

# Conditional Heteroscedasticity

May 30, 2010

Junhui Qian

## 1 Introduction

ARMA( $p, q$ ) models dictate that the conditional mean of a time series depends on past observations of the time series and the past innovations. Let  $\mu_t = \mathbb{E}(X_t | \mathcal{F}_{t-1})$ , we have for an ARMA( $p, q$ ) process,

$$\mu_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}.$$

If we assume  $\varepsilon_t \sim$  i.i.d. with zero mean and finite variance, then the conditional variance of  $X_t$  is a constant, regardless of the order  $p$  or  $q$ ,

$$\text{var}(X_t | X_{t-1}, X_{t-2}, \dots) = \text{var}(\varepsilon_t) < \infty.$$

In this chapter we relax this constraint and consider time-varying conditional variance.

## 2 ARCH and GARCH Models

To introduce time-varying conditional variance to the model, we write

$$X_t = \mu_t + \omega_t,$$

where  $\mu_t$  is the conditional mean as above and  $\omega_t$  is a white noise with time-varying conditional variance (conditional heteroscedasticity). Specifically, we write

$$\omega_t = \sigma_t \varepsilon_t, \tag{1}$$

where  $\varepsilon_t$  is a strong white noise with zero mean and unit variance, ie,  $\varepsilon_t \sim \text{iid}(0, 1)$ . And  $\sigma_t^2$  is the conditional variance of  $X_t$ , ie,  $\sigma_t^2 = \text{var}(X_t | X_{t-1}, X_{t-2}, \dots)$ .

For ARCH and GARCH models,  $\sigma_t^2$  evolves over time in a deterministic manner. For example, in the simplest ARCH(1) model,  $\sigma_t^2$  is specified as

$$\sigma_t^2 = c + a\omega_{t-1}^2, \quad (2)$$

where  $c > 0$  and  $a \geq 0$ . The positiveness of  $a$  implies that the probability of getting a large shock in  $\omega_t$  is high when there is a big shock in  $\omega_{t-1}$ . ARCH(1) model thus describes the volatility clustering to some extent.

More generally, we have ARCH( $p$ ) model which is specified as

**Definition: ARCH( $p$ ) Model**

$$\sigma_t^2 = c + a_1\omega_{t-1}^2 + \dots + a_p\omega_{t-p}^2, \quad (3)$$

where  $c > 0$ ,  $a_i \geq 0$  for all  $i$ .

The ARCH( $p$ ) component  $\omega_t$  has following properties.

- (a) Let  $\eta_t = \omega_t^2 - \sigma_t^2$ .  $(\eta_t)$  is a martingale difference sequence, ie,  $\mathbb{E}(\eta_t | X_{t-1}, X_{t-2}, \dots) = 0$ .
- (b)  $(\omega_t^2)$  has an AR( $p$ ) form.
- (c)  $(\sigma_t^2)$  has an AR( $p$ ) form with random coefficients,

$$\sigma_t^2 = c + \sum_{i=1}^p (a_i \varepsilon_{t-i}^2) \omega_{t-i}^2. \quad (4)$$

- (d)  $(\omega_t)$  is a white noise with a (unconditional) variance of  $\text{var}(\omega_t) = c / (1 - a_1 - \dots - a_p)$ .
- (e) The (unconditional) distribution of  $\omega_t$  is leptokurtic.

For the ARCH(1) model in (2) in particular,  $\text{var}(\omega_t) = c/(1 - a)$ . Since variance has to be positive, we must have  $0 < a < 1$ . And if we assume  $\varepsilon_t \sim \text{iid } N(0, 1)$ , we can calculate the fourth moment of  $\omega_t$ ,

$$\mathbb{E}(\omega_t^4) = \frac{3c^2(1 + a)}{(1 - a)(1 - 3a^2)}.$$

And the (unconditional) kurtosis of  $\omega_t$  is thus,

$$\text{Kurtosis}(\omega_t) = \frac{\mathbb{E}(\omega_t^4)}{[\text{var}(\omega_t)]^2} = 3 \frac{1 - a^2}{1 - 3a^2} > 3. \quad (5)$$

Since the kurtosis of  $\omega_t$  is greater than 3, the kurtosis of normal distribution, the tail of the distribution of  $\omega_t$  is heavier or longer than the normal, which is saying large shocks are more probable for  $\omega_t$  than a normal series. Of course, to ensure the kurtosis in (5) to be positive, we must have  $1 - 3a^2 > 0$ , hence  $a$  is restricted to  $[0, \sqrt{3}/3)$ .

