Financial Risk Management 2

Ta Quoc Bao

Department of Mathematics,
International University-VNUHCM

Chapter 2. Univariate Volatility Modeling

- The volatility plays a crucial role in financial risk management, it is the main measure of risk. On the other hands, the volatility is the key factor in, e.g., Investment decisions, Portfolio construction (Markowitz model) and Derivative pricing (Black-Scholes model).
- In this Chapter we focus on the estimation and forecasting of volatility for a single asset (univariate)

Chapter 2. Univariate Volatility Modeling

- The volatility plays a crucial role in financial risk management, it is the main measure of risk. On the other hands, the volatility is the key factor in, e.g., Investment decisions, Portfolio construction (Markowitz model) and Derivative pricing (Black-Scholes model).
- In this Chapter we focus on the estimation and forecasting of volatility for a single asset (univariate)

2.1. Stationary processes

- A time series is a sequence of observations in chronological order. For example: daily log returns on a stock or monthly values of the Consumer Price Index (CPI)
- A stochastic process is a sequence of random variables and can be viewed as the “theoretical” or “population” analog of a time series, conversely, a time series can be considered a sample from a stochastic process.

Denote $\{X_t, t \in I\}$ the time series, where I is a time index. For example: $I = \{1, 2, 3, ...\}; I = \{2000, 2001, 2002...2021\}$. **Equally** spaced time series are the most common in practice. This is the case of $I = \{t_1, t_2, ..., t_n\}$, where

$$\Delta = t_{i+1} - t_i$$

with Δ is a constant.

Difference from traditional Statistical Inference

- In traditional statistic inference, the data is assumed to be an i.i.d process (random sample).
- In time series, we do not need this assumption and wish to model the dependency among observations which leads to the concept of autocorrelation.

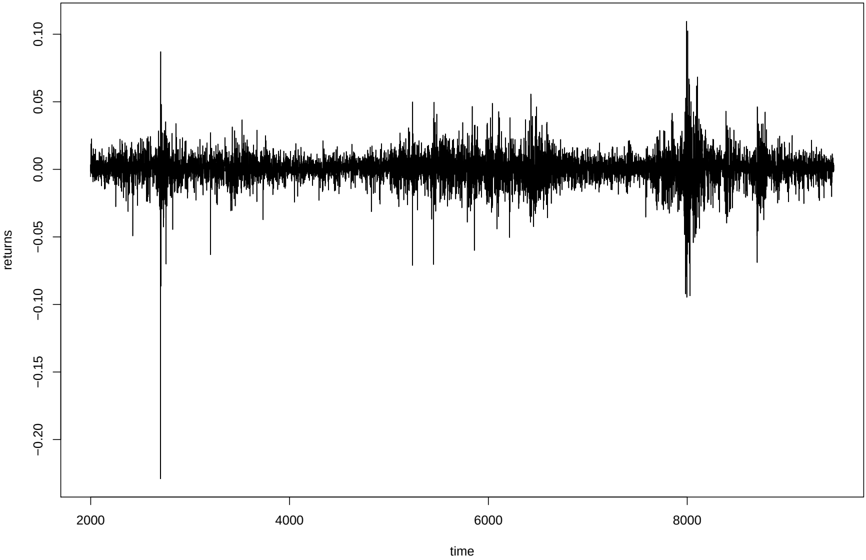
Some main problems in time series

- Formulate and estimate a parametric model for X_t (need to propose methods of estimation and model diagnostics).
- This point is related to the estimation of autoregressive (AR) or ARMA models.
- Estimation of Missing values (fill“gaps”).
- Prediction or Forecasting (“would like to know what a future value is”). For example our data is $x_1, x_2, ..., x_{100}$, we wish to forecast the next 10 values, $x_{101}, ..., x_{110}$. In this case, our forecasting horizon is 10.
- Plotting time series to observe fluctuations of time series, e.g., to find stationarity or non-stationarity, cycles, trends, outliers or interventions. Assisting in the formulation of a parametric model.

Example

Consider Financial Index SP500. The data consists of excess returns $X_t = \log(S_t) - \log(S_{t-1})$. From the plot we see the following properties of X_t

- The mean level of the process seems constant
- There are sections of the data with explosive behavior(high volatility).
- The data corresponds to a non-stationary process. (will define more detailed)
- The variance (or volatility) is not constant in time.
- No linear time series model will be available for this data.



0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
returns
2000
4000
6000
8000
time

Definition 1 (Autocovariance)

The autocovariance function a stochastic process X is defined as

$$\gamma(t,\tau) = \mathbb{E}(X_t - \mu_t)(X_{t-\tau} - \mu_{t-\tau})$$

for $\tau \in \mathbb{Z}$, where $\mu_t = \mathbb{E}(X_t)$.

