Lecture 3: Conditional Heteroscedastic Models

Instructor: Lin Juan

SOE & WISE, Xiamen University

Empirical regularities

Common characteristics in financial time series,

- Asset prices are generally non stationary. Returns are usually stationary.
- Return series usually show no or little autocorrelation.
- Normality has to be rejected in favor of some thick-tailed distribution.
- Serial independence between the squared values of the series is often rejected pointing towards the existence of non-linear relationships between subsequent observations.

What is volatility

- Volatility refers to the spread of asset returns. In finance, volatility is often used as a proxy for the underlying riskiness of an asset.
- A special feature of volatility is that it is not directly observable.
- For example, consider the daily log returns of IBM stock. The daily volatility is not directly observable from the return data because there is only one observation in a trading day.

Volatility

1. Three concepts of volatility:
   - Unconditional volatility is volatility over an entire time period (σ), for example, historical volatility;
   - Conditional volatility is volatility in a given time period, conditional on what happened before (σ_t);
   - Implied volatility (mentioned later)
2. Clear evidence of cyclical patterns in volatility over time, both in the short run and the long run.

Historical Volatility

- Statistically, volatility can be measured by the (unconditional) standard deviation or variance of the returns of a financial asset.
- The simplest measure is historical volatility – the standard deviation or variance over some historical period

Calculation of historical volatility

- Daily volatility

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu)^2}$$

- Annualized volatility
  When volatility is measured over an interval other than a year, such as a day, week or month, it can always be scaled to reflect the volatility of the asset over a year.
  There are 252 trading days in a year, the annualized volatility is $\sqrt{252}\sigma$.

Implied Volatility

- The Black-Scholes formula for an option price includes the variance of stock market prices. One might derive the implied volatility of a stock from the option price by manipulation of the Black-Scholes formula.
- The VIX of Chicago Board Options Exchange (CBOE) is an implied volatility.

Empirical regularities

Although volatility is not directly observable, it has some characteristics that are commonly seen in asset returns.

- Volatility of the return series appears to be clustered, that is, large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes.

Empirical regularities

- Some series exhibit so-called leverage effect, that is, the volatility of stocks tends to increase when the price drops.
- Volatilities of different securities very often move together.

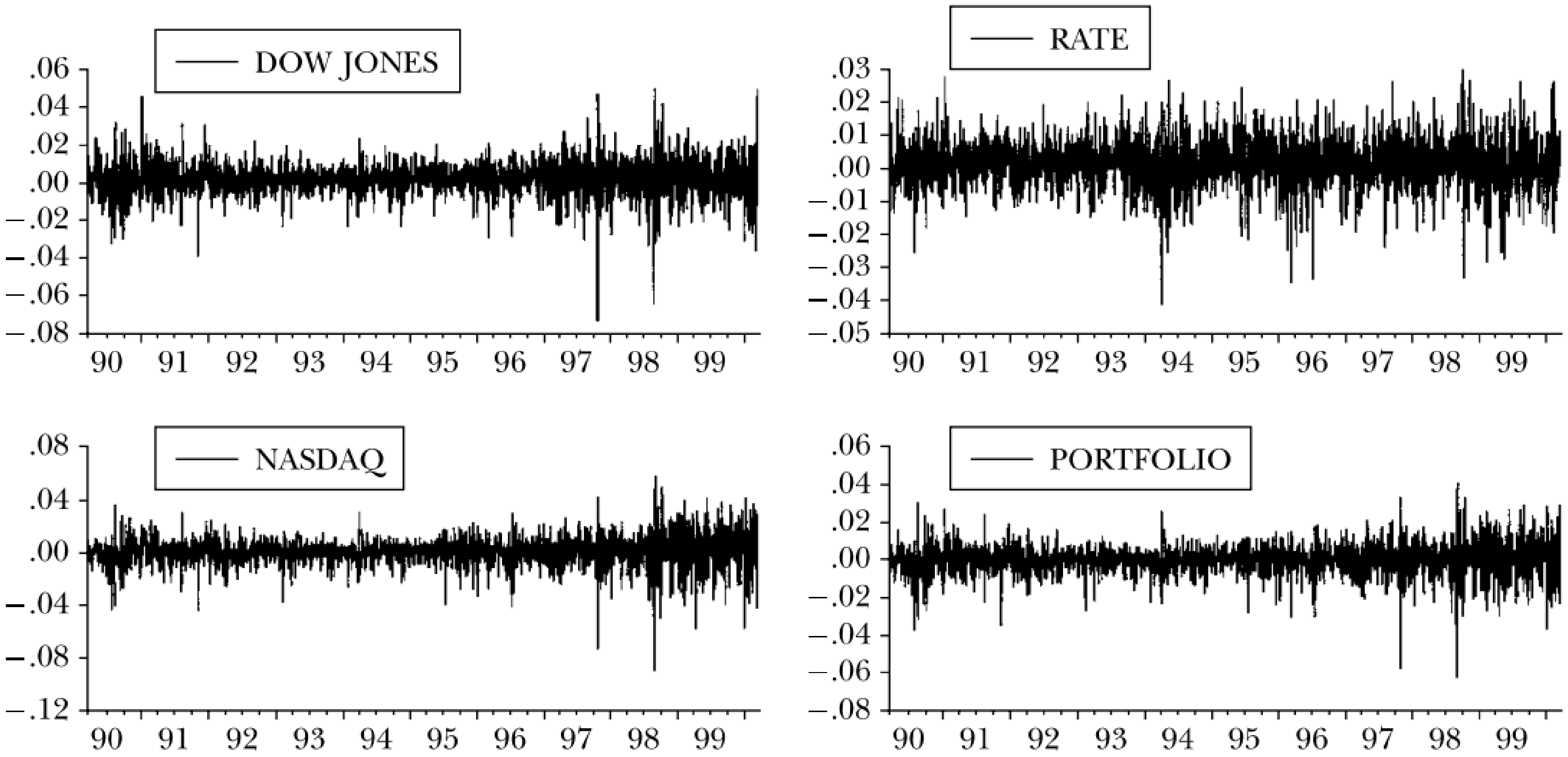

Figure 1: Nasdaq, Dow Jones and Bond Returns

Drawbacks of ARMA models

- ARMA models are used to model the conditional expectation of a process given the past.
- In an ARMA model, the conditional variance given the past is constant.
- Suppose we have noticed that recent daily returns have been unusually volatile. We might expect that tomorrow's return is also more variable than usual.
- However, an ARMA model cannot capture this type of behavior because its conditional variance is constant.

Constant conditional variance models

- Consider the models with a constant conditional variance, $Var(r_t|\Omega_{t-1}) = \sigma^2$, where $\Omega_{t-1} = (r_{t-1}, r_{t-2}, ..., r_1)$.
- The general form of the model is

$$r_t = \mu_t + e_t,$$

where

$$\mu_t = E(r_t|\Omega_{t-1}) = g(\Omega_{t-1})$$

and e_t is independent of Ω_{t-1} and has expectation equal to 0 and constant conditional variance σ^2.

Constant conditional variance models

- The general form of the model with a constant conditional variance is

$$r_t = \mu_t + e_t,$$

- For the linear series r_t, for example, ARMA process, $g(\cdot)$ is a linear function of elements of Ω_{t-1}.

$$\begin{aligned}\mu_t &= E(r_t|\Omega_{t-1}) = g(\Omega_{t-1}) \\ &= \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i e_{t-i},\end{aligned}$$

Conditional heteroskedasticity models

- Consider the models with a conditional heteroskedasticity, $Var(r_t|\Omega_{t-1}) = \sigma_t^2 = h(\Omega_{t-1})$.
- The general form of the model becomes,

$$r_t = \mu_t + e_t = E(r_t|\Omega_{t-1}) + \sigma_t \varepsilon_t,$$

  where $\varepsilon_t = e_t/\sigma_t$ is a standardized shock (or innovation).
- $\varepsilon_t \sim iid(0,1)$, which means that $\{\varepsilon_t\}$ is a sequence of *iid* r.v. with mean 0 and variance 1.

Requirement for σ_t

- σ_t should be nonnegative since it is a standard deviation.
- If the function σ_t is linear, then its coefficients must be constrained to ensure nonnegativity. Such constraints are cumbersome to implement.
- Nonlinear nonnegative functions are usually used instead.

Basic structures

- $r_t = \mu_t + e_t$
- Volatility models are concerned with time-evolution of the conditional variance of the return r_t.

$$\sigma_t^2 = Var(r_t|\Omega_{t-1}) = Var(e_t|\Omega_{t-1}),$$

- The ARCH and GARCH models are important classes of variance function models.

Example

Consider the daily closing index of the S&P500 index from January 03, 2007 to April 14, 2016. The log returns follow approximately an MA(2) model

$$r_t = e_t - 0.109e_{t-1} - 0.054e_{t-2}, \quad \sigma^2 = 0.00018.$$

- How about the volatility?
- Is volatility constant over time?
- NO! See the ACF of squared residuals!

