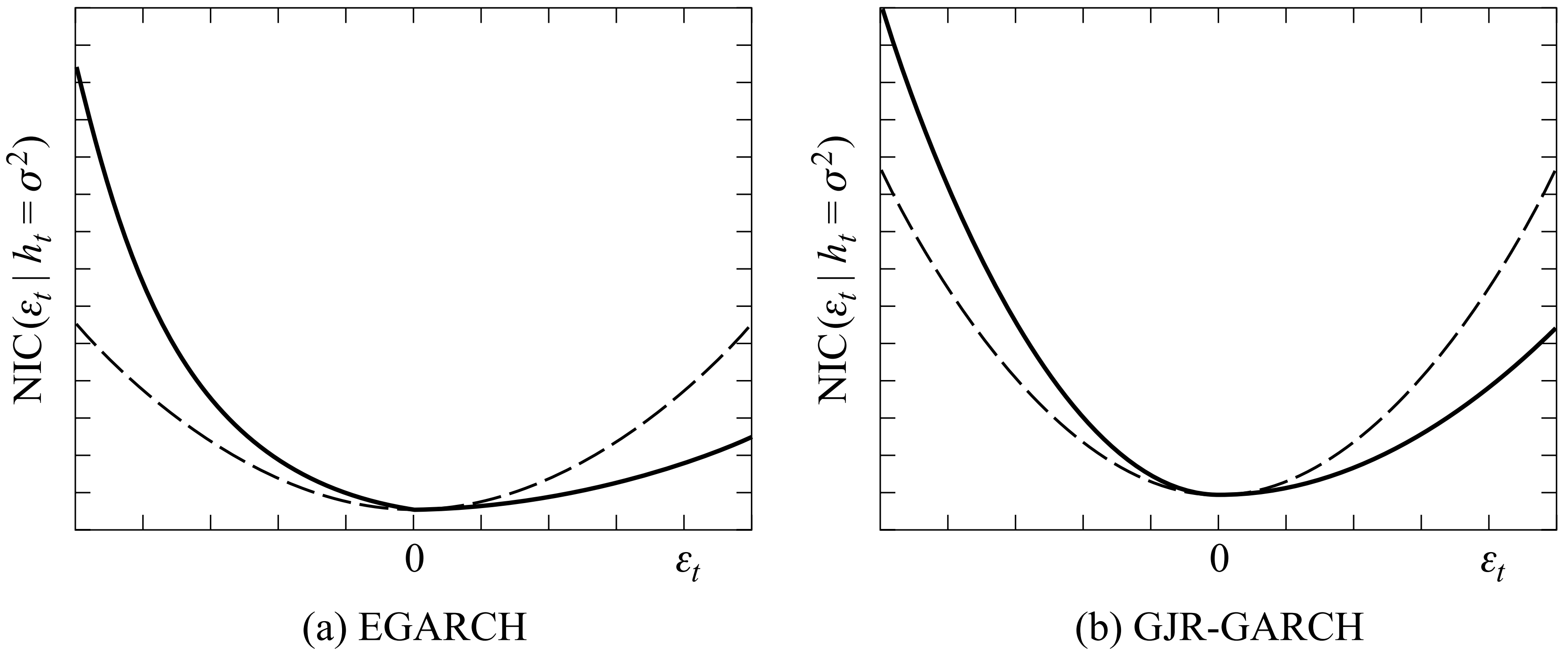

Figure 3 Examples of news impact curves for various asymmetric models (solid lines); for comparison, the news impact curve of a GARCH(1,1) model is also shown in each panel (dashed line).

5 Value at Risk

5.1 Definition of VaR

- VaR is a measure of minimum loss of a financial position within a certain period of time for a given (small) probability.

- A formal definition under a probabilistic framework
  - time period given: $\ell = \Delta t$
  - loss in value: L
  - CDF of the loss: $F_\ell(x)$
  - given (upper tail) probability: p
  - VaR is defined as

$$p = \Pr[L(\ell) \geq \mathrm{VaR}] = 1 - \Pr[L(\ell) < \mathrm{VaR}] = 1 - F_\ell(\mathrm{VaR}), \quad (15)$$

- Quantile: x_q is the 100q-th quantile of the distribution $F_\ell(x)$ if

$$q = F_\ell(x_q), \quad \text{i.e.,} \quad q = \Pr(L \le x_q)$$

and $F_\ell(\cdot)$ is continuous. For discrete distribution, we have

$$x_q = \min\{x \mid \Pr(L \le x) \ge q\}.$$

In real applications, several factors affect VaR:

- The probability p, such as $p = 0.01$ for risk management and $p = 0.001$ in stress testing.
- The time horizon ℓ. 1 day or 10 days for market risk and 1 year or 5 years for credit risk.
- The CDF $F_\ell(x)$ or its quantiles. (or CDF of loss)
- The amount of the financial position or the mark-to-market value (市价计值) of the portfolio.

