12. Conditional heteroscedastic models (ARCH)

MA6622, Ernesto Mordecki, CityU, HK, 2006.

References for this Lecture:

Robert F. Engle. *Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation*, Econometrica 50:987-1008, 1982.

Analysis of Financial Time Series (Chapter 3). Ruey S. Tsay. Wiley (2002) [Available Online]

Main Purposes of Lectures 12 and 13:

- Discuss wheter the studied models satisfy the observed empirical facts of financial time series.
- Introduce the Auto Regressive Conditional Heteroscedastic Model (ARCH)
- Review Maximum Likelihood Estimation method and apply it to the ARCH model to the
- Present further developements of CH models: Generalized ARCH (GARCH) and Exponential GARCH (EGARCH).
- Discuss results of statistical fitting of EGARCH model for the Hang Seng Index

Plan of Lecture 12

(12a) Preliminaries

(12b) Auto Regressive Conditionally Heteroscedastic (ARCH) Model

(12c) Properties of the ARCH model

(12d) Maximum Likelihood Parameter Estimation

12a. Preliminaries

Heteroscedasticity

Let $\mathbf{X} = (X_1, \dots, X_d)$ be a gaussian vector with mean vector μ and variance covariance matrix Σ. More in detail

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \mathbf{cov}_{12} & \dots & \mathbf{cov}_{1d} \\ \mathbf{cov}_{21} & \sigma_2^2 & \dots & \mathbf{cov}_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{cov}_{d1} & \mathbf{cov}_{d2} & \dots & \sigma_d^2 \end{bmatrix}$$

We say that the vector is homoscedastic (or homoskedastic) when

$$\sigma_1 = \sigma_2 = \cdots = \sigma_d.$$

We say that the vector is heteroscedastic (or heteroskedastic) when this assumption does not hold.

Volatility

Given a financial time series $\{X(t)\}$ (in general, the log-returns of a financial instrument) the volatility of the financial instrument at time t is the standard deviation of the random variable $X(t)$ conditional on the registered values

$$\mathcal{F}(t-1) = (X(t-1), X(t-2), \dots).$$

More precisely, denoting the volatility by $\sigma(t)$,

$$\sigma^2(t) = \mathbf{var}(X(t) \mid \mathcal{F}(t))$$

The variance of a random variable X conditional on the values $\mathcal{F}$ can be computed as

$$\begin{aligned}\mathbf{var}(X \mid \mathcal{F}) &= \mathbf{E}\left[\left\{X - \mathbf{E}(X \mid \mathcal{F})\right\}^2 \mid \mathcal{F}\right] \\ &= \mathbf{E}\left[X^2 \mid \mathcal{F}\right] - \left[\mathbf{E}(X \mid \mathcal{F})\right]^2\end{aligned}$$

Remember from Lecture 6, one of the 6 stylized facts:

(4) Volatility appears to vary over time.

Defining the volatility as the conditional standard deviation of the returns given the past information, it is observed that if recent returns have been large, it is expected to have large returns.

Example Assume that our returns follow the random walk hypothesis, with gaussian returns. In other terms,

$$X(t) = \mu + \sigma\varepsilon(t), \quad \text{where } \{\varepsilon(t)\} \text{ GWN with unit variance.}$$

This gives a price process of the form

$$S(t) = S(0) \exp\left[n\mu + \sigma(\varepsilon(1) + \cdots + \varepsilon(t))\right].$$

i.e. a discretization of the Black-Scholes model.

The expected return at time t, conditional on the past information $\mathcal{F}(t-1)$ is

$$\mathbf{E}(X(t) \mid \mathcal{F}(t)) = \mathbf{E}(\mu + \sigma\varepsilon(t) \mid \mathcal{F}(t-1)) = \mu,$$

as $\varepsilon(t)$ is independent of $\mathcal{F}(t-1)$ and centered.

The conditional variance is

$$\begin{aligned}
&\mathbf{var}(X(t)\ |\mathcal{F}(t-1)) \\
&\qquad = \mathbf{E}\left[X(t)^2 \mid \mathcal{F}(t-1)\right] - \left[\mathbf{E}[X(t) \mid \mathcal{F}(t)]\right]^2 \\
&\qquad = \mathbf{E}\left[(\mu+\varepsilon(t))^2 \mid \mathcal{F}(t-1)\right] - \mu^2 = \sigma^2.
\end{aligned}$$

So, Black Scholes model does we verify **(4)**.