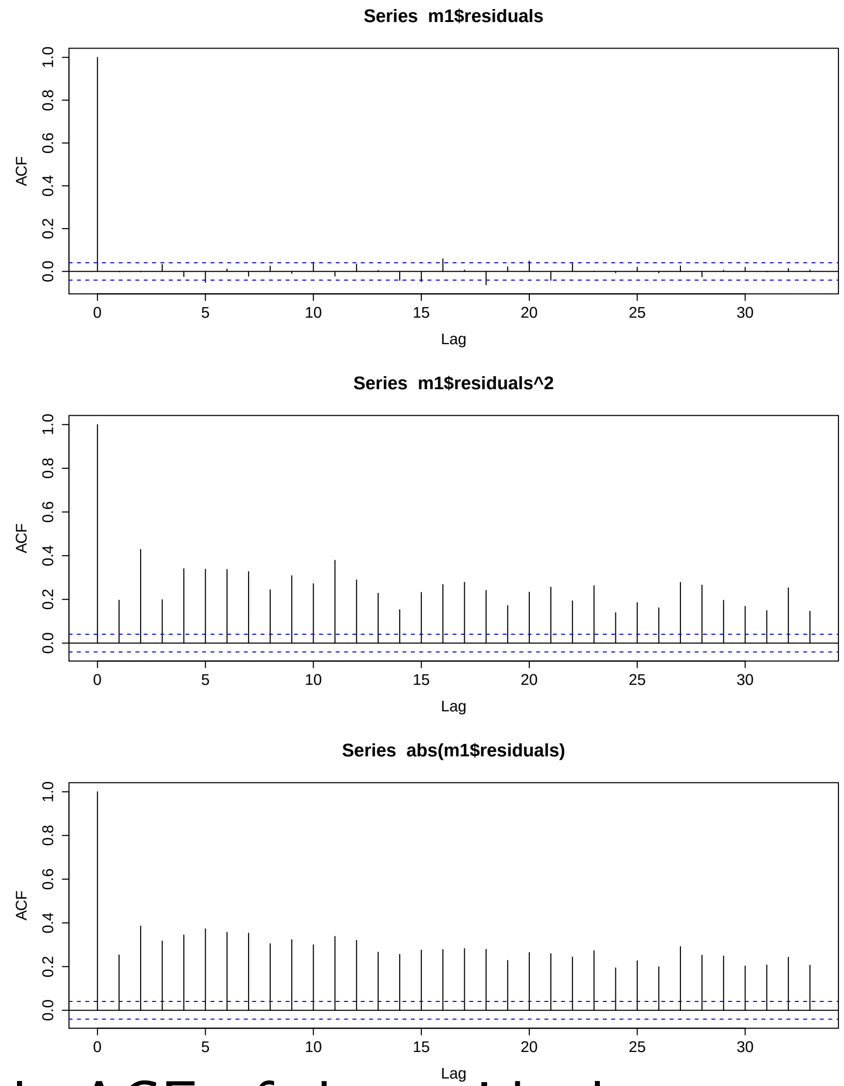

Figure 2: Sample ACF of the residuals, squared residuals and absolute residuals of an MA(2) model fitted to daily log returns of the S&P 500 index from January 3, 2007 to April 14, 2016.

How to model the evolving volatility

How to model the evolving volatility? Two general categories:

- "Fixed function" of the available information;
- Stochastic function of the available information;

Univariate volatility models

Univariate volatility models discussed:

- Autoregressive conditional heteroscedastic (ARCH) model of Engle (1982);
- Generalized ARCH (GARCH) model of Bollerslev (1986);
- GARCH-M models;
- IGARCH models (used by RiskMetrics);
- Exponential GARCH (EGARCH) model of Nelson (1991);

Univariate volatility models

Univariate volatility models discussed:

- Threshold GARCH model of Zakoian (1994) or GJR model of Glosten, Jagannathan, and Runkle (1993);
- Asymmetric power ARCH (APARCH) models of Ding, Granger and Engle (1994), [TGARCH and GJR models are special cases of APARCH models.];
- Stochastic volatility (SV) models of Melino and Turnbull (1990), Harvey, Ruiz and Shephard (1994), and Jacquier, Polson and Rossi (1994).

ARCH(1) model

Basic idea of ARCH model:

- The shock e_t of an asset return is serially uncorrelated, but dependent;
- The dependence of e_t can be described by a simple quadratic function of its lagged values.

Uncorrelation and independence

- X, Y, are said to be uncorrelated if their covariance, $E(XY) - E(X)E(Y)$, is zero.
- If two variables are uncorrelated, there is no linear relationship between them.
- If X and Y are independent, with finite second moments, then they are uncorrelated. However, not all uncorrelated variables are independent.

Uncorrelation and independence

Let X be a random variable that takes the value 0 with probability $1/2$, and takes the value 1 with probability $1/2$.

Let Z be a random variable, independent of X, that takes the value -1 with probability $1/2$, and takes the value 1 with probability $1/2$.

Let U be a random variable constructed as $U = XZ$.

Show that U and X have zero covariance (and thus are uncorrelated), but are not independent.

Uncorrelation and independence

Show that U and X have zero covariance (and thus are uncorrelated), but are not independent.

First note

$$E[U] = E[XZ] = E[X]E[Z] = E[X] * 0 = 0$$

Now

$$\begin{aligned} Cov(U, X) &= E[(U - E(U))(X - E(X))] \\ &= E[U(X - 1/2)] \\ &= E[(X^2 - 1/2X)Z] = E[X^2 - 1/2X]E[Z] = 0 \end{aligned}$$

Uncorrelation and independence

Independence of U and X means that for all a and b, $\Pr(U = a|X = b) = \Pr(U = a)$. This is not true, in particular, for $a = 1$ and $b = 0$,

$$\Pr(U = 1|X = 0) = \Pr(XZ = 1|X = 0) = 0$$
$$\Pr(U = 1) = \Pr(XZ = 1) = 1/4$$

Thus $\Pr(U = 1|X = 0) \neq \Pr(U = 1)$ so U and X are not independent.

ARCH(1) model

- Consider an ARCH(1) model with constant mean of 0,

$$
\begin{aligned}
r_t &= e_t \\
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1), \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2,
\end{aligned}
$$

where e_t denotes the error terms (return residuals, with respect to a mean process), $\{\varepsilon_t\}$ is a sequence of *iid* r.v. with mean 0 and variance 1.

ARCH(1) model

- Consider an ARCH(1) model with constant mean of 0,

$$
\begin{aligned}
r_t &= e_t \\
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1), \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2,
\end{aligned}
$$

- We impose the constraints $\alpha_0 > 0$ and $\alpha_1 \geq 0$ to avoid negative variance.
- α_i must satisfy some regularity conditions to ensure that the unconditional variance of e_t is finite.

ARCH(1) model

- Consider an ARCH(1) model with constant mean of 0,

$$
\begin{aligned}
r_t &= e_t \\
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1), \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2,
\end{aligned}
$$

- Large past squared shocks e_{t-1}^2 imply a large conditional variance σ_t^2 for the innovation e_t (We will show later that $\sigma_t^2 = Var(e_t|\Omega_{t-1})$.). Consequently, e_t tends to assume a large value (in modulus).

- Consider an ARCH(1) model with constant mean of 0,

$$
\begin{aligned}
r_t &= e_t \\
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1), \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2,
\end{aligned}
$$

- e_t tends to assume a large value (in modulus). This means that, under the ARCH framework, large shocks tend to be followed by another large shock.

- Consider an ARCH(1) model with constant mean of 0,

$$
\begin{aligned}
r_t &= e_t \\
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1), \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2,
\end{aligned}
$$

- A large variance does not necessarily produce a large realization. It only says that the probability of obtaining a large variate is greater than that of a smaller variance. This feature is similar to the volatility clusterings observed in asset returns.

Simulate ARCH(1) model

- Simulate from an ARCH(1) model, where

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1), \\
\sigma_t^2 &= 0.1 + 0.8 e_{t-1}^2,
\end{aligned}
$$