One weakness of ARCH( $p$ ) models is that it may need many lags, ie, a big  $p$ , to fully absorb the correlation in  $\omega_t^2$ . In the same spirit of the extension from AR to ARMA models, Bollerslev (1986) proposes GARCH model, which specifies the conditional variance  $\sigma_t^2$  as follows.

**Definition: GARCH( $p, q$ ) Model**

$$\sigma_t^2 = c + a_1\omega_{t-1}^2 + \cdots + a_p\omega_{t-p}^2 + b_1\sigma_{t-1}^2 + \cdots + b_q\sigma_{t-q}^2, \quad (6)$$

where  $c > 0$ ,  $a_i \geq 0$ ,  $b_i \geq 0$  for all  $i$ , and  $\sum_{i=1}^{\max(p,q)} (a_i + b_i) < 1$ .

It is obvious that the GARCH model is a generalization of the ARCH model. If  $b_i = 0$  for all  $i$ , GARCH( $p, q$ ) reduces to ARCH( $p$ ). The GARCH component  $\omega_t$  has following properties.

- (a) Let  $\eta_t = \omega_t^2 - \sigma_t^2$ .  $(\eta_t)$  is a martingale difference sequence, ie,  $\mathbb{E}(\eta_t | X_{t-1}, X_{t-2}, \dots) = 0$ .

(b) Let  $r = \max(p, q)$ ,  $(\omega_t^2)$  has an ARMA( $r, q$ ) form,

$$\omega_t^2 = c + \sum_{i=1}^r (a_i + b_i) \omega_{t-i}^2 + \eta_t - \sum_{i=1}^q b_i \eta_{t-i}, \quad (7)$$

where  $(a_i)$  or  $(b_i)$  were padded with zero to have a length of  $r$  if necessary.

(c)  $(\sigma_t^2)$  has an AR( $r$ ) form with random coefficients,

$$\sigma_t^2 = c + \sum_{i=1}^r (a_i \varepsilon_{t-i}^2 + b_i) \omega_{t-i}^2. \quad (8)$$

(d)  $\omega_t$  is a white noise, with an (unconditional) variance of  $\text{var}(\omega_t) = c / (1 - \sum_i^p a_i - \sum_i^q b_i)$ .

(e) The (unconditional) distribution of  $\omega_t$  is leptokurtic.

GARCH(1,1) is perhaps the most popular model in practice. The conditional variance is specified as follows,

$$\sigma_t^2 = c + a\omega_{t-1}^2 + b\sigma_{t-1}^2, \quad (9)$$

where  $c > 0$ ,  $a > 0$ ,  $b > 0$ , and  $a + b < 1$ .

If  $1 - 2a^2 - (a + b)^2 > 0$ ,

$$\text{Kurtosis}(\omega_t) = \frac{\mathbb{E}(\omega_t^4)}{[\text{var}(\omega_t)]^2} = 3 \frac{1 - (a + b)^2}{1 - (a + b)^2 - 2a^2} > 3.$$

### 3 Identification, Estimation, and Forecasting

Since ARCH model is a special case of GARCH, we will focus on GARCH hereafter.

#### 3.1 Identification

For all GARCH models, the square of the GARCH component,  $\omega_t^2$ , is serially correlated. This gives us a test on whether a given process is GARCH – we may simply use the Ljung-Box test on  $\omega_t^2$ .

We may also use Engle's (1982) Lagrange test. This test is equivalent to the  $F$ -test on  $a_i = 0$  for all  $i$  in the following regression,

$$\omega_t^2 = c + a_1\omega_{t-1}^2 + \cdots + a_m^2\omega_{t-m} + \eta_t,$$

where  $m$  is a predetermined number.

To determine the order of ARCH( $p$ ), we may examine the PACF of  $\omega_t^2$ . If we believe the model is GARCH( $p = 0, q$ ), then we may use the ACF of  $\omega_t^2$  to determine  $q$ .

Finally, we may use information criteria such as AIC to determine the order of GARCH( $p, q$ ).