- The autocovariance function is symmetric, i.e.,

  $$\gamma(t,\tau) = \gamma(t-\tau,-\tau)$$

  For special case $\tau = 0$ then $\gamma(t,0) = Var(X_t)$
- In general $\gamma(t,\tau)$ is depend on t as well as τ

Example

Find autocovariance function of Brownian motion?

Definition 2

A process is said to be strictly stationary if all aspects of its behavior are unchanged by shifts in time. Mathematically, stationarity is defined as the requirement that for every m and n the distribution of $(X_1, X_2, ..., X_n)$ and $(X_{1+m}, X_{2+m}, ..., X_{n+m})$ are the same.

Definition 3

A process is weakly stationary if its mean, variance, and covariance are unchanged by time shifts. More precisely, $X_1, X_2, ...,$ is a weakly stationary process if

(i) $\mathbb{E}(X_t) = \mu$ for all t

(ii) $Var(X_t) = \sigma^2$ (a positive finite constant) for all t

(iii) $Cov(X_t, X_s) = \gamma(\mid t - s \mid)$ for all t, s and some function γ

We see that, the mean and variance do not change with time and the covariance between two observations depends only on the lag, the time distance $\mid t - s \mid$.

- The function γ is the autocovariance function of the process and has symmetric property

$$\gamma(h) = \gamma(-h)$$

- The correlation between X_t and X_{t+h} is denoted by $\rho(h)$, function ρ is called autocorrelation function (ACF). We have $\gamma(0) = \sigma^2$ and, hence

$$\gamma(h) = \sigma^2 \rho(h) \quad \text{hence} \quad \rho(h) = \gamma(h)/\gamma(0).$$

The ACF is normalized on $[-1, 1]$. Since the process is required to be covariance stationary, the ACF depends only on one parameter, lag h.

Example

Consider the random walk X: $X_t = c + X_{t-1} + \epsilon_t$, with c is constant and white noise ϵ_t. We see that if $c \neq 0$, then $Z_t := X_t - X_{t-1} = c + \epsilon_t$ have a non-zero mean. We call it a random walk with drift. Note that since ϵ_t is independent then we call X_t a random walk with independent increments.

For more convenience, assume that c and X_0 are set to zero. We have

$$X_t = \epsilon_t + \epsilon_{t-1} + ... + \epsilon_1$$

Hence

$$\mu_t = \mathbb{E}(X_t) = 0$$

and

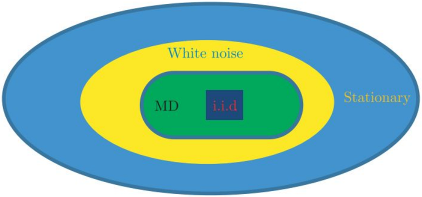
$$Var(X_t) = t\sigma$$

$$\gamma(t,s) = Cov(X_t, X_s) = (t-s)\sigma^2$$

If $s < t$ then

$$\rho(t,s) = \sqrt{1 - \frac{s}{t}}$$

which against ρ depending on t as well as on s, thus the random walk is not covariance stationary. The following figure shows the relationship among different processes: Stationary processes are the largest set, followed by white noise, martingale difference (MD), and i.i.d. processes.



Estimating Parameters of a Stationary Process

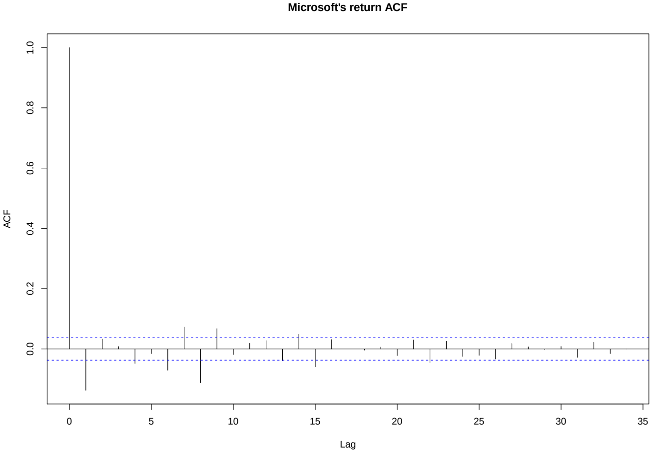
Let $X_1, X_2, ..., X_n$ be observations from weakly stationary process. To estimate the autocovariance function, we use the sample autocovariance function defined by

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X})(X_t - \bar{X})$$

To estimate function ρ, we use the sample autocorrelation function (sample ACF) defined as

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

- To visualize the dependencies of x_t for different lags h, we use the Correlogram.
- A correlogram is a plot of h (x-axis) versus its corresponding value of $\hat{\rho}(h)$ (y-axis).
- The correlogram may exhibit patterns and different degrees of dependency in a time series.
- A “band”of size $2/\sqrt{n}$ is added to the correlogram because asymptotically $\hat{\rho}(h) \sim N(0, 1/n)$ if the data is close to a white noise process.
- This band is used to detect significant autocorrelations, i.e. autocorelations that are different from zero.