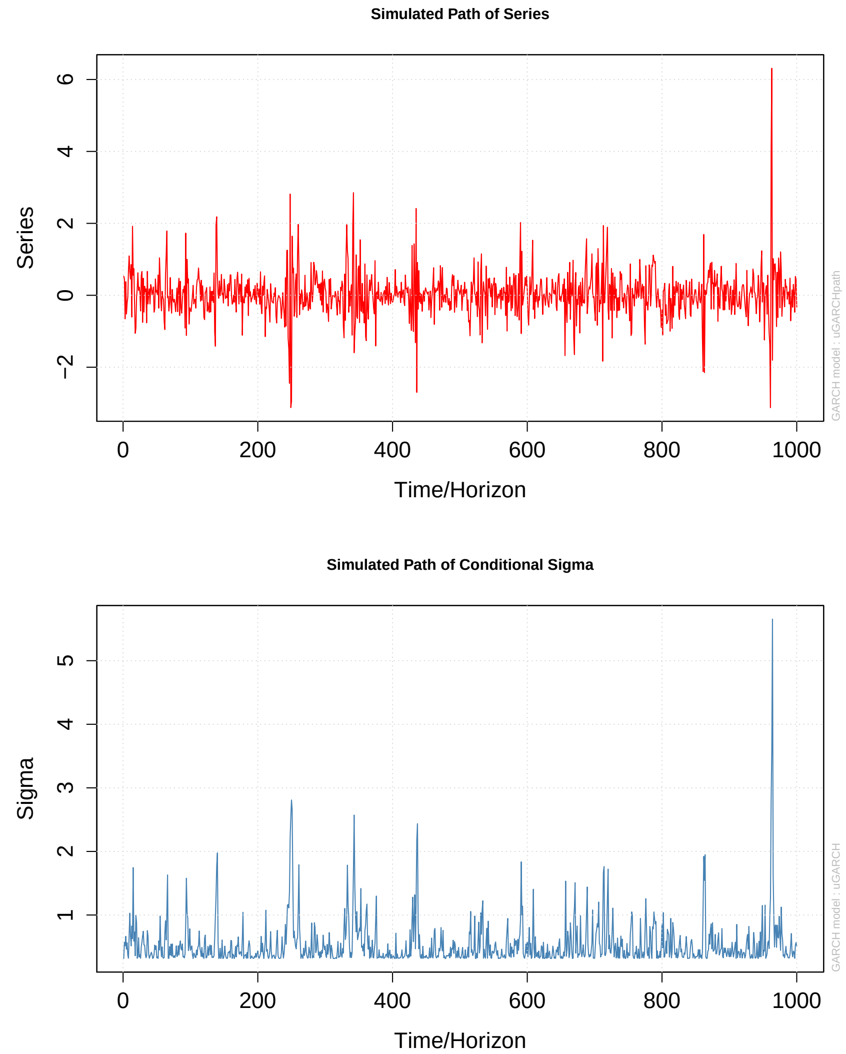

Figure 3: Simulated series and conditional volatilities

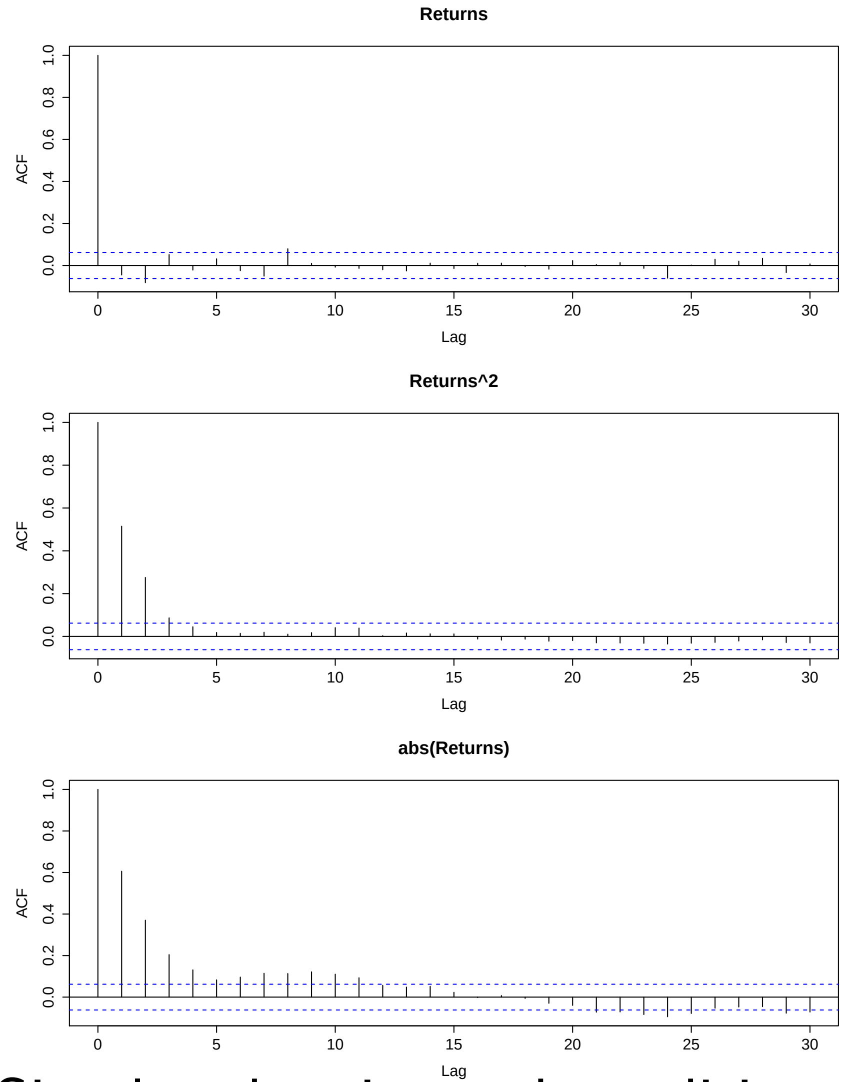

Figure 4: Simulated series and conditional volatilities

Conditional expectation of e_t

- Consider an ARCH(1) model,

$$\begin{aligned} e_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2, \\ \varepsilon_t &\sim iid(0, 1). \end{aligned}$$

- The **conditional expectation** of e_t, $E(e_t|\Omega_{t-1})$ is,

$$\begin{aligned} E(e_t|\Omega_{t-1}) &= E(\sigma_t \varepsilon_t|\Omega_{t-1}) \\ &= \sigma_t E(\varepsilon_t|\Omega_{t-1}) \\ &= \sqrt{\alpha_0 + \alpha_1 e_{t-1}^2} E(\varepsilon_t|\Omega_{t-1}) = 0. \end{aligned}$$

Conditional variance of e_t

- Consider an ARCH(1) model,

$$\begin{aligned} e_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2, \\ \varepsilon_t &\sim iid(0,1). \end{aligned}$$

- The **conditional variance** of e_t, $Var(e_t|\Omega_{t-1})$ is,

$$\begin{aligned} Var(e_t|\Omega_{t-1}) &= Var(\sigma_t \varepsilon_t|\Omega_{t-1}) \\ &= (\alpha_0 + \alpha_1 e_{t-1}^2) Var(\varepsilon_t|\Omega_{t-1}) \\ &= \alpha_0 + \alpha_1 e_{t-1}^2 = \sigma_t^2. \end{aligned}$$

Show $\{e_t\}$ is white noise process

Assume that e_t is stationary, we can show that,

- $E(e_t) = 0$.
- $Var(e_t) = \frac{\alpha_0}{1-\alpha_1}$.
- $Cov(e_t, e_{t-k}) = 0, \; k \neq 0$.

Unconditional expectation of e_t

- Consider an ARCH(1) model,

$$\begin{aligned} e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2. \end{aligned}$$

- The unconditional expectation of e_t is,

$$\begin{aligned} E(e_t) &= E[E(e_t|\Omega_{t-1})] \\ &= E[\sigma_t E(\varepsilon_t)] = 0, \end{aligned}$$

where we have used the law of iterated expectations $E[E(X|Y)] = E[X]$ and the fact that ε_t is independent of e_{t-1}.

Unconditional variance of e_t

- Consider an ARCH(1) model,

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1) \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2.
\end{aligned}
$$

- The **unconditional variance** of e_t is,

$$
\begin{aligned}
Var(e_t) = E(e_t^2) &= E[E(e_t^2|\Omega_{t-1})] \\
&= E[\sigma_t^2 E(\varepsilon_t^2|\Omega_{t-1})] \\
&= E(\alpha_0 + \alpha_1 e_{t-1}^2) = \alpha_0 + \alpha_1 E(e_{t-1}^2),
\end{aligned}
$$

where we have used the fact that ε_t is independent of e_{t-1}.

Unconditional variance of e_t

- The unconditional variance of e_t is,

$$Var(e_t) = \alpha_0 + \alpha_1 E(e_{t-1}^2),$$

- Because e_t is a stationary process with $E(e_t) = 0$, $Var(e_t) = Var(e_{t-1}) = E(e_{t-1}^2)$. Therefore, we have

$$Var(e_t) = \alpha_0 + \alpha_1 Var(e_t),$$

which yields,

$$Var(e_t) = \frac{\alpha_0}{1 - \alpha_1},$$

Unconditional variance of e_t

- The unconditional variance of e_t is,

$$Var(e_t) = \frac{\alpha_0}{1 - \alpha_1},$$

- Since the variance of e_t must be positive, we require $0 \leq \alpha_1 < 1$.

Autocovariances of e_t

- The unconditional autocovariance of e_t is,

$$
\begin{aligned}
Cov(e_t, e_{t-k}) &= E(e_t e_{t-k}) \\
&= E[E(e_t e_{t-k}|\Omega_{t-1})] \\
&= E[e_{t-k}\sigma_t E(\varepsilon_t|\Omega_{t-1})] = 0
\end{aligned}
$$

for $k > 0$.

Kurtosis of e_t

- Under the normality assumption of ε_t, we have

$$E(e_t^4|\Omega_{t-1}) = E[\varepsilon_t^4\sigma_t^4|\Omega_{t-1}]$$
$$= \sigma_t^4 E[\varepsilon_t^4|\Omega_{t-1}] = 3(\alpha_0 + \alpha_1 e_{t-1}^2)^2$$

Therefore,

$$E(e_t^4) = E[E(e_t^4|\Omega_{t-1})] = 3E[(\alpha_0 + \alpha_1 e_{t-1}^2)^2]$$
$$= 3E[\alpha_0^2 + \alpha_1^2 e_{t-1}^4 + 2\alpha_0\alpha_1 e_{t-1}^2]$$

Tip: Kurtosis$(\varepsilon_t) = 3$.

Kurtosis of e_t

- If e_t is fourth-order stationary with $m_4 = E(e_t^4)$, then we have

$$\begin{aligned} m_4 &= 3[\alpha_0^2 + \alpha_1^2 m_4 + 2\alpha_0\alpha_1 Var(e_t)] \\ &= 3\alpha_0^2\left(1 + 2\frac{\alpha_1}{1-\alpha_1}\right) + 3\alpha_1^2 m_4 \end{aligned}$$

Consequently,

$$m_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

Kurtosis of e_t

- $m_4 = E(e_t^4)$, where

$$m_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

- This result has two important implications:
  - Since the fourth moment of e_t is positive, $1 - 3\alpha_1^2 > 0$, that is, $0 \leq \alpha_1^2 < 1/3$.
  - The unconditional kurtosis of e_t is

$$\frac{E(e_t^4)}{[Var(e_t)]^2} = 3\frac{\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \times \frac{(1-\alpha_1)^2}{\alpha_0^2}$$
$$= 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3$$

Kurtosis of e_t

- The unconditional kurtosis of e_t is greater than 3.
- The excess kurtosis of e_t is positive and the tail distribution of e_t is heavier than that of a normal distribution.
- The shock e_t of a conditional Gaussian ARCH(1) model is more likely than a Gaussian white noise series to produce "outliers".