Example AR(1) time series. Assume now that our returns follow the model

$$X(t) = \mu + \phi X(t-1) + \varepsilon(t),$$

where

- $|\phi| < 1$,
- $\{\varepsilon(t)\}$ is strict white noise with variance σ_ε^2.

Let us compute the volatility begining by the conditional expectation:

$$\begin{aligned}\mathbf{E}(X(t)\,|\mathcal{F}(t-1)) &= \mathbf{E}[\mu + \phi X(t-1) + \varepsilon(t) \mid \mathcal{F}(t-1)] \\ &= \mu + \mathbf{E}[\phi X(t-1) \mid \mathcal{F}(t-1)] + \mathbf{E}[\varepsilon(t) \mid \mathcal{F}(t-1)] \\ &= \mu + \phi X(t-1).\end{aligned}$$

So

$$X(t) - \mathbf{E}(X(t) \mid \mathcal{F}(t-1)) = \varepsilon(t),$$

that is independent of $\mathcal{F}(t-1)$, so the volatility satisfies

$$\begin{aligned}\sigma(t)^2 &= \mathbf{var}\left[(X(t) - \mathbf{E}(X(t) \mid \mathcal{F}(t-1)))^2 \mid \mathcal{F}(t-1)\right] \\ &= \mathbf{var}\left[\varepsilon(t) \mid \mathcal{F}(t-1)\right] = \sigma_\varepsilon^2.\end{aligned}$$

So the AR(1) model does not satisfy **(4)**.

The same can be verified for any ARMA(p,q) model.

12b. Auto Regressive Conditionally Heteroscedastic (ARCH) Model

In order fulfill the observed empirical characteristics of financial time series, in 1982 Robert Engle[1] introduced the ARCH model[2]

The ARCH model assumes that $\{X(t)\}$ is a stationary process satisfying:

$$X(t) = \sigma(t)\varepsilon(t), \qquad \sigma(t)^2 = \omega + \alpha X(t-1)^2,$$

where $\{\varepsilon(t)\}$ is a strict white noise with unit variance, and the parameters satisfy

$$\omega > 0, \qquad 0 \leq \alpha < 1.$$

Note that for $\alpha = 0$ we obtain a strict white noise model.

[1] 2003 Nobel Laureate in Economics.

[2] Robert F. Engle. "Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation", Econometrica 50:987-1008, 1982. .

12c. Properties of the ARCH model

Let us examine the statistical properties of this model.

Conditional Mean:

$$\begin{aligned}\mathbf{E}(X(t) \mid \mathcal{F}(t-1)) &= \mathbf{E}\left(\sigma(t)\varepsilon(t) \mid \mathcal{F}(t-1)\right) \\ &= \sigma(t)\,\mathbf{E}\left(\varepsilon(t) \mid \mathcal{F}(t-1)\right) = \sigma(t)\,\mathbf{E}\,\varepsilon(t) = 0\end{aligned}$$

(Unconditional) Mean

$$\mathbf{E}\,X(t) = \mathbf{E}\left(\,\mathbf{E}(X(t) \mid \mathcal{F}(t-1)\right) = 0.$$

This means that the sequence $\{X(t)\}$ forms a martingale difference. From this it follows that

$$\mathbf{E}\,X(t)X(t-1) = 0,$$

i.e. the values are uncorrelated (but not independent!)

Conditional Variance and Volatility

$$\begin{aligned}\mathbf{var}(X(t) \mid \mathcal{F}(t-1)) &= \mathbf{E}(X(t)^2 \mid \mathcal{F}(t-1)) \\ &= \mathbf{E}\left(\sigma(t)^2\varepsilon(t)^2 \mid \mathcal{F}(t-1)\right) \\ &= \sigma(t)^2\,\mathbf{E}[\varepsilon(t)^2 \mid \mathcal{F}(t-1)] \\ &= \omega + \alpha X(t-1)^2.\end{aligned}$$

We obtain a time varying volatility

$$\sigma(t)^2 = \sqrt{\mathbf{var}(X(t) \mid \mathcal{F}(t-1))} = \sqrt{\omega + \alpha X(t-1)^2}$$

(Unconditional) Variance

$$\begin{aligned}\mathbf{var}\,X(t) &= \mathbf{E}[\mathbf{var}(X(t) \mid \mathcal{F}(t-1))] \\ &= \mathbf{E}[\omega + \alpha X(t-1)^2] \\ &= \omega + \alpha\,\mathbf{var}\,X(t-1).\end{aligned}$$

