Time Series Models of Heteroskedasticity

Zhiyuan Pan

Southwestern University of Finance and Economics

April 2018

Outline

- Chapter 21 in Hamilton(1994) and Chapter 3 in Fan and Yao (2006).
- ARCH model: Properties.
- Estimation: MLE, QMLE
- Extensions: GARCH, EGARCH, IGARCH, GJR, etc.
- Example

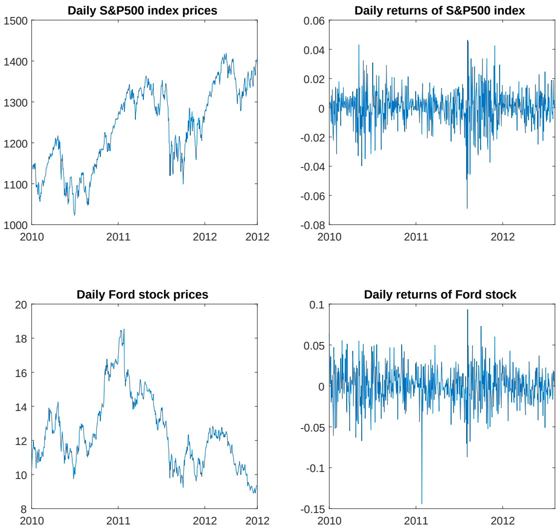
Motivation

Changes in the variance are quite important for understanding financial markets, since investors require higher expected returns as compensation for holding riskier assets.

Large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes (Mandelbrot, 1963). That is called **volatility clustering**.

- option pricing (e.g., Black-Scholes formula)
- risk measures (e.g., value-at-risk)
- risk-adjusted return (e.g., Sharpe ratio)
- securities regulations (e.g., capital requirement in Basel III)
- portfolio allocation (e.g., mean-variance model)
- hedging strategy (e.g., optimal hedging ratio)

Volatility Clustering



Modeling the asset return and conditional volatility

Conditional volatility:

$$\sigma_t = \sqrt{Var[r_t | I_{t-1}]} \tag{1}$$

where I_{t-1} is the **information set** at time $t-1$, $Var[\cdot | I_{t-1}]$ is variance conditional on I_{t-1}.

Asset return models:

$$\begin{aligned} r_t &= \mu_t + \epsilon_t \qquad (2) \\ \epsilon_t &= \sigma_t z_t, \qquad (3) \end{aligned}$$

- $\mu_t = E[r_t | I_{t-1}]$ is the conditional mean of asset return,
- σ_t is a nonnegative function of I_{t-1},
- ϵ_t is a **white noise**,
- z_t is **i.i.d.**, $E[z_t] = 0$, $Var[z_t] = 1$.

Modeling the asset return and conditional volatility

Specification of μ_t

- Constant: $\mu_t = \mu$
- ARMA process:

$$\mu_t = \phi_0 + \sum_{i=1}^{P} \phi_i r_{t-i} + \sum_{i=1}^{Q} \theta_i \epsilon_{t-i}. \tag{4}$$

In practice, we often assume $\mu_t \approx 0$ for high frequency data (i.e., daily or higher frequency returns).

In our course, we mainly focus on the volatility model. **That is how to specify the term**, σ_t

Modeling the asset return and conditional volatility

Specification of σ_t

A simple specification for the volatility function σ_t in (3) is the **autoregressive conditional heteroscedastic (ARCH) model**:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2. \tag{5}$$

where $\alpha_0 > 0, \alpha_j \geq 0 (1 \leq j \leq p)$ are constants, and p is a positive integer. We write $ARCH(p)$ for the processes defined by (5).

The ARCH model was initially introduced by Engle (1982) for modeling the predictive variance for the U.K. inflation returns. Since then it has been widely used for modeling volatilities of financial returns.

Properties of ARCH model

To gain further appreciation of the dynamics of ARCH structure, let us look at a simple $ARCH(1)$ model:

$$\begin{aligned} r_t &= \sigma_t z_t, && (6) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 && (7) \\ z_t &\sim i.i.d.(0,1), && (8) \end{aligned}$$

where $\alpha_0 > 0$ and $\alpha_1 \geq 0$.