AR representation of ARCH(1)

Let e_t be an ARCH(1) process

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2, \\
\varepsilon_t &\sim N(0,1).
\end{aligned}
$$

It holds that

- e_t^2 is an AR(1) process with $e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \eta_t$.
- $\eta_t = e_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1)$ is white noise.

Proof: $e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \eta_t$

- By adding $e_t^2 - \sigma_t^2$ to both sides of the volatility equation

$$
\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2, \\
\sigma_t^2 + e_t^2 - \sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + e_t^2 - \sigma_t^2, \\
e_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + e_t^2 - \sigma_t^2, \\
e_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \eta_t,
\end{aligned}
$$

AR representation of ARCH(1)

- ARCH(1) can be equivalently expressed as an $AR(1)$ for e_t^2.

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + e_t^2 - \sigma_t^2,$$
$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \eta_t,$$

where the error term, η_t represents the volatility surprise,

$$\eta_t = e_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1).$$

AR representation of ARCH(1)

- The error term, η_t represents the volatility surprise,

$$\eta_t = e_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1).$$

Note that η_t is a new white noise process:

$$E(\eta_t) = 0$$

$$E(\eta_t \eta_s) = \begin{cases} \text{constant} & \text{for } t = s, \\ 0 & \text{otherwise} . \end{cases}$$

Proof: $E(\eta_t) = 0$

Note that $\eta_t = e_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1)$.

$$
\begin{aligned}
E[\eta_t] &= E[\sigma_t^2(\varepsilon_t^2 - 1)] \\
&= E\{\sigma_t^2 E(\varepsilon_t^2 - 1|\Omega_{t-1})\} \\
&= 0
\end{aligned}
$$

Proof: $Var(\eta_t) = \text{constant}$

Note that $\eta_t = e_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1)$.

$$\begin{aligned} Var(\eta_t) &= E[\sigma_t^4(\varepsilon_t^2 - 1)^2] \\ &= E\{\sigma_t^4 E[(\varepsilon_t^2 - 1)^2|\Omega_{t-1}]\} \end{aligned}$$

Recall that $E(\varepsilon_t^4|\Omega_{t-1}) = 3$ and $E(\varepsilon_t^2|\Omega_{t-1}) = 1$.

$$\begin{aligned} Var(\eta_t) &= 2E[\sigma_t^4] = 2E[(\alpha_0 + \alpha_1 e_{t-1}^2)^2] \\ &= 2E[\alpha_0^2 + \alpha_1^2 e_{t-1}^4 + 2\alpha_0\alpha_1 e_{t-1}^2] \end{aligned}$$

Proof: $Var(\eta_t) = \text{constant}$

$$Var(\eta_t) = 2E[\alpha_0^2 + \alpha_1^2 e_{t-1}^4 + 2\alpha_0\alpha_1 e_{t-1}^2]$$

Recall that we have shown

$$E(e_t^2) = \frac{\alpha_0}{1-\alpha_1},$$
$$E(e_t^4) = m_4 < \infty$$

Therefore,

$$Var(\eta_t) = 2\alpha_0^2 + 2\alpha_1^2 m_4 + \frac{4\alpha_0^2\alpha_1}{1-\alpha_1}$$

Proof: $Cov(\eta_t, \eta_s) = 0$, $t \neq s$

Without loss of generality, we assume $t < s$.

$$\begin{aligned} Cov(\eta_t, \eta_s) &= E[\sigma_t^2(\varepsilon_t^2 - 1)\sigma_s^2(\varepsilon_s^2 - 1)] \\ &= E\{\sigma_t^2(\varepsilon_t^2 - 1)\sigma_s^2 E[\varepsilon_s^2 - 1|\Omega_{s-1}]\} \\ &= 0 \end{aligned}$$

AR model with ARCH(1) Disturbance

- An AR(p) process for an observed variable r_t takes the form

$$r_t = \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_p r_{t-p} + e_t$$

- We assume that e_t is an ARCH(1) process

$$e_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0, 1)$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2.$$

AR model with ARCH(1) Disturbance

- Recall that for an ARCH(1) model, $E(e_t) = 0$.
- The **unconditional mean** of r_t is,

$$E(r_t) = \phi_0 + \phi_1 E(r_{t-1}) + ... + \phi_p E(r_{t-p}),$$

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1 - ... - \phi_p}.$$

AR model with ARCH(1) Disturbance

- Recall that for an ARCH(1) model, $Var(e_t) = \frac{\alpha_0}{1-\alpha_1}$.
- Using $\phi_0 = \mu(1 - \phi_1 - ... - \phi_p)$, we can rewrite AR(p) model as,

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + ... + \phi_p(r_{t-p} - \mu) + e_t$$

AR model with ARCH(1) Disturbance

- Multiplying the prior equation by $r_t - \mu$, we have,

$$(r_t - \mu)^2 = \phi_1(r_{t-1} - \mu)(r_t - \mu) + \dots \\ + \phi_p(r_{t-p} - \mu)(r_t - \mu) + e_t(r_t - \mu)$$

- Taking expectation on both sides of the prior equation, the **unconditional variance** of r_t is,

$$Var(r_t) = \gamma_0 = \phi_1\gamma_1 + \dots + \phi_p\gamma_p + \frac{\alpha_0}{1 - \alpha_1}$$

AR model with ARCH(1) Disturbance

- Recall that $E(e_t|\Omega_{t-1}) = 0$.
- The **conditional mean** of r_t, $E(r_t|\Omega_{t-1})$ is given by,

$$E(r_t|\Omega_{t-1}) = \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_p r_{t-p}$$

AR model with ARCH(1) Disturbance

- Recall that $Var(e_t|\Omega_{t-1}) = \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2$.
- The **conditional variance** of r_t, $Var(r_t|\Omega_{t-1})$ is,

$$\begin{aligned} Var(r_t|\Omega_{t-1}) &= E[(r_t - E(r_t|\Omega_{t-1}))^2|\Omega_{t-1}] \\ &= E(e_t^2|\Omega_{t-1}) = \alpha_0 + \alpha_1 e_{t-1}^2 \end{aligned}$$

ARCH(p) model

- The ARCH(p) model is given by,

$$
\begin{aligned}
r_t &= \mu_t + e_t, \\
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1), \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + ... + \alpha_p e_{t-p}^2,
\end{aligned}
$$

where μ_t can be any adapted model for the conditional mean.

ARCH(p) model

- μ_t can be any adapted model for the conditional mean.
- A model is adapted if everything required to model the mean at time t is known at time $t-1$.
- Standard examples of adapted mean processes include a constant mean or anything in the family of ARMA processes or any exogenous regressors known at time $t-1$.

ARCH(p) model

- The ARCH(p) model is given by,

$$\begin{aligned} r_t &= \mu_t + e_t, \\ e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1), \\ \sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_p e_{t-p}^2, \end{aligned}$$

- Distribution of ε_t: Standard normal, standardized Student-t, generalized error dist (ged), or their skewed counterparts.

ARCH(p) model

- The ARCH(p) model is given by,

$$
\begin{aligned}
r_t &= \mu_t + e_t, \\
e_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_p e_{t-p}^2.
\end{aligned}
$$

- The key feature of this model is that the variance of the shock, e_t , is time varying and depends on the past p shocks, e_{t-1}, e_{t-2}, ..., e_{t-p} through their squares.
- σ_t^2 is the time $t-1$ conditional variance and it is in the time $t-1$ information set Ω_{t-1}.

ARCH(p) model

- The ARCH(p) model is given by,

$$
\begin{aligned}
r_t &= \mu_t + e_t, \\
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1), \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_p e_{t-p}^2,
\end{aligned}
$$

where we impose the constraints $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i > 0$ to avoid negative variance.

AR representation of ARCH(p)

- Define $\eta_t = e_t^2 - \sigma_t^2$, which represents the volatility surprise,
- Note that η_t is a new white noise process:

$$E(\eta_t) = 0$$

$$E(\eta_t \eta_s) = \begin{cases} \text{constant} & \text{for } t = s, \\ 0 & \text{otherwise} . \end{cases}$$

AR representation of ARCH(p)

- The ARCH model then becomes,

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_m e_{t-p}^2 + \eta_t,$$

which is in the form of an AR(p) model for e_t^2.

- If an ARCH effect is found to be significant, one can use the PACF of e_t^2 to determine the ARCH order.

Pure ARCH(p) model estimation

- The pure ARCH(p) model, which is the model with constant mean of 0, is given by,

$$e_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_p e_{t-p}^2.$$

- For the likelihood function of an ARCH(p) model, given a sample of size T, we consider two cases,
  - assume ε_t follows the standard normal distribution;
  - assume ε_t follows the standardized Student-t distribution.