5.2 Riskmetrics

- The RiskMetrics methodology to VaR calculation is developed by J. P. Morgan. [3]

- RiskMetrics assumes that r_t follows an *IGARCH process* below

$$r_t \mid \mathcal{F}_{t-1} \sim N(\mu_t, \sigma_t^2), \tag{16}$$

$$\mu_t = 0, \quad \sigma_t^2 = \beta\sigma_{t-1}^2 + (1-\beta)r_{t-1}^2, \quad 0 < \beta < 1. \tag{17}$$

  Often, β is in the interval $(0.9, 1)$ with a typical value of 0.94.

- Assuming $\mu_t = 0$ and consider the VaR for the next trading day $(\ell = 1)$
  - $\mathrm{VaR} = 1.65\sigma_{t+1}$ if $p = 0.05$.

[3] Morgan, J.P., 1996, RiskMetrics Technical Document, fourth edition, New York.

– $\mathrm{VaR} = 2.326\sigma_{t+1}$ if $p = 0.01$.
– In general, $\mathrm{VaR} = z_{1-p}\sigma_{t+1}$, where z_{1-p} is the $100(1-p)$th quantile of the standard normal distribution.
– The VaR for the portfolio under RiskMetrics if $p = 0.01$ is then

$$\mathrm{VaR}(1) = \text{Amount of position} \times 2.326\sigma_{t+1}.$$

- For a k-period horizon,
  – k-horizon return: $r_t[k] = r_{t+1} + \cdots + r_{t+k-1} + r_{t+k}$.
  – Using the independence assumption of ε_t and the IGARCH model, we have

$$\sigma_t^2[k] = \mathrm{Var}(r_t[k] \mid \mathcal{F}_t) = \sum_{i=1}^{k} \mathrm{Var}(r_{t+i} \mid \mathcal{F}_t),$$

where $\mathrm{Var}(r_{t+i} \mid \mathcal{F}_t) = E(\sigma_{t+i}^2 \mid \mathcal{F}_t)$ can be obtained recursively.

– For IGARCH(1,1) model, we can easily show that

$$E(\sigma^2_{t+i} \mid \mathcal{F}_t) = E(\sigma^2_{t+i-1} \mid \mathcal{F}_t) \quad \text{for} \quad i = 2, \ldots, k.$$

This implies that $\mathrm{Var}(r_{t+i} \mid \mathcal{F}_t) = \sigma^2_{t+1}$ for $i \geq 1$ and hence $\sigma^2_t[k] = k\sigma^2_{t+1}$.

– The results show that $r_t[k] \mid \mathcal{F}_t \sim N(0, k\sigma^2_{t+1})$.

– Therefore, for the next k trading days, $\mathrm{VaR}[k] = 1.65\sqrt{k}\sigma_{t+1}$ if $p = 0.05$. Similarly, $\mathrm{VaR}[k] = 2.326\sqrt{k}\sigma_{t+1}$ if $p = 0.01$.

– The VaR for the portfolio under RiskMetrics if $p = 0.01$ is then

$$\mathrm{VaR}(k) = \text{Amount of position} \times 2.326\sqrt{k}\sigma_{t+1}$$

for the next trading day.

– Consequently, under RiskMetrics, we have

$$\mathrm{VaR}(k) = \sqrt{k} \times \mathrm{VaR}(1).$$

This is referred to as the square root of time rule in VaR calculation under RiskMetrics.

Example 4. Consider modelling daily log returns of IBM stock, July 3, 1962 - Dec. 31, 1998 (totally 9190 observations)

- Model: $r_t = \varepsilon_t$, $\varepsilon_t = \sigma_t z_t$, and

$$\sigma_t^2 = 0.9396\sigma_{t-1}^2 + (1 - 0.9396)\varepsilon_{t-1}^2$$

- From data and fitted model, $r_{9190} = -0.0128$ and $\hat{\sigma}_{9190}^2 = 0.0003472$.

- 1-step-ahead volatility forecast is $\hat{\sigma}_{9190}^2(1) = 0.000336$.

- The 95% quantile of $r_{9191}|\mathcal{F}_{9190}$ is $1.65 \times \sqrt{0.000336} = 0.03025$.

- Consequently, 1-day horizon 5% VaR of a long position of $10 millions is

$$\mathrm{VaR} = \$10,000,000 \times 0.03025 = \$302,500.$$

5.3 Expected Shortfall (期望损失)

From the prior discussion, VaR is simply the $100(1-p)$-th quantile of the loss function, where p is the upper tail probability. When an extreme loss occurs, i.e. VaR is exceeded, the actual loss can be much higher than VaR.