For any $k \geq 1$,

$$Var_t[r_{t+k}] = E_t[r_{t+k}^2] = \alpha_0 + \alpha_1 E_t[r_{t+k-1}^2], \tag{9}$$

then, it follows from iterating the above expression that

$$Var_t[r_{t+k}] = \frac{\alpha_0(1-\alpha_1^k)}{1-\alpha_1} + \alpha_1^k r_t^2. \tag{10}$$

Properties of ARCH model

This indicates that a **large value** of $|r_t|$ will lead to **large volatilities** for a while in the immediate future. Hence ARCH models can produce **the volatility clustering**, a stylized feature of financial returns.

To measure the **heaviness of tail distributions**, let us compare the **kurtosis** of the return series $\{r_t\}$ with that of the innovation series $\{z_t\}$.

Under the assumption that $E[z_t^4] < \infty$ and $E[r_t^4] < \infty$, using the predictability of σ_t, it holds that,

$$E[r_t^2] = E[E_{t-1}\sigma_t^2 z_t^2] = E[\sigma_t^2 E_{t-1} z_t^2] = E[\sigma_t^2]E[z_t^2], \qquad (11)$$

where $E_{t-1}[z_t^2] = E[z_t^2]$ follows from the independence of z_t and $r_{t-1}, r_{t-2}, \cdots$.

Properties of ARCH model

Similarly,

$$E[r_t^4] = E[\sigma_t^4]E[z_t^4]. \tag{12}$$

Recall that the kurtosis is defined as

$$\kappa_r = \frac{E[r_t^4]}{(E[r_t^2])^2} \quad \text{and} \quad \kappa_z = \frac{E[z_t^4]}{(E[z_t^2])^2}, \tag{13}$$

for r_t and z_t, respectively. It follows from (12) and (13) that

$$\kappa_r = \kappa_z \frac{E[\sigma_t^4]}{(E[\sigma_t^2])^2} \geq \kappa_z, \tag{14}$$

where the last inequality follows from the fact that $E[\sigma_t^4] - (E[\sigma_t^2])^2 = Var[\sigma_t^2] \geq 0$. **This implies that the tails of the distribution of r_t is always heavier, in terms of Kurtosis, than that of z_t regardless of the distribution of z_t.**

Estimation

Maximum likelihood estimation with Gaussian distribution

$$
\begin{aligned}
r_t &= x_t'\beta + \epsilon_t && (15)\\
\epsilon_t &= \sigma_t z_t && (16)\\
\sigma_t &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2, && (17)\\
z_t &\sim i.i.d.\mathcal{N}(0,1). && (18)
\end{aligned}
$$

Then the conditional distribution of r_t is Gaussian with mean $x_t'\beta$ and variance σ_t^2:

$$
f(r_t; x_t, \mathbf{r}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(r_t - x_t'\beta)^2}{2\sigma_t^2}\right) \qquad (19)
$$

Estimation

Maximum likelihood estimation with Gaussian distribution

The sample log likelihood conditional on the first p observation is then

$$
\begin{aligned}
L &= \sum_{t=1}^{T} \log f(r_t; x_t, \mathbf{r}_{t-1}) && (20)\\
&= -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log(\sigma_t^2) - \sum_{t=1}^{T}\frac{(y_t - x_t'\beta)^2}{2\sigma_t^2}. && (21)
\end{aligned}
$$

This can then be maximized numerically using many statistical softwares, e.g., Matlab.

Estimation

Maximum likelihood estimation with t distribution

The preceding formulation of the likelihood function assumed that z_t has a Gaussian distribution. However, the unconditional distribution of many financial time series seems to have **fatter tails** than allowed by the Gaussian family.

Thus, assume that ϵ_t has a t distribution with ν degree of freedom and scale parameter M_t, then its density is given by

$$f(\epsilon_t) = \frac{\Gamma[(\nu+1)/2]}{(\pi\nu)^{1/2}\Gamma(\nu/2)} M_t^{-1/2}[1 + \frac{\epsilon_t^2}{M_t\nu}]^{-(u+1)/2}, \qquad (22)$$

where $\Gamma(\cdot)$ is the gamma function. If $\nu > 2$, then ϵ_t has mean zero and variance:

$$E[\epsilon_t^2] = M_t \frac{\nu}{\nu - 2}. \qquad (23)$$

Estimation

Maximum likelihood estimation with t distribution

Hence, a t variable with ν degree of freedom and variance σ_t^2 is obtained by taking the scale parameter M_t to be

$$M_t = \sigma_t^2 \frac{\nu - 2}{\nu}, \tag{24}$$

for which the density (22) becomes

$$f(\epsilon_t) = \frac{\Gamma[(\nu+1)/2]}{(\pi)^{1/2}\Gamma(\nu/2)}(\nu-2)^{-1/2}\sigma_t^{-1}[1+\frac{\epsilon_t^2}{\sigma_t^2(\nu-2)}]^{-(u+1)/2}, \tag{25}$$

