Lecture 10. Time series with conditional heteroscedasticity

Muyi Li

Xiamen University

Outline

- Reference:
  - *Analysis of Financial Time Series (3ed)*, by Ruey S. Tsay, 2010. (Chapter 3)
  - *GARCH Models*, by C. Francq and J.M. Zakoian, 2010.
  - *Handbook of Volatility Models anf Their Applications*, 2012.
- Outline:
  - Motivations
  - ARCH model
  - GARCH model
  - Other derivatives of GARCH models

Empirical properties of asset returns

Figures of illustration

- Asset prices are generally *non-stationary*. Returns are usually stationary. Some financial time series are fractionally integrated.
- Return series usually show no or little *autocorrelation*.
- There exist *volatility clusters*.
- Normality has to be rejected in favor of some *thick-tailed* distribution.
- Volatility is *stationary*, does not diverge to infinity.
- Volatility seems to react differently to a big price increase or a big price drop, referred to as the *leverage effect*.
- Reference reading: Empirical properties of assets returns: stylized facts and statistical issues. Quantitative Finance, 2001, 223-236.

R package: "quantmod"

library(quantmod)

Figure: HSI daily price and volume

Motivation of volatility models

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

$$y_t = \mu_t + \varepsilon_t,$$

where

$$\mu_t = c_0 + c_1 t (or \quad f(t)),$$

$$\varepsilon_t = \theta(L)\eta_t, \quad \theta(L) = \sum_{i=0}^{\infty} \theta_i L^i, \qquad \theta_0 = 1, \sum_{i=0}^{\infty} \theta_i^2 < \infty,$$

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

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

Motivation of volatility models

- In practice, it's too restrictive to assume innovations η_t to be i.i.d. sequence.
- Suppose y_t is the log-return series of some financial asset. If the efficient market hypothesis (EMH) is true, i.e., $E(y_t|F_{t-1}) = 0$.
- This only implies that y_t is an uncorrelated sequence.
- What is $E(y_t^2|F_{t-1})$?
- Empirical properties: volatility clustering implies that the *conditional variance of y_t is not a constant but time varying.*

Motivation of volatility models

- Consider the conditional mean and variance simultaneously:

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \\
\varepsilon_t &= \sigma_t z_t, \quad z_t \sim i.i.d.(0,1)
\end{aligned}
$$

  where $\sigma_t > 0$, z_t is independent of F_{t-1}.

$$
\begin{aligned}
&\mu_t = E[y_t|\mathcal{F}_{t-1}], ---\text{Conditional Mean} \\
&\sigma_t^2 = E[(y_t - \mu_t)^2|\mathcal{F}_{t-1}] = E[\varepsilon_t^2|F_{t-1}]. ---\text{Conditional Variance}
\end{aligned}
$$

- The dynamics of the conditional mean μ_t may be an ARMA(p,q) process or could consist of seasonality features.
- How to model σ_t^2?

- The ARCH(q) model, originally introduced by Engle (1982), is defined as follows:

$$\varepsilon_t = \sigma_t z_t, \qquad z_t \sim i.i.d.(0,1),$$
$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2.$$

  where $\omega > 0$, and $\alpha_i \geq 0$, $i = 1, \cdots, q$.
- The conditional variance(volatility) σ_t^2 is a a linear function of past squared disturbances.
- If all $\alpha_i = 0$, then the conditional variance σ_t^2 reduces to a constant.
- Robert Fry Engle III is the winner of the 2003 Nobel Memorial Prize in Economic Sciences, sharing the award with Clive Granger, "for methods of analyzing economic time series with time-varying volatility (ARCH)".

ARCH Model

- The innovation ε_t is a MDS, i.e. $E(\varepsilon_t | F_{t-1}) = 0$, hence they are serially uncorrelated but not independent.
- We can rewrite this ARCH(q) model as an AR(q) model for the squared innovation ε_t^2:

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

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

$$v_t = \varepsilon_t^2 - \sigma_t^2 = \sigma_t^2(z_t^2 - 1),$$

where $E[v_t | F_{t-1}] = 0$.