Pure ARCH(p) model estimation

- Under the normality assumption, $\varepsilon_t \sim N(0,1)$,

$$\begin{aligned} f(e_T, e_{T-1}, ..., e_1) &= f(e_T|\Omega_{T-1})f(e_{T-1}|\Omega_{T-2}) \\ &\quad ...f(e_{p+1}|\Omega_p)f(e_p, e_{p-1}, ..., e_1) \\ &= \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp(-\frac{(e_t - 0)^2}{2\sigma_t^2}) \\ &\quad \times f(e_p, e_{p-1}, ..., e_1), \end{aligned}$$

where we have used the fact that the conditional mean and variance of e_t, i.e. $E(e_t|\Omega_{t-1}) = 0$ and $Var(e_t|\Omega_{t-1}) = \sigma_t^2$.

Pure ARCH(p) model estimation

- Since the exact form of $f(e_p, e_{p-1}, ..., e_1)$ is complicated, sometimes it is dropped from the likelihood function, especially when the sample size is sufficiently large.
- This results in using the conditional likelihood function,

$$f(e_T, e_{T-1}, ..., e_1) = \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp(-\frac{(e_t - 0)^2}{2\sigma_t^2}),$$

where $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + ... + \alpha_p e_{t-p}^2$ can be evaluated recursively.

Pure ARCH(p) model estimation

- ε_t follows a standardized Student-t distribution. The probability density of ε_t is

$$f(\varepsilon_t;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi}}\left(1+\frac{\varepsilon_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}},$$

where $\nu > 2$ and $\Gamma(x)$ is the gamma function,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx.$$

Student-t

If X follows a Student-t distribution with ν degrees of freedom ($X \sim t_\nu$),

- its probability density is,

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

- its mean is

$$E(x) = 0.$$

- its variance is,

$$Var(x) = \frac{\nu}{\nu-2}, \quad \nu > 2.$$

Change-of-Variable Technique

- Let X be a continuous random variable with a generic pdf $f(x)$ defined over the support $c_1 < x < c_2$. And let $Y = u(X)$ be a continuous, increasing function of X with inverse function $X = v(Y)$. Then

$$f_Y(y) = F_Y'(y) = f_X(v(y)) \cdot v'(y)$$

  for $d_1 = u(c_1) < y < u(c_2) = d_2$.
- See `https://newonlinecourses.science.psu.edu/stat414/node/157/`.

Standardized Student-t

- If Y follows a standardized Student-t distribution with ν degrees of freedom. Then $Y = X/\sqrt{\nu/(\nu-2)}$. The probability density of Y is

$$f(y;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi}}\left(1+\frac{y^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \quad \nu > 2.$$

Location-scale t distribution

If $X \sim t_\nu$, then $Y = \mu + \sigma X \sim t_\nu(\mu, \sigma^2)$, where μ and σ are location and scale parameters,

- the probability density function of Y is,

$$f(y;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}\sigma}\left(1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}.$$

- the expectation of Y is

$$E(Y) = \mu$$

- the variance of Y is

$$Var(Y) = \frac{\nu}{\nu-2}\sigma^2$$

Pure ARCH(p) model estimation

- ε_t follows a standardized Student-t distribution.
- Using $e_t = \sigma_t \varepsilon_t$, we obtain the conditional likelihood function of e_t as

$$
\begin{aligned}
& f(e_T, e_{T-1}, \ldots, e_1) \\
& \quad = \prod_{t=p+1}^{T} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{(\nu-2)\pi}} \frac{1}{\sigma_t} \left(1 + \frac{e_t^2}{(\nu-2)\sigma_t^2}\right)^{-(\nu+1)/2},
\end{aligned}
$$

Testing for ARCH Effects

- Compute the residuals e_t from mean equation regression, where $e_t = r_t - \mu_t$;
- The squared series e_t^2 is then used to check for conditional heteroscedasticity, which is also known for ARCH effects.
- Two tests are available: Ljung-Box test and Engle's Lagrange Multiplier test.

Ljung-Box test

- Apply the Ljung-Box statistics $Q(p)$ to the $\{e_t^2\}$ series;
- The null is the first p lags of ACF of the e_t^2 series are zero.

Engle's Lagrange Multiplier test

- Consider the linear regression,

  $$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_p e_{t-p}^2 + \nu_t, \quad t = p+1, \ldots, T,$$

  where ν_t denotes the error term, p is a prespecified positive integer, and T is the sample size.
- The null (No ARCH) is
  $\alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$;
  The alternative (ARCH) is at least one $\alpha_i \neq 0$.
- This test is equivalent to the usual F statistic.

Engle's Lagrange Multiplier test

- Let $SSR_0 = \sum_{t=p+1}^{T}(e_t^2 - \bar{\omega})^2$, where $\bar{\omega} = (1/T)\sum_{t=1}^{T} e_t^2$ is the sample mean of e_t^2 and $SSR_1 = \sum_{t=p+1}^{T} \hat{\nu}_t^2$, where $\hat{\nu}_t$ is the least-squares residual of the prior linear regression. Then we have

$$LM_{ARCH} = \frac{(SSR_0 - SSR_1)/p}{SSR_1/(T - 2p - 1)}$$

- Under H_0: (No ARCH)

$$LM_{ARCH} \xrightarrow{d} \chi^2(p)$$

Model Checking

- For a properly specified ARCH model, the standardized residuals

$$\varepsilon_t = \frac{e_t}{\sigma_t}$$

form a sequence of iid random variables.

- One can check the adequacy of a fitted ARCH model by examining the series $\{\varepsilon_t\}$.
- The Ljung-Box statistics of ε_t can be used to check the adequacy of the mean equation.

Model Checking

- The Ljung-Box statistics of ε_t^2 can be used to check the adequacy of the volatility equation.
- The skewness, kurtosis and quantile-to-quantile plot (QQ plot) of ε_t can be used to check the adequacy of the distribution assumption.

Advantages of ARCH models

- Simplicity of the model
- Generate volatility clustering
- Heavy tails (high kurtosis)

Weaknesses of ARCH models

- Symmetric between positive & negative prior returns
  - The model assumes that positive and negative shocks have the same effects on volatility because it depends on the square of the previous shocks.
  - In practice, it is well known that the price of a financial asset responds differently to positive and negative shocks.

Weaknesses of ARCH models

- Restrictive
  - For example, α_1^2 of an ARCH(1) model must be in the interval $[0, 1/3]$ if the series has a finite fourth moment.
  - The constraint becomes complicated for higher order ARCH models.
  - In practice, it limits the ability of ARCH models with Gaussian innovations to capture excess kurtosis.

Weaknesses of ARCH models

- Provides no explanation
  - The ARCH model does not provide any new insight for understanding the source of variations of a financial time series.
  - It merely provides a mechanical way to describe the behavior of the conditional variance.
  - It gives no indication about what causes such behavior to occur.
- Not sufficiently adaptive in prediction
  - ARCH models are likely to overpredict the volatility because they respond slowly to large isolated shocks to the return series.

Building an ARCH model

- Modeling the mean effect and testing for ARCH effects
  H_0: no ARCH effects versus H_a : ARCH effects
  Use Q-statistics of squared residuals; Engle's LM test
- Order determination
  Use PACF of the squared residuals. (In practice, simply try some reasonable order).
- Estimation: Conditional MLE

Building an ARCH model

- Model checking: Q-stat of standardized residuals and squared standardized residuals. Skewness & Kurtosis of standardized residuals. R provides many plots for model checking and for presenting the results.
- Software: We use R with the package fGarch. (Other software available).

Forecasting with ARCH(p) model

The ARCH(p) model is,

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_p e_{t-p}^2
\end{aligned}
$$

At the forecast origin h, the 1-step-ahead forecast of σ_{h+1}^2 is

$$
\begin{aligned}
\sigma_h^2(1) = E[\sigma_{h+1}^2 | \Omega_h] &= \sigma_{h+1}^2 \\
&= \alpha_0 + \alpha_1 e_h^2 + \ldots + \alpha_p e_{h+1-p}^2.
\end{aligned}
$$

Forecasting with ARCH(p) model

The ARCH(p) model is,

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_p e_{t-p}^2
\end{aligned}
$$

At the forecast origin h, the 2-step-ahead forecast is,

$$
\begin{aligned}
\sigma_h^2(2) &= E[\sigma_{h+2}^2 | \Omega_h] \\
&= \alpha_0 + \alpha_1 E[e_{h+1}^2 | \Omega_h] + \alpha_2 e_h^2 \ldots + \alpha_p e_{h+2-p}^2 \\
&= \alpha_0 + \alpha_1 \sigma_h^2(1) + \alpha_2 e_h^2 \ldots + \alpha_p e_{h+2-p}^2
\end{aligned}
$$

Forecasting with ARCH(p) model

The ARCH(p) model is,

$$\begin{aligned} e_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_p e_{t-p}^2 \end{aligned}$$

At the forecast origin h, the ℓ-step-ahead forecast is,

$$\begin{aligned} \sigma_h^2(\ell) &= E[\sigma_{h+\ell}^2 | \Omega_h] \\ &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \sigma_h^2(\ell - i), \end{aligned}$$

where $\sigma_h^2(\ell - i) = e_{h+\ell-i}^2$ if $\ell - i \leq 0$.

GARCH models

- The Generalized Autoregressive Conditional Heteroskedasticity (GARCH(m,s)) model is given by,

$$r_t = \mu_t + e_t$$
$$e_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1),$$
$$\sigma_t^2 = \omega + \sum_{i=1}^{m} \alpha_i e_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2,$$

GARCH models

- The GARCH(m, s) model is given by,

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

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i e_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2,$$

  - we put the constraints $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$ to avoid the negative variance.
  - we put the constraints $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$ to ensure that the unconditional variance of e_t is finite.