Microsoft's return ACF
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
35
Lag

Testing stationarity

Augmented Dickey-Fuller Test (ADF) (also called Unit Root Test)

The test uses the following null and alternative hypotheses:

- H_0 : The time series is non-stationary, i.e., it has some time-dependent structure and does not have constant variance over time.
- H_1 : The time series is stationary.
  In R, use: *adf.test*

KPSS test

The ideas of KPSS test comes from the regression model with time trend

$$X_t = c + \mu t + k \sum_{i=1}^{t} \xi_i + \eta_t$$

with stationary η_t and i.i.d ξ with mean 0 and variance 1. Note that the third term is a random walk. So we set the hypothesis and alternative as follows

- $H_0 : k = 0$, i.e., the test is that the data is stationary
- $H_1 : k \neq 0$

in R, use *kpss.test*. Test results for Microsoft data

```
        Augmented Dickey-Fuller Test

data:  y1
Dickey-Fuller = -14.566, Lag order = 14, p-value = 0.01
alternative hypothesis: stationary


> kpss.test(y1)

        KPSS Test for Level Stationarity

data:  y1
KPSS Level = 0.38818, Truncation lag parameter = 9,
p-value = 0.08225
```

Ljung–Box Test for autocorrelations

Sample ACF with test bounds.

- These bounds are used to test the null hypothesis that an autocorrelation coefficient is 0. The null hypothesis is rejected if the sample autocorrelation is outside the bounds.
- The usual level of the test is 0.05

Example (The First-order Autoregression Model (AR(1))) The time series $X = (X_t)$ is called AR(1) if the value of X at time t is a linear function of the value of X at time $t-1$ as follows

$$X_t = \delta + \phi_1 X_{t-1} + w_t = .. = \delta + \sum_{h=0}^{\infty} \phi_1^h w_{t-h}$$

where

(i) the error $w_t \sim N(0, \sigma_w^2)$ and are i.i.d

(ii) w_t is independent of X_t

(iii) $\mid \phi_1 \mid < 1$ this condition grantees that X_t is weakly stationary

We have

- The (theoretical) mean of X_t

$$\mathbb{E}(X_t) = \mu = \frac{\delta}{1 - \phi_1}$$

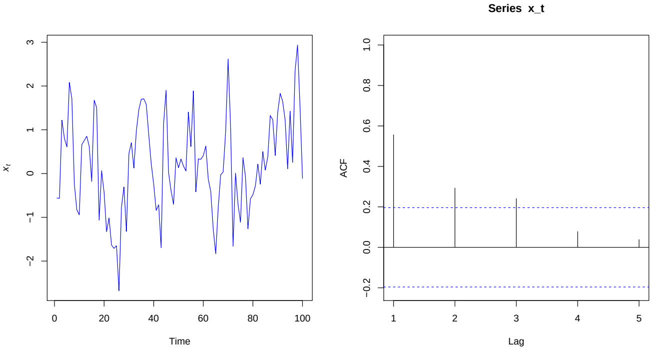
- The variance of X_t is

$$Var(X_t) = \frac{\sigma_w^2}{1 - \phi_1^2}$$

- The covariance: $Cov(X_t, X_{t+h}) = \gamma(h) = \phi_1^h \frac{\sigma_w^2}{1-\phi_1^2}$
- The correlation between observations h time periods apart is

$$\rho(h) = \phi_1^h$$

Note that the magnitude of its ACF decays geometrically to zero, either slowly as when $\phi_1 = 0.95$, moderately slowly as when $\phi_1 = 0.75$, or rapidly as when $\phi_1 = 0.25$. We now simulate $AR(1)$ and plot the ACF with $\phi_1 = 0.64$ and $\sigma_w^2 = 1$



Ljung–Box Test. The null hypothesis of the Ljung–Box test is

$$H_0 : \rho(1) = \rho(2) = ...\rho(m) = 0$$

for some m. If the Ljung–Box test rejects, then we conclude that one or more of $\rho(1), \rho(2), ..., \rho(m)$ is nonzero. The Ljung–Box test is sometimes called simply the Box test.

$$Q(m) = n(n+2)\sum_{i=j}^{m} \frac{\hat{\rho}^2(j)}{n-j} \sim \chi^2(m)$$

Example: Consider AR(1) with $\phi_1 = 0.64$ and $\sigma_w^2 = 1$, we have the results of Box test in R

Box.test(x_t, lag = 10, type = "Ljung-Box")

X-squared = 50.935, df = 10, p-value = 1.796e-07

Example (Nonstationary AR(1) Processes)

If $|\phi_1| > \geq 1$ then AR(1) process is nonstationary, and the mean, variance, covariances and and correlations are not constant.