- In this case, the unconditional variance of ε_t is constant

$$E[\varepsilon_t^2] = \frac{\omega}{1 - \alpha_1 - \cdots - \alpha_q}.$$

Properties of ARCH model

- ARCH models are able to generate excess kurtosis (heavy tails).
- Consider an ARCH(1) model

$$
\begin{aligned}
\varepsilon_t &= \sigma_t z_t, z_t \sim i.i.d.(0,1), \\
\sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2,
\end{aligned}
$$

- Assume $z_t \sim i.i.d.N(0,1)$. The conditional fourth moment of ε_t:

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

- The unconditional fourth moment is

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

Properties of ARCH model

- The unconditional kurtosis is

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

for $3\alpha_1^2 < 1 (\alpha_1 > 0\ )$.

Hence, even if the innovation z_t is normally distributed, the ARCH model driven by z_t can have heavy tails.

Testing for ARCH effects: Test 1

- Under the null hypothesis of no ARCH effect, the error term ε_t is assumed to have constant conditional variance.
- The alternative hypothesis is that the error term is driven by an ARCH(q) model, so that

$$\varepsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2.$$

- $$H_0 : \alpha_1 = \cdots = \alpha_q = 0, \quad H_A : \alpha_1 \geq 0 \text{ or } \cdots \text{ or } \alpha_q \geq 0.$$
- The corresponding LM test $= T \times R^2$, where T is the sample size and the R^2 is obtained from a regression (Engle, 1982)

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \cdots + \alpha_q \hat{\varepsilon}_{t-q}^2 + v_t.$$

- When H_0 is true, the LM test statistics are asymptotically distributed as a

$$S = TR^2 \sim \chi^2(q).$$

Test for ARCH effect: Test 2

- There is no ARCH effect in the series $\{\varepsilon_t\}$ is equivalent that there are no autocorrelations in the squared sequence ε_t^2.
- Joint test for the significance of autocorrelations of ε_t^2 at first M lags:

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_M = 0 \leftrightarrow H_1 : \text{at least} \exists j \ni \rho_j \neq 0$$

where

$$\rho_k = \frac{\frac{1}{T-k}\sum_{t=k+1}^{T} \varepsilon_t^2 \varepsilon_{t-k}^2}{\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t^2 - \bar{\varepsilon_t^2})^2} \quad (\bar{\varepsilon_t^2} = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2)$$

- Define the Ljung-Box typed portmanteau test statistics:

$$LB_M = T(T+2)\sum_{k=1}^{M} \frac{\rho_k^2}{T-k}$$

- Under H_0, when sample size T is large enough, $LB_M \sim \chi^2(M)$ asymptotically.

Model checking

- Assume y_t is fitted by ARMA$(1,1)$-ARCH(2), i.e.

$$
\begin{aligned}
y_t &= \phi_1 y_{y-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \\
\varepsilon_t &= \sigma_t z_t, z_t \sim i.i.d.(0,1) \\
\sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2
\end{aligned}
$$

- Whether this model is adequate to capture the dynamics in both conditional mean and variance of y_t?
- The residual of the mean model is $\hat{\varepsilon}_t = y_t - \hat{\phi}_1 y_{t-1} - \hat{\theta}_1 \hat{\varepsilon}_{t-1}$,
- The (standardized) residual of the volatility model is $\hat{z}_t = \dfrac{\hat{\varepsilon}_t}{\hat{\sigma}_t}$. where $\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha}_1 \hat{\varepsilon}_{t-1}^2 + \hat{\alpha}_2 \hat{\varepsilon}_{t-2}^2$.
- The ACF of $\hat{z}_t$ and $\hat{z}_t^2$ can be used to check the adequacy of mean and volatility equations respectively. (Recall Ljung-Box-typed portmanteau tests)