GARCH(1,1) model

- The GARCH(1,1) is given by,

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

- Weakly stationarity conditions: $0 \leq \alpha_1, \beta_1 \leq 1$, $\alpha_1 + \beta_1 < 1$.
- Volatility clustering: a large e_{t-1}^2 or σ_{t-1}^2 gives rise to a large σ_t^2. This means that a large e_{t-1}^2 tends to be followed by another large e_t^2, generating the behavior of volatility clustering.
- GARCH(1,1) is very often the best model (See paper by Hansen and Lund, JOE)

Simulate GARCH(1,1) model

- Simulate from an GARCH(1,1) model, where

$$e_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1),$$
$$\sigma_t^2 = 0.1 + 0.1 e_{t-1}^2 + 0.7 \sigma_{t-1}^2,$$

  This is an ARMA form for the squared series e_t^2.
  Use it to understand properties of GARCH models, e.g. moment equations, forecasting, etc.

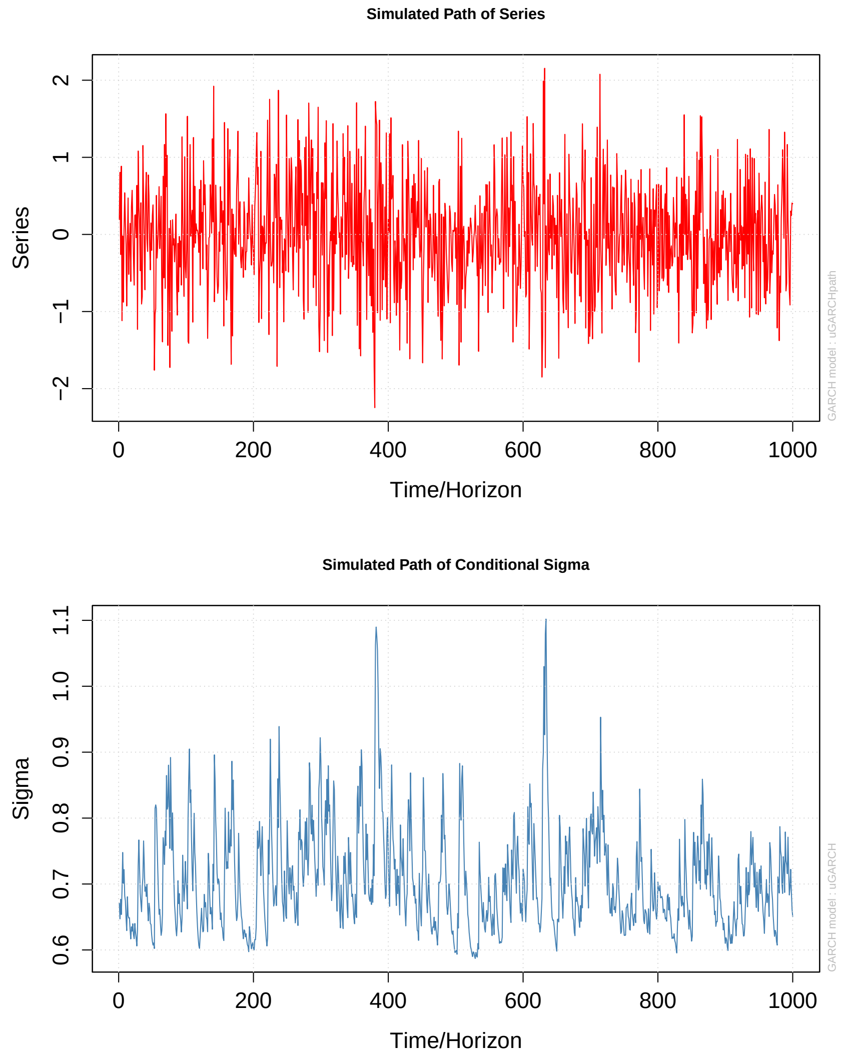

Figure 5: Simulated series and conditional volatilities

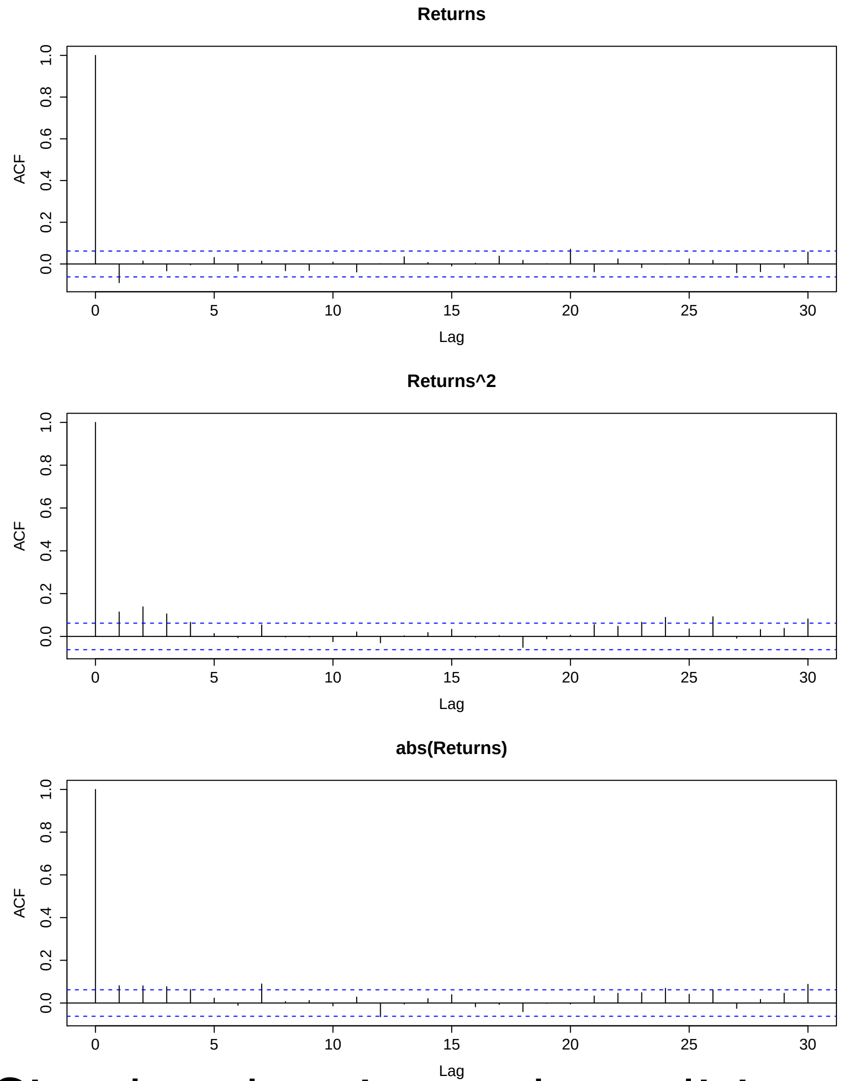

Figure 6: Simulated series and conditional volatilities

Conditional expectation

- Consider an GARCH(1,1) model,

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \\
\varepsilon_t &\sim iid(0,1).
\end{aligned}
$$

- The **conditional expectation** of e_t, $E(e_t|\Omega_{t-1})$ is,

$$
\begin{aligned}
E(e_t|\Omega_{t-1}) &= E(\sigma_t \varepsilon_t|\Omega_{t-1}) \\
&= \sigma_t E(\varepsilon_t|\Omega_{t-1}) \\
&= \sqrt{\alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2} E(\varepsilon_t|\Omega_{t-1}) = 0.
\end{aligned}
$$

Conditional variance

- Consider an GARCH(1,1) model,

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \\
\varepsilon_t &\sim iid(0,1).
\end{aligned}
$$

- The **conditional variance** of e_t, $Var(e_t|\Omega_{t-1})$ is,

$$
\begin{aligned}
Var(e_t|\Omega_{t-1}) &= Var(\sigma_t \varepsilon_t|\Omega_{t-1}) \\
&= (\alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2) Var(\varepsilon_t|\Omega_{t-1}) \\
&= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \sigma_t^2.
\end{aligned}
$$

Unconditional expectation

- Consider an GARCH(1,1) model,

$$e_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$
$$\varepsilon_t \sim iid(0, 1).$$

- The **unconditional expectation** of e_t is,

$$E(e_t) = E[E(e_t|\Omega_{t-1})]$$
$$= E[\sigma_t E(\varepsilon_t)] = 0,$$

where we have used the law of iterated expectations $E[E(X|Y)] = E[X]$ and the fact that ε_t is independent of e_{t-1}.

Re-parameterization of GARCH(m,s)

- Let $\eta_t = e_t^2 - \sigma_t^2$.
- It's easy to check that $\{\eta_t\}$ is martingale difference series, i.e. $E(\eta_t) = 0$ and $cov(\eta_t, \eta_{t-j}) = 0$ for $j \geq 1$. (Why?)
- However, $\{\eta_t\}$ in general is not an iid sequence.

Re-parameterization of GARCH(m, s)

- The GARCH(m, s) model is given by,

$$e_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1),$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i e_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2,$$

- By plugging $\sigma_{t-i}^2 = e_{t-i}^2 - \eta_{t-i}$, (i=0,...,s) into the prior equation, the GARCH model becomes

$$e_t^2 = \alpha_0 + \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) e_{t-i}^2 + \eta_t - \sum_{j=1}^{s} \beta_j \eta_{t-j}.$$

- It can be regarded as an application of the ARMA idea to the squared series e_t^2.