- **Random Walk ($\phi = 1$)**

$$X_t = X_{t-1} + w_t = X_0 + w_1 + w_2 + ... + w_t$$

and the process is not stationary. Hence, for all t

$$\mathbb{E}(X_t \mid X_0) = X_0$$

which is constant but depends entirely on the arbitrary starting point. Moreover,

$$Var(X_t) = t\sigma_w^2$$

which is not stationary but rather increases linearly with time and makes the random walk “wander”, i.e.,X_t takes increasingly longer excursions away from its conditional mean of X_0, and therefore is not mean-reverting.

AR, MA and ARMA models

- The autoregressive process of order p or $AR(p)$ is defined by the equation

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \omega_t$$

  where $\omega_t \sim N(0, \sigma^2)$

- The AR model establishes that a realization at time t is a linear combination of the p previous realization plus some noise term. If $p = 0$ then $X_t = \omega_t$ there is no autoregression term.
- The lag operator (Back-shift operator) is denoted by B and used to express lagged values of the process so, $B^j X_t = X_{t-j}$
- Define

$$\Phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j$$

Then we have AR(p) can be written by

$$\Phi(B)X_t = \omega_t, \quad t = 1, 2, ..., n$$

- $\Phi(B)$ is known as the characteristic polynomial of the process and its roots determine when the process is stationary or not.

The moving average process of order q, denoted MA(q), defined as

$$X_t = \omega_t + \sum_{j=1}^{q} \theta_j \omega_{t-j}$$

- Under this model, the observed process depends on previous ω_t
- MA(q) can define correlated noise structure in our data and goes beyond the traditional assumption where errors are iid.
- We have $X_t = \Theta(B)(\omega_t)$, where $\Theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j$

The general autoregressive moving average process of orders p and q or ARMA(p, q) combines both AR(p) and MA(q) models into a unique representation.

Consider the AR(p). The process X_t is stationary if and only if all root of the characteristic equation

$$\Phi(\alpha) = 1 - \sum_{j=1}^{p} \phi_j \alpha^j = 0$$

are greater than one, i.e., $\mid \alpha_j \mid > 1$ for all $j = 1, 2, .., p$

Partial Autocorrelation Function (PACF)

A partial correlation is a conditional correlation. It is the correlation between two variables under the assumption that we know and take into account the values of some other set of variables.

- The partial autocorrelation function (PACF) of a process X_t is defined as

$$p_k = Corr(X_t, X_{t+k} \mid X_{t+1}, X_{t+2}..., X_{t+k-1}), k = 0, 1, 2, ..$$

- The PACF can also be derived through an autoregressive model of order k

$$X_{t+k} = \phi_{k1}X_{t+k-1} + \phi_{k2}X_{t+k-2} + ... + \phi_{kk}X_t + \omega_{t+k}$$

  ω_{t+k} is normal error term uncorrelated with $X_{t+k-j}, j \geq 1$
- The coefficients $\phi_{k1}, \phi_{k2}, ..., \phi_{kk}$ define the PACF
- In R we use: *acf(data, type='partial')*

- Suppose that X_t is zero mean stationary process.
- Multiplying X_{t+k-j} on both sides of the above regression equation and taking the expectation, we get

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + ... + \phi_{kk}\gamma_{j-k}$$

- If we divide by γ_0 we get,

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + ... + \phi_{kk}\rho_{j-k}$$

- Now for $j = 1, 2, ..., k$, we have the following system of equations with variables:

$$\begin{aligned}
\rho_1 &= \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + ... + \phi_{kk}\rho_{k-1} \\
\rho_2 &= \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + ... + \phi_{kk}\rho_{k-2} \\
&\vdots \\
\rho_k &= \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + ... + \phi_{kk}\rho_0
\end{aligned}$$

- We can write the system of equations as

$$P_k \phi_k = \rho$$

where

$$P_k = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-1} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix}$$

$\phi_k = (\phi_{k1}, \phi_{k2,...,\phi_{kk}})^T$ and $\rho = (\rho_1, \rho_2, ..., \rho_k)$. We get

$$\phi_{kk} = \frac{| P_k^* |}{| P_k |}$$

where P_k^* is is matrix obtained from P_k by replaced $k-$th with ρ.