Forecasting based on ARCH(q) model

- Let t be the starting date for forecasting (forecasting origin). Then, the 1-step ahead forecast for σ^2_{t+1} is

$$\sigma^2_t(1) = \hat{\omega} + \hat{\alpha}_1 \hat{\varepsilon}^2_t + \cdots + \hat{\alpha}_q \hat{\varepsilon}^2_{t+1-q} = \hat{\omega} + \sum_{i=1}^{q} \hat{\alpha}_i \hat{\varepsilon}^2_{t+1-i}.$$

- For the 2-step ahead forecast for σ^2_{t+2}, we need a forecast of ε^2_{t+1}. So

$$\sigma^2_t(2) = \hat{\omega} + \hat{\alpha}_1 \sigma^2_t(1) + \hat{\alpha}_2 \hat{\varepsilon}^2_t + \cdots + \hat{\alpha}_q \hat{\varepsilon}^2_{t+2-q}.$$

- The κ-step ahead forecast for $\sigma^2_{t+\kappa}$ is

$$\sigma^2_t(\kappa) = \hat{\omega} + \hat{\alpha}_1 \hat{\sigma}^2_t(\kappa - 1) + \cdots + \hat{\alpha}_p \hat{\sigma}^2_t(\kappa - q) = \hat{\omega} + \sum_{i=1}^{q} \hat{\alpha}_i \hat{\sigma}^2_t(\kappa - i).$$

with $\sigma^2_t(\kappa - i) = \hat{\varepsilon}^2_{t+\kappa-i}$ if $\kappa - i \leq 0$.

GARCH model

Due to the large persistence in volatility, the ARCH model often requires a large p to fit the data.

In such cases, it is more parsimonious to use the GARCH (Generalized ARCH) model (Bollerslev ,1986). See Table 1 for an illustration with S&P 500 stock returns.

Table: Model selection with information criteria

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

GARCH (p, q) model

- The process ε_t is called a GARCH(p, q) if

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

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

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

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

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

- When $q = 0$, it reduces to the ARCH(p) model .

GARCH(p, q) to ARCH(∞)

- Rewriting the GARCH(p, q) model as an ARCH(∞):

$$\begin{aligned}\sigma_t^2 &= [1-\beta(L)]^{-1}[\omega + \alpha(L)\varepsilon_t^2] = \omega^* + \psi(L)\varepsilon_t^2 \\ &= \omega^* + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}^2,\end{aligned}$$

  where $\psi(L) = \dfrac{\alpha(L)}{1-\beta(L)}$.

- Considering a GARCH(1,1): $\sigma_t^2 = \omega + 0.1\sigma_{t-1}^2 + 0.8\varepsilon_{t-1}^2$, we have

$$\begin{aligned}(1-0.1L)\sigma_t^2 &= \omega + 0.8\varepsilon_{t-1}^2, \\ \sigma_t^2 &= \frac{\omega}{0.9} + \sum_{j=0}^{\infty}(0.9L)^j 0.8\varepsilon_{t-1}^2.\end{aligned}$$

ARMA-in-squared

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

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

where $m = \max(p, q)$, $v_t \equiv \varepsilon_t^2 - \sigma_t^2$ (m.d.s. and uncorrelated).