Unconditional variance

- The GARCH(1,1) model becomes

$$e_t^2 = \alpha_0 + (\alpha_1 + \beta_1)e_{t-1}^2 + \eta_t - \beta_1\eta_{t-1}.$$

- Using the unconditional mean of an ARMA model, the **unconditional variance** of GARCH(1,1) is,

$$E(e_t^2) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}.$$

Kurtosis

The GARCH(1,1) is given by,

$$e_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

Assume $\varepsilon_t \sim iid\ N(0,1)$, we have $E(\varepsilon_t^4) = 3$. We have the following

- $Var(e_t) = E(e_t^2) = E(\sigma_t^2) = \frac{\alpha_0}{1-(\alpha_1+\beta_1)}$;
- $E(e_t^4) = 3E(\sigma_t^4)$ provided that $E(\sigma_t^4)$ exists;

Proof of $E(e_t^2) = E\{\sigma_t^2\}$ and $E(e_t^4) = 3E\{\sigma_t^4\}$

$$
\begin{aligned}
E(e_t^2) &= E\{E(e_t^2|\Omega_{t-1})\} \\
&= E\{\sigma_t^2 E(\varepsilon_t^2|\Omega_{t-1})\} \\
&= E\{\sigma_t^2\}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
E(e_t^4) &= E\{E(e_t^4|\Omega_{t-1})\} \\
&= E\{\sigma_t^4 E(\varepsilon_4^2|\Omega_{t-1})\} \\
&= 3E\{\sigma_t^4\}
\end{aligned}
$$

Kurtosis

- Taking the square of the volatility model, we have,

$$\sigma_t^4 = \alpha_0^2 + \alpha_1^2 e_{t-1}^4 + \beta_1^2 \sigma_{t-1}^4 \\ + 2\alpha_0\alpha_1 e_{t-1}^2 + 2\alpha_0\beta_1\sigma_{t-1}^2 + 2\alpha_1\beta_1\sigma_{t-1}^2 e_{t-1}^2$$

Taking expectation of the equations, we have

$$E(\sigma_t^4) = \frac{\alpha_0^2(1+\alpha_1+\beta_1)}{[1-(\alpha_1+\beta_1)][1-2\alpha_1^2-(\alpha_1+\beta_1)^2]}$$

provided that $0 \leq \alpha_1 + \beta_1 < 1$ and $1 - 2\alpha_1^2 - (\alpha_1+\beta_1)^2 > 0$.

Kurtosis

- Heavy tails: Under the normality assumption of ε_t, the Kurtosis of e_t, if it exists, is then

$$\frac{E(e_t^4)}{[E(e_t^2)]^2} = \frac{3[1-(\alpha_1+\beta_1)^2]}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2} > 3$$

- When we use GARCH models, we can model both the conditional heteroskedasticity and the heavy-tailed distributions of financial markets data.

Kurtosis

- Many financial time series have tails that are heavier than implied by a GARCH process with Gaussian $\{\varepsilon_t\}$.
- To handle such data, one can assume that, instead of being Gaussian white noise, $\{\varepsilon_t\}$ is an i.i.d. white noise process with a heavy-tailed distribution.

Estimation of AR-GARCH model

- Assume the data $\{r_1, ..., r_T\}$ come from the AR(1)-GARCH(1,1) model,

$$
\begin{aligned}
r_t &= \phi r_{t-1} + e_t, \\
e_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2,
\end{aligned}
$$

  where $\varepsilon_t \sim i.i.d.\ N(0,1)$. Give the conditional log-likelihood function for $\theta = (\phi, \alpha_0, \alpha_1, \beta_1)'$.
- Given Ω_{t-1}, r_{t-1} is a constant and $r_t \sim N(\phi r_{t-1}, \sigma_t^2)$ for every $t > 0$.

Estimation of AR-GARCH model

- So the conditional likelihood function for θ at t is,

$$f(r_t|\Omega_{t-1}) = \frac{1}{\sqrt{2\pi}\sigma_t}\exp\left\{-\frac{1}{2}\frac{(r_t-\phi r_{t-1})^2}{\sigma_t^2}\right\},$$

where

$$\sigma_t^2 = \alpha_0 + \alpha_1(r_{t-1}-\phi r_{t-2})^2 + \beta_1\sigma_{t-1}^2$$

- The log-likelihood function is:

$$l_t(\theta) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{1}{2}\frac{(r_t-\phi r_{t-1})^2}{\sigma_t^2}$$

Estimation of AR-GARCH model

- For the sample data $\{r_1, \ldots, r_T\}$, we have

$$L(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\theta)$$

- We maximize $L(\theta)$ to get the MLE for θ.

Forecasting with GARCH(1, 1) model

- Consider a GARCH(1, 1) model,

$$\begin{aligned} e_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned}$$

- Assume that the forecast origin is h. For 1-step ahead forecast, we have

$$\begin{aligned} \sigma_h^2(1) = E(\sigma_{h+1}^2 | \Omega_h) &= \sigma_{h+1}^2 \\ &= \alpha_0 + \alpha_1 e_h^2 + \beta_1 \sigma_h^2, \end{aligned}$$

where h is the forecast origin.

Forecasting with GARCH(1, 1) model

- For multiple-step ahead forecast, use $e_t^2 = \sigma_t^2 \varepsilon_t^2$ and rewrite the model as,

$$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 + \alpha_1 \sigma_t^2(\varepsilon_t^2 - 1)$$

- When $t = h + 1$, the equation becomes,

$$\sigma_{h+2}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_{h+1}^2 + \alpha_1 \sigma_{h+1}^2(\varepsilon_{h+1}^2 - 1)$$

- For 2-step ahead volatility forecast, the best linear predictor of σ_{h+2}^2 is

$$\begin{aligned}\sigma_h^2(2) &= E[\sigma_{h+2}^2|\Omega_h] \\ &= \alpha_0 + (\alpha_1 + \beta_1)E[\sigma_{h+1}^2|\Omega_h] + \alpha_1 E[\sigma_{h+1}^2(\varepsilon_{h+1}^2 - 1)|\Omega_h]\end{aligned}$$

Forecasting

- For 2-step ahead volatility forecast, the best linear predictor of σ^2_{h+2} is

$$\sigma^2_h(2) = \alpha_0 + (\alpha_1 + \beta_1)E[\sigma^2_{h+1}|\Omega_h] + \alpha_1 E[\sigma^2_{h+1}(\varepsilon^2_{h+1} - 1)|\Omega_h]$$

- Since $E(\varepsilon^2_{h+1} - 1|\Omega_h) = 0$, the 2-step-ahead volatility forecast at the forecast origin h satisfies the equation,

$$\sigma^2_h(2) = \alpha_0 + (\alpha_1 + \beta_1)\sigma^2_h(1)$$

Forecasting

- In general, we have

$$\sigma_h^2(\ell) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(\ell - 1), \quad \ell > 1$$

  This result is exactly the same as that of an ARMA(1,1) model with AR polynomial $1 - (\alpha_1 + \beta_1)L$.
- By repeated substitutions, we obtain that the ℓ-step-ahead forecast can be written as,

$$\sigma_h^2(\ell) = \frac{\alpha_0[1 - (\alpha_1 + \beta_1)^{\ell-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{\ell-1}\sigma_h^2(1)$$

Forecasting

- By repeated substitutions, we obtain that the ℓ-step-ahead forecast can be written as,

$$\sigma_h^2(\ell) = \frac{\alpha_0[1-(\alpha_1+\beta_1)^{\ell-1}]}{1-\alpha_1-\beta_1} + (\alpha_1+\beta_1)^{\ell-1}\sigma_h^2(1)$$

- Therefore,

$$\sigma_h^2(\ell) \to \frac{\alpha_0}{1-\alpha_1-\beta_1}, \qquad \text{as } \ell \to \infty$$

provided $\alpha_1 + \beta_1 < 1$

Forecasting

- Provided $\alpha_1 + \beta_1 < 1$,

$$\sigma_h^2(\ell) \to \frac{\alpha_0}{1 - \alpha_1 - \beta_1}, \qquad \text{as } \ell \to \infty$$

- The multistep-ahead volatility forecasts of a GARCH(1,1) model converge to the unconditional variance of e_t as the forecast horizon increases to infinity provided that $Var(e_t)$ exists.

Drawbacks of GARCH model

- GARCH(1, 1) model,

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2
\end{aligned}
$$

  The sign of ε_t does not have different impacts on σ_t^2.
- One of the main weaknesses of the GARCH models is that the conditional variance responds symmetrically to "positive "and "negative "past innovations;

EGARCH(m,s) model

- Nelson (1991) proposed the exponential GARCH (EGARCH) models;
- An EGARCH(m,s) model is given by,

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t, \\
\ln(\sigma_t^2) &= \alpha_0 + \frac{1 + \beta_1 L + \ldots + \beta_{s-1} L^{s-1}}{1 - \alpha_1 L - \ldots - \alpha_m L^m} g(\varepsilon_{t-1}),
\end{aligned}
$$

  where $g(\varepsilon_t)$ is the weighted innovation.
- Since log-transformation is taken, there is no positiveness requirement on the model parameters.

Weighted innovation

- $g(\varepsilon_t)$ is the weighted innovation,

$$g(\varepsilon_t) = \theta\varepsilon_t + \gamma[|\varepsilon_t| - E(|\varepsilon_t|)],$$

  where both ε_t and $|\varepsilon_t| - E(|\varepsilon_t|)$ are zero-mean iid sequences with continous distributions.
- $E[g(\varepsilon_t)] = 0$.