As the process is stationary,

$$\mathbf{var}\, X(t) = \mathbf{var}\, X(t-1) = \frac{\omega}{1-\alpha}.$$

Skewness and Kurtosis

For $m = 3, 4$ we obtain

$$\mathbf{E}(X(t)^m \mid \mathcal{F}(t-1)) = \sigma(t)^m \, \mathbf{E}\, \varepsilon^m(t).$$

In particular, if the white noise has normal distributions, after some computations, we obtain

- $\gamma_{X(t)} = 0$ (we have no skewness).
- $\kappa_{X(t)} = \frac{6\alpha^2}{1-3\alpha^2}$, if $\alpha < \sqrt{1/3}$
- $\kappa_{X(t)} = \infty$, if $\alpha \geq \sqrt{1/3}$
- We always have positive kurtosis, (i.e. $X(t)$ is leptokurtic)

If the white noise has a different distribution, for instance a t-student distribution, $X(t)$ inherits non-vanishing skewness. The kurtosis remains positive.

Correlation of Squares

It can be computed that

$$\mathbf{var}[X(t)^2] = \frac{2}{1-3\alpha^2}\Big(\frac{\omega}{1-\alpha}\Big)^2.$$

Furthermore,

$$\mathbf{E}\, X(t)^2 X(t-1)^2 = \frac{1+3\alpha}{1-3\alpha^2}\frac{\omega^2}{1-\alpha}.$$

And these computations allow to compute the correlation between the squared values, that is

$$\rho(X(t)^2, X(t-1)^2) = \alpha.$$

12d. Maximum Likelihood Parameter Estimation

Assume that we have historical data of certain financial instrument

$$X(0), X(1), \dots X(n),$$

and we want to fit certain model depending on a (possibly vectorial) parameter θ.

Maximum Likelihood Estimation (ML) is one method to perform the estimation of the parameter θ. The ML estimator $\bar{\theta}$ is the one that maximizes the density function of our model when we plug in our empirical data in the places of the variables of the density function of the model.

From the general theory of statistics, it is known that (under mild assumptions) the ML estimator $\bar{\theta}$ has two important properties:

- Estimators are consistent:

$$\bar{\theta} \to \theta, \quad n \to \infty.$$

  This means that, if the sample is enough large (and the model is true) our estimations are near to the true values of the parameters.

- Estimators are asymptotically normal:

$$\sqrt{n}(\bar{\theta} - \theta) \sim \mathcal{N}(\mathbf{0}, \sigma^2)$$

  meaning that, estimating σ (this is a number if we have one parameter, and a matrix if we have more that one), we can construct confidence intervals for our estimators.

Example To see how ML works, we first examine a simpler example. Suppose that we observe a sample with 10 independent values:

$$x(1) = -0.38,\ x(2) = 0.11,\ 2.2,\ 1.2,\ -0.33,\ 1.3,\\ -0.38,\ 2.1,\ 1.5,\ x(10) = 1.8,$$

that we we want to model through a $\mathcal{N}(\mu, 1)$.

Step 1: Compute the joint density of the sample of our model:

$$\begin{aligned} f(x_1, \ldots, x_{10}) &= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x_1-\mu)^2\right] \\ &\quad \ldots \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x_{10}-\mu)^2\right] \\ &= \frac{1}{(2\pi)^5} \exp\left[-\frac{1}{2}\sum_{t=1}^{10}(x_t-\mu)^2\right] \end{aligned}$$

Step 2: Plug the observed values in the density, to obtain the likelihood function:

$$\begin{aligned}\mathcal{L}(\mu) &= f(-0.38, \ldots, 1.8) \\ &= \frac{1}{(2\pi)^5} \exp\Big[-\frac{1}{2}\sum_{t=1}^{10}(x(t)-\mu)^2\Big]\end{aligned}$$

Step 3: Take the logarithm to obtain the log-likelihood function:

$$\ell(\mu) = \log \mathcal{L}(\mu) = 5\log(2\pi) - \frac{1}{2}\sum_{t=1}^{10}(x(t)-\mu)^2.$$

Step 4: To find the maximum we differentiate with respect to μ:

$$\frac{\partial \ell(\mu)}{\partial \mu} = -\sum_{t=1}^{10}(x(t)-\mu) = n\mu - \sum_{t=1}^{10} x(t),$$

that vanishes when

$$\mu = \frac{1}{10}\sum_{t=1}^{10} x(t).$$

(that can be checked to produce a minimum).

Step 5: In this way we obtain our estimator

$$\bar{\mu} = \frac{1}{10}(-0.38 + \cdots + 1.8) = 0.9$$

Remark

The data was taken from a simulated sample of $\mathcal{N}(1,1)$.