- For PACF and ACF order 1, there is no difference, we have

$$\phi_{11} = \rho_1$$

- For order 2 we have

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

Example: The AR(1), process $X_t = \alpha X_{t-1} + \omega_t$ has ACF $\rho_\tau = \alpha^\tau$. For the PACF, we have $\phi_{11} = \rho_1 = \alpha$ and

$$\phi_{22} = \frac{\alpha^2 - \alpha^2}{1 - \alpha^2} = 0$$

and $\phi_{kk} = 0$ for all $k > 1$.
In general, for AR(p), we have $\rho_{kk} = 0$ for all $k > p$

Example: Consider MA(1); process $X_t = \beta\omega_{t-1} + \omega_t$, with $Var(\omega_t = \sigma^2$. It holds that $\gamma_0 = \sigma^2(1+\beta^2), \rho_1 = \frac{\beta}{1+\beta^2}$, and $\rho_k = 0$ for all $k > 1$. We have

$$\phi_{11} = \rho_1 = \frac{\beta}{1+\beta^2}$$

and

$$\phi_{22} = -\frac{\rho_1^2}{1-\rho_1^2}$$

For MA(1) process, it strictly holds that $\phi_{22} < 0$. If we continue to calculate with $k > 2$ we could determine that PACF will not reach zero.

2.2 Volatility models

- Volatility plays important role in modelling financial system, it is directly used to measure risk.
- Stocks, exchange rates, interest rates and other financial time series have some stylized facts that are difference from other time series. Stylized facts, generally speaking, are the result of many independent empirical studies statistical properties of financial markets that have been proven to be common across financial markets.
- These stylized facts are arised from Effective Market Hypothesis (EMH).
- A good candidate for modelling of financial time series should represent the properties of stochastic process. e.g., AR or ARMA can fulfill this task.
- Some volatility models, e.g., ARCH or GARCH can replicate these stylized facts appropriately.

- The efficient markets hypothesis (EMH) in finance assumes that asset prices are fair, information is accessible for everybody and is assimilated rapidly to adjust prices, and people (including traders) are rational.
- The price change $P_t - P_{t-1}$ is only due to the arrival of "news" between $t$and $t+1$. Hence individuals have no opportunities for making an investment with return greater than a fair payment for undertaking riskiness of the asset. i.e., the price is right, and there exist no arbitrage opportunities. This is called strong form the EMH
- A semi-strong form states that security prices reflect efficiently all public information, leaving rooms for the value of private information. The weak form merely assumes security prices reflect all past publicly available information.

- **Stylized fact 1.** *Time series of share prices P_t and other basic financial instruments are not stationary time series and possess a local trend at the least.*
- **Stylized fact 2.** *Returns r_t have a leptokurtic distribution. The empirically estimated kurtosis is mostly greater than 3.*
- **Stylized fact 3.** *The return process is white noise since the sampleautocorrelation $\hat{\rho}_{k,n}, k \neq 0$ is not significantly different from 0. Furthermore the white noise is not independent since the sample autocorrelations of squared and absolute returns are clearly greater than 0.*
- **Stylized fact 4.** *Volatility tends to form clusters: After a large (small) price change (positive or negative) a large (small) price change tends to occur. This effect is called volatility clustering.*

- Under the EHM, an asset return process may be expressed as

$$r_t = \mu_t + \epsilon_t$$

where $\mu_t = \mathbb{E}(r_t \mid \mathcal{F}_{t-1})$ is the conditional mean or the rational expectation of log-return, and $\epsilon_t \sim (0, \sigma_t^2)$, more precisely, $Var(\epsilon_t \mid \mathcal{F}_{t-1}) = Var(r_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$ is the conditional variance. Note that μ_t and σ_t^2 are known at time $t-1$.

- Note that $\mathcal{F}_{t-1} = \sigma(r_0, r_1, .., r_{t-1})$. So μ_t usually can be followed ARMA model. The term ϵ_t is the innovation or random component of the log-return and in practice, the conditional variance of this innovation, σ_t^2, is time varying and stochastic. So EWMA, ARCH/GARCH models may be used to model this dynamic behavior of conditional variances.