Weighted innovation

- The asymmetry of $g(\varepsilon_t)$ can be rewritten as,

$$g(\varepsilon_t) = \begin{cases} (\theta+\gamma)\varepsilon_t - \gamma E(|\varepsilon_t|), & \text{if } \varepsilon_t \geq 0, \\ (\theta-\gamma)\varepsilon_t - \gamma E(|\varepsilon_t|), & \text{if } \varepsilon_t < 0. \end{cases}$$

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

Weighted innovation

- $g(\varepsilon_t)$ is the weighted innovation,

$$g(\varepsilon_t) = \theta\varepsilon_t + \gamma[|\varepsilon_t| - E(|\varepsilon_t|)],$$

- For the standard Gaussian random variable ε_t, $E(|\varepsilon_t|)] = \sqrt{2/\pi}$.
- For the standardized Student-t distribution,

$$E(|\varepsilon_t|)] = \frac{2\sqrt{\nu - 2}\Gamma[(\nu + 1)/2]}{(\nu - 1)\Gamma(\nu/2)\sqrt{\pi}}$$

EGARCH(1,1) model

- Under normality assumption of ε_t, $E(|\varepsilon_t|) = \sqrt{2/\pi}$ and the model becomes,

$$e_t = \sigma_t \varepsilon_t$$

$$(1 - \alpha L)\ln(\sigma_t^2) = \begin{cases} \alpha_* + (\gamma + \theta)\varepsilon_{t-1} & \text{if } \varepsilon_{t-1} \geq 0, \\ \alpha_* + (\gamma - \theta)(-\varepsilon_{t-1}) & \text{if } \varepsilon_{t-1} < 0, \end{cases}$$

where $\alpha_* = (1 - \alpha)\alpha_0 - \sqrt{2/\pi}\gamma$.

EGARCH(1,1) model

- We have,

$$\sigma_t^2 = \begin{cases} \sigma_{t-1}^{2\alpha} \exp(\alpha_*) \exp[(\gamma + \theta)\varepsilon_{t-1}], & \text{if } \varepsilon_{t-1} \geq 0, \\ \sigma_{t-1}^{2\alpha} \exp(\alpha_*) \exp[(\gamma - \theta)|\varepsilon_{t-1}|], & \text{if } \varepsilon_{t-1} < 0, \end{cases}$$

- The coefficients $(\gamma + \theta)$ and $(\gamma - \theta)$ show the asymmetry in response to positive and negative e_{t-1}.
- The model is, therefore, nonlinear if $\theta \neq 0$. Thus, θ is referred to as the leverage parameter.
- Since negative shocks tend to have larger impacts, we expect θ to be negative.

Alternative form of EGARCH(m,s)

- Another parameterization of EGARCH(m,s) model is,

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{s} \alpha_i \frac{|e_{t-i}| + \gamma_i e_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{m} \beta_j \ln(\sigma_{t-j}^2),$$

- A positive e_{t-i} contributes $\alpha_i(1+\gamma_i)|\varepsilon_{t-i}|$ to the log volatility, whereas a negative e_{t-i} gives $\alpha_i(1-\gamma_i)|\varepsilon_{t-i}|$, where $\varepsilon_{t-i} = e_{t-i}/\sigma_{t-j}$.
- γ_i signifies the leverage effect of e_{t-i}.
- We expect γ_i to be negative in real applications.

IGARCH model

- An IGARCH(1,1) model,

$$\begin{aligned} e_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) e_{t-1}^2 \end{aligned}$$

- For an IGARCH(1,1) model,

$$\sigma_h^2(\ell) = \sigma_h^2(1) + (\ell - 1)\alpha_0, \quad \ell \geq 1.$$

IGARCH model

- For an IGARCH(1,1) model,

$$\sigma_h^2(\ell) = \sigma_h^2(1) + (\ell - 1)\alpha_0, \quad \ell \geq 1.$$

- Effect of $\sigma_h^2(1)$ on future volatilities is persistent, and the volatility forecasts form a straight line with slope α_0. See Nelson (1990) for more info.
- Special case: $\alpha_0 = 0$. Volatility forecasts become a constant.

GARCH-in-Mean model (GARCH-M)

- Idea: Modern finance theory suggests that volatility may be related to risk premia on assets.
- The GARCH-M model allows time-varying volatility to be related to expected returns

$$
\begin{aligned}
r_t &= \mu + cg(\sigma_t) + e_t, \\
e_t &= \sigma_t \varepsilon_t, \\
\sigma_t &\sim GARCH
\end{aligned}
$$

where μ and c are constants.

GARCH-in-Mean model (GARCH-M)

- The GARCH-M model allows time-varying volatility to be related to expected returns

$$
\begin{aligned}
r_t &= \mu + c g(\sigma_t) + e_t, \\
e_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2
\end{aligned}
$$

where

$$
g(\sigma_t) = \begin{cases} \sigma_t, \\ \sigma_t^2, \\ \ln(\sigma_t^2) \end{cases}
$$

GARCH-in-Mean model (GARCH-M)

- The GARCH-M model allows time-varying volatility to be related to expected returns

$$
\begin{aligned}
r_t &= \mu + cg(\sigma_t) + e_t, \\
e_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2
\end{aligned}
$$

- The parameter c is called the risk premium parameter. The positive c indicates that the return is positively related to its volatility.

GARCH(1,1)-M model

- The GARCH(1,1)-M model is given by,

$$
\begin{aligned}
r_t &= \mu + cg(\sigma_t) + e_t, \\
e_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2
\end{aligned}
$$

  where the parameter c is referred to as risk premium parameter.
- The parameter c is expected to be positive, which indicates that the return is positively related to its volatility.

GARCH(1,1)-M model

- The GARCH(1,1)-M model is given by,

$$
\begin{aligned}
r_t &= \mu + cg(\sigma_t) + e_t, \\
e_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2
\end{aligned}
$$

- It implies that there are serial correlations in the return series r_t. These serial correlations are introduced by those in the volatility process $\{\sigma_t^2\}$.
- The existence of risk premium is, therefore, another reason that some historical stock returns have serial correlations.

The Threshold GARCH (TGARCH)

- A TGARCH(s, m) or GJR(s, m) model is defined as

$$
\begin{aligned}
r_t &= \mu_t + e_t \\
e_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{s} (\alpha_i + \gamma_i N_{t-i}) e_{t-i}^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2,
\end{aligned}
$$

where N_{t-i} is an indicator variable such that

$$
N_{t-i} = \begin{cases} 1 & \text{if } e_{t-i} < 0, \\ 0 & \text{otherwise} . \end{cases}
$$

The Threshold GARCH (TGARCH)

- A TGARCH(s, m) or GJR(s, m) model is defined as

$$
\begin{aligned}
r_t &= \mu_t + e_t \\
e_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{s} (\alpha_i + \gamma_i N_{t-i}) e_{t-i}^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2,
\end{aligned}
$$

- One expects γ_i to be positive so that prior negative returns have higher impact on the volatility.

The Asymmetric Power ARCH (APARCH) Model

- This model was introduced by Ding, Engle and Granger (1993) as a general class of volatility models.
- The basic form is

$$
\begin{aligned}
r_t &= \mu_t + e_t \\
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1) \\
\sigma_t^\delta &= \omega + \sum_{i=1}^{s} \alpha_i (|e_{t-i}| - \gamma_i e_{t-i})^\delta + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^\delta,
\end{aligned}
$$

where δ is a non-negative real number.

The Asymmetric Power ARCH (APARCH) Model

- The basic form is

$$
\begin{aligned}
r_t &= \mu_t + e_t \\
e_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0,1) \\
\sigma_t^\delta &= \omega + \sum_{i=1}^{s} \alpha_i (|e_{t-i}| - \gamma_i e_{t-i})^\delta + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^\delta,
\end{aligned}
$$

- $\delta = 2$ gives rise to the TGARCH model and $\delta \to 0$ corresponds to using $\log(\sigma_t)$.

Stochastic volatility model

- A (simple) SV model is

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t, \\
(1 - \alpha_1 L - \dots - \alpha_m L^m) \ln(\sigma_t^2) &= \alpha_0 + \nu_t,
\end{aligned}
$$

  where ε_t's are iid N(0,1), ν_t's are iid $N(0, \sigma_\nu^2)$, $\{\varepsilon_t\}$ and $\{\nu_t\}$ are independent.

Long-memory SV model

- A (simple) LMSV model is

$$
\begin{aligned}
e_t &= \sigma_t \varepsilon_t, \\
\sigma_t &= \sigma \exp(u_t/2) \\
(1-L)^d u_t &= \eta_t,
\end{aligned}
$$

where $\sigma > 0$, ε_t are iid N(0,1), η_t are iid $N(0, \sigma_\eta^2)$ are independent of ε_t and $0 < d < 0.5$.

Long-memory SV model

- The model says

$$\begin{aligned}\ln(e_t^2) &= \ln(\sigma^2) + u_t + \ln(\varepsilon_t^2), \\ &= [\ln(\sigma^2) + E(\ln(\varepsilon_t^2))] + u_t + [\ln(\varepsilon_t^2) - E(\ln(\varepsilon_t^2))] \\ &= \mu + u_t + a_t.\end{aligned}$$

Thus, the $\ln(e_t^2)$ series is a Gaussian long-memory signal plus a non- Gaussian white noise.