The sample log likelihood conditional on the first p observations then becomes

$$L = \sum_{t=1}^{T} \log f(r_t | x_t, \mathbf{r}_{t-1}; \nu, \beta, \alpha). \tag{26}$$

The log likelihood is then maximized numerically with respect to ν, β, α subject to the constraint $\nu > 2$.

Estimation

Quasi-Maximum likelihood estimation

Even if the assumption that z_t is i.i.d.$\mathcal{N}(0,1)$ is invalid, we saw in (17) that the ARCH specification can still offer a reasonable model on which to base a linear forecast of the squared value of z_t. As shown in Weiss(1984, 1986), Bollerslev and Wooldridge (1992), Maximization of the Gaussian log likelihood function (21) can provide consistent estimates of the parameters even when the distribution of ϵ_t is **non-Gaussian**, provided that z_t in (16) satisfies

$$E[z_t | x_t, \mathbf{r}_{t-1}] = 0 \tag{27}$$

and

$$E[z_t^2 | x_t, \mathbf{r}_{t-1}] = 1. \tag{28}$$

However, the standard errors have to be adjusted.

Estimation

Quasi-Maximum likelihood estimation

To more details, Let $\hat{\theta}$ be the estimates in the model, and θ be the true value. Then even when z_t is actually non-Gaussian, under certain regularity conditions

$$\sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow{d} \mathcal{N}(0, D^{-1}SD^{-1}). \tag{29}$$

where

$$S = \text{plim}_{T\to\infty} Var[T^{-1/2}\sum_{t=1}^{T} s_t(\theta)] \tag{30}$$

and

$$D = \text{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} -E[\frac{\partial s_t(\theta)}{\theta'}] \tag{31}$$

where $s_t(\theta)$ is the tth score.

Usually, we can use their sample counterpart to obtain S, D, denoting $\hat{S}, \hat{D}$. For more details, see **the P661-663 in the textbook**.

Standard errors for $\hat{\theta}_T$ that are robust to misspecification of the family of densities can thus be obtained from the square root of diagonal element of

$$T^{-1}\hat{D}_T^{-1}\hat{S}_T\hat{D}_T^{-1}. \tag{32}$$

Recall that if the model is correctly specified so that the data were really generated by a Gaussian model, then $S = D$, and this simplifies to the usual asymptotic variance matrix for maximum likelihood estimation.

Testing for ARCH

Fortunately, it is simple to test whether the residuals ϵ_t from a regression model exhibit time-varying heteroskedasticity **without actually having to estimate the ARCH parameters**. Engle (1982) derived the following test based on the Lagrange multiplier principle.

- First the regression of (15) is estimated by OLS for observation $t = -p+1, -p+2, \cdots, T$ and the OLS sample residuals $\hat{\epsilon_t}$ are saved.
- Next, $\hat{\epsilon}_t^2$ is regressed on a constant and p of its own lagged values:

$$\hat{\epsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\epsilon}_{t-1}^2 + \cdots + \alpha_p \hat{\epsilon}_{t-p}^2 + e_t \tag{33}$$

- The statistics $T \times R_u^2 \xrightarrow{d} \chi^2(p)$ under the null hypothesis that ϵ_t is actually i.i.d.$\mathcal{N}(0, \sigma_t^2)$, where R_u^2 is the coefficient of determination from the regression of (33).

Extension

GARCH model

Although ARCH models catch some stylized features in financial return data, fitting real financial returns with an ARCH model often lead to an ARCH(p) **with a large value of p**. Similar to the relation between ARMA models and purely AR models, a much more parsimonious representation for volatility could be obtained by using GARCH models due to Bollerslev (1986).