2.2. Exponentially weighted moving average (EWMA)

Denote y_t the return of stock at time t. Then

- Volatility a weighted sum of past returns, with weights ω_i, is defined by
$$\hat{\sigma}_t^2 = \omega_1 y_{t-1}^2 + \omega_2 y_{t-2}^2 + ... + \omega_L y_{t-L}^2$$
where L is the length of the estimation window, i.e., the number of observations used in the calculation. This is called MA model
- An extension of MA model is Exponentially weighted moving average. Let the weights be exponentially declining, and denote them by λ^i
$$\hat{\sigma}_t^2 = \lambda y_{t-1}^2 + \lambda^2 y_{t-2}^2 + ... + \lambda^L y_{t-L}^2$$
where $0 < \lambda < 1$. If L is large enough, the term α^n are negligible for all $n > L$. So we set $L = \infty$

Note that the sum of weights is

$$\frac{\lambda}{1-\lambda} = \sum_{i=1}^{\infty} \lambda^i$$

So the exponentially weighted moving average is defined by

$$\hat{\sigma}_t^2 = \frac{1-\lambda}{\lambda} \sum_{i=1}^{\infty} \lambda_i y_{t-i}^2$$

and, hence, we get the EWMA equation (why???)

$$\hat{\sigma}_t^2 = \lambda \hat{\sigma}_{t-1}^2 + (1-\lambda) y_{t-1}^2 \tag{1}$$

Note that JP Morgan set for daily data with $\lambda = 0.94$

Example

Suppose that $\lambda = 0.9$, the volatility estimated for a market variable for day $n-1$ is 1% per day, and during day $n-1$ the market variable increased by 2%. This means that $\sigma_{n-1}^2 = 0.01^2 = 0.0001$ and $y_{n-1}^2 = 0.02^2 = 0.0004$. From equation (1) we get

$$\sigma_n^2 = 0.9 \times 0.0001 + 0.1 \times 0.0004 = 0.00013$$

The estimate of the volatility for day n is $\sigma_n = \sqrt{0.00013} = 1.4\%$ per day. Note that the expected value of y_{n-1}^2 is $\sigma_{n-1}^2 = 0.0001$. Hence, realized value of $y_{n-1}^2 = 0.0002$ is greater that expected value, and as a result our volatility estimate increase. If the realized value of y_{n-1}^2 has been less than its expected valued, our estimate of the volatility would have decreased.

2.3 The ARCH and GARCH models

ARCH model.

- The ARCH model was proposed by Robert Engle in 1982 called autoregressive conditionally heteroscadastic
- most volatility models derive from this
- Returns are assumed to have conditional distribution (here assumed to be normal)

$$y_t \sim N(0, \sigma_t^2)$$

or we can write

$$y_t = \sigma_t \epsilon_t$$

where $\epsilon_t \sim N(0, 1)$ is called residual.

- ARCH(L_1) is defined by

$$Var(y_t \mid y_{t-1}, y_{t-2}, ..., y_{t-L_1}) = \sigma_t^2 = \omega + \sum_{i=1}^{L_1} \alpha_i y_{t-i}^2$$

where L_1 is called the lag of the model. It is seen that in the ARCH model, the volatility is weighted average of past returns. The most common form is ARCH (1)

$$Var(y_t \mid y_{t-1}) = \sigma_t^2 = \omega + \alpha y_{t-1}^2$$

where ω and α are parameters that can be estimated by maximum likelihood

If we assume that the series has mean = 0 (this can always be done by centering), then the ARCH model could be written as

$$
\begin{aligned}
& y_t = \sigma_t \epsilon_t \\
& with \quad \sigma_t = \sqrt{\omega + \alpha y_{t-1}^2} \\
& and \quad \epsilon_t \sim N(0,1), i.i.d
\end{aligned}
$$

We require that $\omega, \alpha > 0$ so that $\omega + \alpha y_{t-1}^2 > 0$ for all t. We also require that $\alpha < 1$ in order to the process to be stationary with a finite variance. Now we have

$$y_t^2 = \epsilon_t^2(\omega + \alpha y_{t-1}^2)$$

which is similar to an AR(1) for variable y_t^2 and with multiplicative noise with a mean of 1 rather than additive noise with a mean of 0.

Generalized ARCH (GARCH) model.