To better quantify the loss and to employ a *coherent risk measure (一致风险度量)*, we consider the *expected loss* once the VaR is exceeded.

- For a given upper tail probability p, let $q = 1 - p$ and VaR_q be the associated VaR, i.e., VaR_q is the q-th quantile of X.

- The expected shortfall (ES) under the standard normal distribution $X \sim N(0,1)$ is then

$$
\begin{aligned}
\text{ES}_q = E[X \mid X > \mathrm{VaR}_q] &= \frac{1}{p}\int_{\mathrm{VaR}_q}^{+\infty} x\phi(x)dx = \frac{1}{p\sqrt{2\pi}}\int_{\mathrm{VaR}_q}^{+\infty} xe^{-\frac{x^2}{2}}dx \\
&= -\frac{1}{p\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Big|_{\mathrm{VaR}_q}^{+\infty} = \frac{1}{p\sqrt{2\pi}}e^{-\frac{\mathrm{VaR}_q^2}{2}} \\
&= \frac{\phi(\mathrm{VaR}_q)}{p}
\end{aligned}
$$

- Consequently, the expected shortfall for a log return r_t with conditional distribution $N(0, \sigma_t^2)$ is then

$$
\text{ES}_q = \frac{\phi(\mathrm{VaR}_q)}{p}\sigma_t \quad \text{or} \quad \text{ES}_{1-p} = \frac{\phi(\mathrm{VaR}_{1-p})}{p}\sigma_t,
$$

where σ_t is the volatility.

Example 5.

- if $p = 0.05$, then $\mathrm{VaR}_{0.95} \approx 1.645$ and $\phi(\mathrm{VaR}_q)/p = \phi(1.645)/0.05 = 2.0627$ so that the expected shortfall under RiskMetrics is $\text{ES}_{0.95} = 2.0627\sigma_t$.

- If $p = 0.01$, then $\text{ES}_{0.99} = 2.6652\sigma_t$.

In general, for a normal distribution $N(\mu_t, \sigma_t^2)$, the ES is

$$\text{ES}_q = \mu_t + \frac{\phi(x_q)}{p} \times \sigma_t,$$

where $q = 1 - p$ and x_q is the q-th quantile of the standard normal distribution.

For instance, if $p = 0.01$, then $\text{ES}_{0.99} = \mu_t + 2.6652\sigma_t$.

5.4 Econometric Approach

For simplicity, we consider the use of GARCH models as an econometric approach to VaR calibration. Other volatility models, including SV models, the nonlinear ones, can also be used.

A general time series model for the log return r_t can be written as

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j},$$

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim iid(0,1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{u} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=q}^{v} \beta_j \sigma_{t-j}^2.$$

(1) One-period

We can use the above ARMA-GARCH model to obtain *1-step-ahead forecasts* of the conditional mean and conditional variance of r_t, denoted by $\hat{r}_t(1)$ and $\hat{\sigma}_t^2(1)$.

- If z_t is Gaussian, then the conditional distribution of r_t is $N(\hat{r}_t(1), \hat{\sigma}_t^2(1))$. The quantile of this conditional distribution can easily be obtained for VaR calculation, that is,

$$\hat{r}_t(1) + Z_{1-p}\hat{\sigma}_t(1),$$

  where Z_{1-p} is the $1-p$-th quantile of standard normal. For example, the 95% quantile is $\hat{r}_t(1) + 1.65\hat{\sigma}_t(1)$.

- If z_t is a standardized Student-t with ν, then the quantile is

$$\hat{r}_t(1) + t^*_\nu(1-p)\hat{\sigma}_t(1),$$

where $t^*_\nu(1-p)$, the $1-p$-th quantile of a standardized Student-t, can be computed by

$$p = \Pr[t_\nu \le q] = \Pr\left[\frac{t_v}{\sqrt{\nu/(\nu-2)}} \le \frac{q}{\sqrt{\nu/(\nu-2)}}\right] = \Pr\left[t^*_v \le \frac{q}{\sqrt{\nu/(\nu-2)}}\right],$$

where q is p-th quantile of a Student-t. Therefore, the $1-p$-th quantile used to calculate the 1-period horizon VaR at time t is

$$\hat{r}_t(1) + \frac{t_v(1-p)}{\sqrt{\nu/(\nu-2)}}\hat{\sigma}_t(1),$$

where $t_\nu(1-p)$, the $1-p$-th quantile of a Student-t with ν.

Example 6. Consider again the daily IBM log return as before.