A GARCH(p,q) model is defined as

$$\begin{aligned} r_t &= \sigma_t \epsilon_t && (34) \\ \sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, && (35) \end{aligned}$$

where $\omega > 0, \alpha_i \geq 0$ and $\beta_j \geq 0$, and $\epsilon_t \sim i.i.d(0,1)$.

GARCH model

Stationarity of GARCH models

Let $\eta_t \equiv \sigma_t^2(\epsilon_t^2 - 1) = r_t^2 - \sigma_t^2$. Then η_t is a sequence of martingale differences. model (34) becomes

$$r_t^2 = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \eta_t. \tag{36}$$

Substituting $\sigma_t^2 = r_t^2 - \eta_t$ into the above equation, we have

$$\begin{aligned} r_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j (r_{t-j}^2 - \eta_{t-j}) + \eta_t \\ &= \omega + \sum_{i=1}^{p\vee q} (\alpha_i + \beta_i) r_{t-i}^2 + \eta_t - \sum_{j=1}^{q} \beta_j \eta_{t-j}, \end{aligned} \tag{37}$$

where $\alpha_{p+j} = \beta_{q+j} = 0$ for $j \geq 1, p \vee q = \max(p, q)$.

GARCH model

Stationarity of GARCH models

Obviously, the model (37) is an ARMA process for r_t^2. Thus, the processes is stationarity if we impose the condition as

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \tag{38}$$

To sum up, we should impose some conditions given below when we estimate GARCH(1,1) model:

$$\omega > 0 \tag{39}$$

$$\alpha_i, \beta_i \geq 0 \tag{40}$$

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \tag{41}$$

GARCH model

Model diagnostics

The model you adopt might not be a good fit the data and model diagnostics are needed. Taking GARCH model for example, the raw materials are the residuals after GARCH fit:

$$\hat{z}_t = \frac{r_t}{\hat{\sigma}_t}. \tag{42}$$

Ideally, $\hat{z}_t$ should behave like an **i.i.d** with density the same as the density you used in estimation. Therefore, we ask the following questions:

- Is $\{\hat{z}_t\}$ an **i.i.d.** series? This is usually accomplished by the examination of the ACF plot and **performing the Ljung-Box test**.
- Is $\{\hat{z}_t^2\}$ an **i.i.d.** series? This is again be checked by using the ACF plot and **the Ljung-Box test**.

GARCH model

Model diagnostics

- Dose the distribution $\{\hat{z}_t\}$ look like the distribution used to derive the log likelihood? For example,
- if t distribution is used, we would ideally like to have the distribution of $\{\hat{z}_t\}$ look like t. This can be checked by looking at **the Q-Q plot** of the residuals against t distribution, or by applying **the Kolmogorov-Smirnov test**.
- When normal distribution is used, we may also test the Gaussianity of $\{\hat{z}_t\}$ using **the Jarque-Bera test**.

GARCH model

Forecasting GARCH model

Consider GARCH(1,1) model in (37), i.e., $p = q = 1$. Let $\sigma_T^2(k) = E_T[r_{T+k}^2]$ be the conditional volatility in k-period from time T. Then

$$r_T^2 = \omega + (\alpha_1 + \beta_1) r_{T-1}^2 + \eta_T - \beta_1 \eta_{T-1} \tag{43}$$

For $k = 1$,

$$\sigma_T^2(1) = \omega + (\alpha_1 + \beta_1) r_T^2 - \beta \hat{\eta}_T \tag{44}$$

For $k > 1$,

$$\sigma_T^2(k) = \omega + (\alpha_1 + \beta_1) \sigma_T^2(k-1) \tag{45}$$

Iterating this k time, we get

$$\sigma_T^2(k) = \frac{\omega(1 - (\alpha_1 + \beta_1)^k)}{1 - (\alpha_1 + \beta_1)} + (\alpha_1 + \beta_1)^k \sigma_T^2, \tag{46}$$

Hence,

$$\sigma_T^2(k) = \frac{\omega}{1 - (\alpha_1 + \beta_1)} \quad \text{as} \quad k \to \infty \tag{47}$$

Integrated GARCH model (IGARCH)

Suppose that $r_t = \sigma_t \epsilon_t$, where ϵ_t is i.i.d. with zero mean and unit variance and where σ_t^2 obeys the $GARCH(p,q)$ specification:

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha_1 r_{t-1}^2 + \cdots + \alpha_p r_{t-p}^2 \\ &\quad + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2 \end{aligned} \tag{48}$$

As show in (37), this ARMA process for r_t^2 would **have a unit root** if

$$\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j = 1. \tag{49}$$

Engle and Bollerslev (1986) referred to a model satisfying (49) as an **integrated** GARCH process, denoted IGARCH.