- It turns out that ARCH model is not a very good model and almost nobody uses it. Because, it needs to use information from many days before t to calculate volatility on day t. That is, it needs a lot of lags
- The GARCH (L_1, L_2) model is defines as

$$\sigma_t^2 = \omega + \sum_{i=1}^{L_1} \alpha_i y_{t-i}^2 + \sum_{i=1}^{L_2} \beta_i \sigma_{t-i}^2$$

  and, hence, GARCH (1,1)

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

- GARCH(1,1) is the most common specification

GARCH (1,1) unconditional volatility

- The unconditional volatility (so-called the long-run variance rate) is the unconditional expectation of volatility on given time

$$\sigma^2 = \mathbb{E}(\sigma_t^2)$$

so we have

$$\sigma^2 = \mathbb{E}(\omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2) = \omega + \alpha \sigma^2 + \beta \sigma^2$$

Hence,

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

- So to ensure positive volatility forecasts we need the condition

$$\omega, \alpha, \beta \geq 0$$

Because if any parameter is negative σ_{t+1} may be negative

- For stationary we need condition

$$\alpha + \beta < 1$$

Setting $\gamma := 1 - \alpha - \beta$ and $V := \sigma^2$ (called long-run variance rate). We have

$$\sigma_t^2 = \gamma V + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$$

Meaning of Parameters in GARCH model

- The parameter α is news, it shows that how the volatility reacts to new information
- The parameter β is memory, it shows that how much volatility remembers from the past
- The sum $\alpha + \beta$ determines how quickly the predictability (memory) of the process dies out:
    - if $\alpha + \beta \approx 0$ predictability will die out very quickly,
    - if $\alpha + \beta \approx 1$ predictability will die out very slowly

Example

Suppose that a GARCH(1,1) model is estimated from *daily data* is

$$\sigma_n = 0.000002 + 0.13y_{n-1}^2 + 0.86\sigma_{n-1}^2$$

This corresponds to $\omega = 0.000002, \alpha = 0.13, \beta = 0.86$. We have

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta} = 0.0002$$

or $\sigma = \sqrt{0.0002} = 0.014 = 1.4\%$ per day.
Suppose that the estimate of the volatility on day $n-1$ is 1.6% per day so that $\sigma_{n-1}^2 = 0.016^2 = 0.000256$, and on that day $n-1$ the market variable decreased by 1% so that $y_{n-1}^2 = 0.01^2 = 0.0001$. Then

$$\sigma_n^2 = 0.000002 + 0.13 \times 0.0001 + 0.86 \times 0.000256 = 0.00023516$$

the new estimate of the volatility is: $\sqrt{0.00023516} = 0.0153$ or 1.53% per day.

2.4. Maximum likelihood

Maximum likelihood is the most important and widespread method of estimation. What is maximum likelihood?

- Ask the question which parameters most likely generated the data we have
- Suppose we have a sample of

$$\{-0.2, 3, 4, -1, 0.5\}$$

- in the following three possibilities, which is most likely for parameters?

case	μ	σ
1	1	5
2	-2	2
3	1	2

Let $Y = (y_1, y_2, ..., y_n)$ be a vector of data and let $\theta = (\theta_1, \theta_2, ..., \theta_p)$ be a vector of parameters. Let $f(Y \mid \theta)$ be the density of Y which depends on the parameters. The function

$$L(\theta) := f(Y \mid \theta)$$

is viewed as the function of θ with Y fixed at the observed data is called the likelihood function.

- *The maximum likelihood estimator (MLE)* is the value of θ that maximizes the likelihood function. We denote the MLE by $\hat{\theta}_{ML}$.
- it is mathematically easier to maximize $\log L(\theta)$, which is called the log-likelihood. If the data are independent, then the likelihood is the product of the marginal densities

Application to ARCH(1)

Consider ARCH(1) model:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega + \alpha y_{t-1}^2 \\ \epsilon_t &\sim N(0,1) \end{aligned}$$

For $t = 2$ we have the density??

$$f(y_2 \mid y_1) = \frac{1}{\sqrt{2\pi(\omega + \alpha y_1^2)}} \exp\Big(-\frac{1}{2} \frac{y_2^2}{\omega + \alpha y_1^2} \Big)$$

Hence, the joint density

$$\prod_{t=2}^{T} f(y_t \mid y_{t-1}) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi(\omega + \alpha y_{t-1}^2)}} \exp\Big(-\frac{1}{2} \frac{y_t^2}{\omega + \alpha y_{t-1}^2} \Big)$$

and, the log likelihood

$$\log(L(\omega,\alpha)) = -\frac{T-1}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\Big(\log(\omega+\alpha y_{t-1}^2) + \frac{y_t^2}{\omega+\alpha y_{t-1}^2}\Big)$$

Application to GARCH(1,1)

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

the density

$$f(y_2 \mid y_1) = \frac{1}{\sqrt{2\pi(\omega+\alpha y_1^2+\beta\hat{\sigma}_1^2)}}\exp\Big(-\frac{1}{2}\frac{y_2^2}{\omega+\alpha y_1^2+\beta\hat{\sigma}_1^2}\Big)$$

and the log likelihood

$$\log(L(\omega,\alpha)) = -\frac{T-1}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\left(\log(\omega+\alpha y_{t-1}^2+\beta\hat{\sigma}_{t-1}^2) + \frac{y_t^2}{\omega+\alpha y_{t-1}^2+\beta\hat{\sigma}_{t-1}^2}\right)$$