- Autocovariance function:

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

- Autocorrelation function:for $j \geq q + 1$

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

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

Maximum Likelihood Estimation

- *Likelihood function:* Under the normality assumption on innovations z_t, $\varepsilon_t | F_{t-1} \sim N(0, \sigma_t^2)$, hence the likelihood function of a GARCH(p, q) model is

$$
\begin{aligned}
f(\varepsilon_1, \cdots, \varepsilon_T | F_0) &= f(\varepsilon_T | F_{T-1}) f(\varepsilon_{T-1} | F_{T-2}) \cdots f(\varepsilon_2 | F_1) f(\varepsilon_1 | F_0) \\
&= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2(\theta)}{2\sigma_t^2(\theta)} \right),
\end{aligned}
$$

- $\varepsilon_t = \begin{cases} y_t - \mu_t(\theta), & \text{if mean equation is considered} \\ y_t, & \text{if mean is 0.} \end{cases}$

Maximum Likelihood Estimation

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

$$L_T(\theta|_T) = \sum_{t=1}^{T} l_t(\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2}.$$

- **CAN (Consistency and Asymptotic Normality):** As shown by Weiss (1986), provided standardized innovations have finite fourth moments, the MLE is *consistent* and has the following asymptotic distribution

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

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

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

Weaknesses of GARCH

- Black (1976): Volatility tends to rise in response to "bad news" and to fall in response to "good news" *(leverage effect)*;
- The volatility in the GARCH process is *symmetric*, determined only by the magnitude of the previous return and shock, not by its sign.
- The parameters in GARCH are restricted to be positive to ensure positivity of σ_t^2. *(nonnegativity)* When estimating, however, sometimes best fits are achieved for negative parameters.
- The normality assumption on innovations are too restrictive in practice.

The EGARCH model

- The Exponential GARCH(p, q) (EGARCH, Nelson (1991)):

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

  with $\alpha_1 = 1$ and $g(z_t) = \varphi z_t + \gamma(|z_t| - E|z_t|)$. ($E|z_t| = \sqrt{2/\pi}$ if $z_t \sim N(0,1)$).
- The parameters ω, β_i and α_i are not restricted to be non-negative.
- The components of $g(z_t)$ are φz_t and $\gamma(|z_t| - E|z_t|)$, each with (conditional) mean zero. (Assumption on regression)
- $$g(z_t) = \begin{cases} (\varphi + \gamma) z_t - \gamma\sqrt{2/\pi}, & \text{when } z_t > 0; \\ (\varphi - \gamma) z_t - \gamma\sqrt{2/\pi}, & \text{when } z_t < 0. \end{cases}$$

  Asymmetric effect: $g(z_t)$ allows for the conditional variance process $\{\sigma_t^2\}$ to respond asymmetrically to rises and falls in stock price.

Non-Gaussian Error Distributions

- Assume innovations $z_t \sim i.i.d.t(\nu)$. The pdf of z_t is

$$t(z_t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}}\left[1+\frac{x^2}{\nu}\right]^{-\frac{\nu+1}{2}}$$

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

- Since $Var(z_t) = \frac{\nu}{\nu-2}$, we should first standardize it by $\sqrt{\frac{\nu-2}{\nu}}z_t$.
- We also have $z_t = \varepsilon_t/\sigma_t$, hence $t(z_t) = t(\frac{\varepsilon_t}{\sigma_t}\sqrt{\frac{\nu}{\nu-2}})\sqrt{\frac{\nu}{\nu-2}}\frac{1}{\sigma_t}$.
- The pdf of ε_t takes the form

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

Summary

- The evolution of GARCH family contains two aspects:
  1. The specifications of volatility structures: (G)ARCH, IGARCH, EGARCH, GARCH-M, APGARCH, etc.
  2. The distribution of innovations: Normal, Student-t, Skewed Student-t, GED.
- From discrete GARCH to continuous GARCH, from univariate GARCH to multivariate GARCH...
- Alternative volatility approach: SV model, realized volatility.

Summary

Model building: Building a volatility model for an asset return series consists of four steps:

1. Specify a mean equation by testing for serial dependence in the data and, if necessary, building an econometric model (e.g., an ARMA model) for the return series to remove any linear dependence.
2. Use the residuals of the mean equation to test for ARCH effects.
3. Specify a volatility model if ARCH effects are statistically significant, and perform a joint estimation of the mean and volatility equations.
4. Check the fitted model carefully and refine it if necessary.