Integrated GARCH model (IGARCH)

Remark: If r_t follows an IGARCH process, then the unconditional variance of r_t is infinite, so neither r_t nor r_t^2 satisfies the definition of a covariance-stationary process. **However, it is still possible for r_t to come from a strictly stationary process in the sense that the unconditional density of r_t is the same for all t**, see Nelson (1990).

ARCH-in-Mean Specification

Finance theory suggests that an asset with a higher perceived risk would pay a higher return on average. Motivated by this, Engle, Lilien and Robins (1987) proposed the ARCH-in-mean, or ARCH-M model as

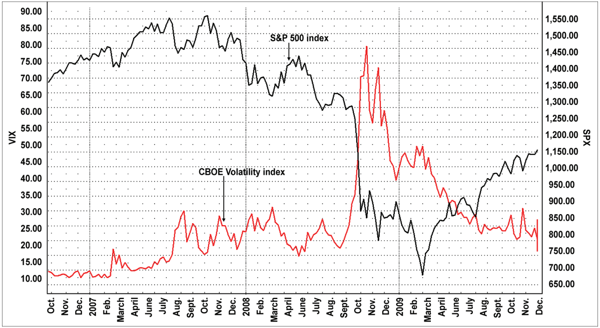
$$
\begin{aligned}
r_t &= x_t'\beta + \delta\sigma_t^2 + u_t && (50)\\
u_t &= \sigma_t\epsilon_t && (51)\\
\sigma_t^2 &= \zeta + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 && (52)
\end{aligned}
$$

for ϵ_t i.i.d. with zero mean and unit variance. The effect that higher perceived variability of u_t has on the level of r_t is captured by the parameter δ.

Exponential GARCH (EGARCH)

A number of researcher have found evidence of **asymmetry** in stock price behavior: **negative surprises seem to increase volatility more than positive surprises**.

FIGURE 1: S&P 500 & THE VIX, 2006-2009



The S&P 500 tends to move in opposite directions as volatility rises as stocks fall and vice versa. However, the VIX fell sharply in late 2008 as stocks continued to slide, a sign that no concrete relationship holds.

Source: eSignal

Exponential GARCH (EGARCH)

As before, Nelson (1991) proposed the following model for the evolution of the conditional variance of u_t:

$$\log \sigma_t^2 = \zeta + \sum_{j=1}^{\infty} \pi_j \{ |\epsilon_{t-j}| + E[|\epsilon_{t-j}|] + \gamma \epsilon_{t-j} \}. \tag{53}$$

Nelson's model is sometimes referred to as **exponential GARCH**, or EGARCH. if $\pi_j > 0$, Nelson's model implies that a deviation of $|\epsilon_{t-j}|$ from its expected value causes the variance of u_{t-j} to be larger than otherwise. The γ **parameter allow this effect to be asymmetric**.

Exponential GARCH (EGARCH)

- If $\gamma = 0$, then a positive surprise ($\epsilon_{t-j} > 0$) has the same effect on volatility as a negative surprise of the same magnitude.
- If $\gamma < 0$, a negative surprise can cause a larger volatility than a positive surprise's, sometime described as the **leverage effect**.
- The variance itself (σ_t^2) will be positive regardless of whether the π_j coefficients are positive.

GJR model

Asymmetric consequences of positive and negative innovations can also be captured with a simple modification of the linear GARCH framework. Glosten, Jagannathan, and Runkle (1989) proposed modeling $u_t = \sigma_t \epsilon_t$, where ϵ_t is i.i.d. with zero mean and unit variance and

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \gamma u_{t-1}^2 I_{t-1} + \beta_1 \sigma_t^2. \tag{54}$$

Here, $I_{t-1} = 1$ if $u_{t-1} < 0$ and $I_{t-1} = 0$ if $u_{t-1} \geq 0$. Again, if the leverage effect holds, we expect to find $\gamma > 0$. The nonnegativity condition is satisfied provided that $\beta_1 \geq 0$ and $\alpha_1 + \gamma \geq 0$.