The importance of σ_1

- σ_1 can make a large difference
- Especially when the sample size is small
- Typically set $\sigma_1 = \hat{\sigma}$

Volatility targeting

- Since we have the long-run variance rate

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

- we can set

$$\omega = \hat{\sigma}^2 (1 - \alpha - \beta)$$

  where $\hat{\sigma}^2$ is is the sample variance
- Hence we save one parameter in the estimation

2.5. Forecasting future volatility

The variance rate estimated at the end of day $n-1$ for n day when apply GARCH(1,1) model is

$$\sigma_n^2 = \omega + \alpha y_{n-1}^2 + \beta\sigma_{n-1}^2 = \sigma^2(1-\alpha-\beta) + \alpha y_{n-1}^2 + \beta\sigma_{n-1}^2$$

or

$$\sigma_n^2 - \sigma^2 = \alpha(y_{n-1}^2 - \sigma^2) + \beta(\sigma_{n-1}^2 - \sigma^2)$$

On day $n+t$ in the future we have

$$\sigma_{n+t}^2 - \sigma^2 = \alpha(y_{n+t-1}^2 - \sigma^2) + \beta(\sigma_{n+t-1}^2 - \sigma^2)$$

Hence,

$$\mathbb{E}[\sigma_{n+t}^2 - \sigma^2] = (\alpha+\beta)\mathbb{E}[\sigma_{n+t-1}^2 - \sigma^2]$$

By induction we obtain

$$\mathbb{E}(\sigma_{n+t}^2) = \sigma^2 + (\alpha + \beta)^t(\sigma_n^2 - \sigma^2) \tag{2}$$

Example

For the S&P data consider earlier, $\alpha + \beta = 0.9935$, the long-run variance rate $\sigma^2 = 0.0002075$ (or $\sigma = 1.44\%$ per day). Suppose that our estimate of the current variance rate per day is 0.0003 (This corresponds to a volatility of 1.732% per day). In $t = 10$ days, calculate the expected variance rate??
We have $\sigma_n^2 = 0.0003$
Hence

$$\mathbb{E}(\sigma_{n+10}^2) = 0.0002075 + 0.9935^{10} \times (0.0003 - 0.0002075) = 0.0002942$$

or the expected volatility per day is $\sqrt{0.0002942} = 1.72\%$, still above the long-term volatility of 1.44% per day.

Volatility term structures

Suppose it is day n. We define

$$V(t) = \mathbb{E}(\sigma_{n+1}^2)$$

and

$$a := \log\left(\frac{1}{\alpha+\beta}\right)$$

From (2) we have

$$V(t) = \sigma^2 + e^{-at}(V(0) - \sigma^2)$$

Then we have the average variance rate per day between today and time T

$$\frac{1}{T}\int_0^T V(t)dt = \frac{1}{T}\int_0^T \Big(\sigma^2 + e^{-at}(V(0)-\sigma^2)\Big)dt$$
$$= \sigma^2 + \frac{1-e^{-aT}}{aT}[V(0)-\sigma^2]$$

Now we define $\sigma(T)$ the volatility per annum that should be used to price a T-day option under GARCH(1,1) model. Then we have

$$\sigma^2(T) = 252\Big(\sigma^2 + \frac{1-e^{-aT}}{aT}[V(0)-\sigma^2]\Big) \tag{3}$$

This relationship between the volatility of options and their maturities is referred to as the volatility term structure.

Example

For S&P data, using GARCH(1,1) model we obtain the coefficients $\omega = 0.0000013465, \alpha = 0.083394$ and $\beta = b = 0.910116$. So from (3), assume that $V(0) = 0.0003$ we have

$$\sigma^2 = \frac{0.0000013465}{1 - 0.083394 - 0.910116} = 0.0002073$$

and $a = \log(1/0.99351) = 0.00651$. Hence,

$$\sigma^2(T) = 252\left(0.0002073 + \frac{1 - e^{-0.00651 \times T}}{0.00651 \times T}[0.0003 - 0.0002073]\right)$$

For the option life (days) $T = 10, 30, 50, 100, 500$, we obtain the option volatility (% per annum)

Option life (days)	10	30	50	100	500
option volatility	27.36	27.10	26.87	26.35	24.32

Calculate the skewness of the distribution

$$f(x) = \frac{3}{8}x^2$$

for $0 < x < 2$ and $f(x) = 0$ otherwise