- Case 1: (Gaussian) Assume that z_t is normal. The fitted model is

$$r_t = -0.00066 - 0.0247 r_{t-2} + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t,$$

$$\sigma_t^2 = 0.00000389 + 0.0799 \varepsilon_{t-1}^2 + 0.9073 \sigma_{t-1}^2.$$

From $r_{9189} = 0.00201$, $r_{9190} = 0.0128$, $\sigma_{9190}^2 = 0.00033455$, and $\varepsilon_{9190}^2 = 0.00017109$, we have

$$\hat{r}_{9190}(1) = -0.00071 \quad \text{and} \quad \hat{\sigma}_{9190}^2(1) = 0.0003211.$$

The 95% quantile is then

$$-0.00071 + 1.6449 \times \sqrt{0.0003211} = 0.02877,$$

and the VaR for a long position of \$10 million with probability 0.05 is

$$\mathrm{VaR} = \$10,000,000 \times 0.02877 = \$287,700.$$

If the tail probability is 0.01, then the 99% quantile is

$$-0.00071 + 2.3262 \times \sqrt{0.0003211} = 0.0409738.$$

The VaR for the position becomes \$409,738.

- Case 2: (Student-t) Assume that z_t is a standardized Student-t distribution with $\nu = 5$. The fitted model is

$$r_t = -0.0003 - 0.0335 r_{t-2} + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t,$$

$$\sigma_t^2 = 0.000003 + 0.0559\varepsilon_{t-1}^2 + 0.9350\sigma_{t-1}^2.$$

From the data, we have $r_{9189} = 0.00201$, $r_{9190} = 0.0128$, $\sigma_{9190}^2 = 0.000349$, and $\varepsilon_{9190}^2 = 0.0001661$. The 1-step-ahead forecasts

$$\hat{r}_{9190}(1) = -0.000367 \quad \text{and} \quad \hat{\sigma}_{9190}^2(1) = 0.0003386.$$

If $p = 0.05$, then the quantile is

$$-0.000367 + (2.015/\sqrt{5/3})\sqrt{0.0003386} = 0.028354.$$

Therefore, the VaR for a long position of $10 milion is

$$\text{VaR} = \$10,000,000 \times 0.028352 = \$283,520.$$

If $p = 0.01$, the 99% quantile of the conditional distribution is

$$-0.000367 + (3.3649/\sqrt{5/3})\sqrt{0.0003386} = 0.0475943.$$

The corresponding VaR is $475,943.

- Comparison: We see the heavy-tail effect of using a Student-t distribution increases the VaR when the tail probability becomes smaller.

(2) Multiple periods

Suppose that at time h we want to compute the k-horizon VaR.

This needs to get the forecast values of $r_h[k] = r_{h+1} + \cdots + r_{h+k}$ and its variance.

- The conditional mean $E(r_h[k]|\mathcal{F}_h)$ can be obtained by the forecasting method of ARMA model. Specifically, we have

$$\hat{r}_h[k] = r_h(1) + \cdots + r_h(k),$$

where $r_h(\ell)$ is the ℓ-step-ahead forecast of r_t at forecast origin h.

- Meanwhile, we can obtain the ℓ-step-ahead forecast error

$$e_h(\ell) = r_{h+\ell} - r_h(\ell) = \varepsilon_{h+\ell} + \psi_1\varepsilon_{h+\ell-1} + \cdots + \psi_{\ell-1}\varepsilon_{h+1}$$

and hence the forecast error of the expected k-period return $\hat{r}_h[k]$ is

$$
\begin{aligned}
e_h[k] &= e_h(1) + e_h(2) + \cdots + e_h(k) \\
&= \varepsilon_{h+k} + (1+\psi_1)\varepsilon_{h+k-1} + \cdots + \sum_{i=0}^{k-1} \psi_i \varepsilon_{h+1}.
\end{aligned}
$$

- Under independent assumption of z_{t+i} for $i = 1, \ldots, k$, where $\varepsilon_{t+i} = \sigma_{t+i} z_{t+i}$, the conditional variance of the forecast error of $\hat{r}_h[k]$ is then

$$
\begin{aligned}
V_h(e_h[k]) &= V_h(\varepsilon_{h+k}) + (1+\psi_1)^2 V_h(\varepsilon_{h+k-1}) + \cdots + \left(\sum_{i=0}^{k-1} \psi_i\right)^2 V_h(\varepsilon_{h+1}) \\
&= \sigma_h^2(k) + (1+\psi_1)^2 \sigma_h^2(k-1) + \cdots + \left(\sum_{i=0}^{k-1} \psi_i\right)^2 \sigma_h^2(1)
\end{aligned}
$$

- If z_t is Gaussian, then the quantiles need in VaR calculations are readily available.

- If the conditional distribution of ε_t is not Gaussian, simulation can be used to obtain the multiperiod VaR.