### 3.2 Estimation

Maximum likelihood estimation is commonly used in estimating GARCH models. Assume  $\varepsilon_t \sim N(0, 1)$ , the log likelihood function of GARCH( $p, q$ ) is

$$\begin{aligned} l(\theta|\omega_1, \dots, \omega_T) &= \log [f(\omega_T|\mathcal{F}_{T-1})f(\omega_{T-1}|\mathcal{F}_{T-2}) \cdots f(\omega_{p+1}|\mathcal{F}_p)f(\omega_1, \dots, \omega_p; \theta)] \\ &= \log f(\omega_1, \dots, \omega_p; \theta) \log \left[ \prod_{t=p+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \left( -\frac{\omega_t^2}{2\sigma_t^2} \right) \right] \\ &= \log f(\omega_1, \dots, \omega_p; \theta) \left[ -\frac{1}{2} \sum_{t=p+1}^T \left( \log(2\pi) + \log(\sigma_t^2) + \frac{\omega_t^2}{\sigma_t^2} \right) \right], \end{aligned}$$

where  $\theta$  is the set of parameters to be estimated,  $f(\omega_s|\mathcal{F}_{s-1})$  is the density of  $\omega_t$  conditional on the information contained in  $(\omega_t)$  up to time  $s - 1$ , and  $f(\omega_1, \dots, \omega_p; \theta)$  is the joint distribution of  $\omega_1, \dots, \omega_p$ .

Since the form of  $f(\omega_1, \dots, \omega_p; \theta)$  is rather complicated, the usual practice is to ignore this term and to use conditional log likelihood instead,

$$l(\theta|\omega_1, \dots, \omega_T) = -\frac{1}{2} \sum_{t=p+1}^T \left( \log(2\pi) + \log(\sigma_t^2) + \frac{\omega_t^2}{\sigma_t^2} \right). \quad (10)$$

Note that the  $\sigma_t^2$  in the above log likelihood function is not observable and has to be estimated recursively,

$$\sigma_t^2 = c + a_1\omega_{t-1}^2 + \cdots + a_p\omega_{t-p}^2 + b_1\sigma_{t-1}^2 + \cdots + b_q\sigma_{t-q}^2.$$

The initial values of  $\sigma_t^2$  are usually assigned to be the unconditional variance of  $\omega_t$ , which is  $c/(1 - \sum_i a_i - \sum_i b_i)$ .

To check whether a model is adequate, we may examine the following series,

$$\hat{\varepsilon}_t = \frac{\omega_t}{\hat{\sigma}_t}.$$

If the model is adequate and it is appropriately estimated,  $(\hat{\varepsilon}_t)$  should be iid normal. We may apply Ljung-Box test to  $(\hat{\varepsilon}_t)$  to see if the conditional mean,  $\mu_t$  in (??), is correctly specified. We may apply Ljung-Box test to  $(\hat{\varepsilon}_t^2)$  to see if the model of  $\omega_t$  is adequate. Finally, we may use Jarque-Bera Test and QQ-plot to check whether  $\varepsilon_t$  is normal.

We may, of course, use other distribution for the specification of  $\varepsilon_t$ . For example, one popular choice is the Student-t, which has heavier tails than the normal distribution. For the purpose of consistently estimating GARCH parameters such as  $(a_i)$  and  $(b_i)$ , the choice of distribution does not matter much. It can be shown that maximizing the log likelihood in (10) yields consistent estimator even when the distribution of  $\varepsilon_t$  is not normal. This is called quasi-likelihood estimation.

### 3.3 Forecasting

Forecasting volatility is perhaps the most interesting aspect of GARCH model in practice. For one-step-ahead forecast, we have

$$\sigma_{T+1}^2 = c + a_1\omega_T^2 + \cdots + a_p\omega_{T-p+1}^2 + b_1\sigma_T^2 + \cdots + b_q\sigma_{T-q+1}^2,$$

where  $(\omega_T^2, \dots, \omega_{T-p+1}^2)$  and  $(\sigma_T^2, \dots, \sigma_{T-q+1}^2)$  are known at time  $T$ . Note that the one-step-ahead forecast is deterministic.

For two-step-ahead forecasting, we have

$$\begin{aligned}\hat{\sigma}_{T+2}^2 &= c + a_1 \mathbb{E}(\omega_{T+1}^2 | \mathcal{F}_T) + a_2 \omega_T^2 + \dots + a_p \omega_{T-p+2}^2 + b_1 \sigma_{T+1}^2 + b_2 \sigma_T^2 + \dots + b_q \sigma_{T-q+2}^2 \\ &= c + a_2 \omega_T^2 + \dots + a_p \omega_{T-p+2}^2 + (a_1 + b_1) \sigma_{T+1}^2 + b_2 \sigma_T^2 + \dots + b_q \sigma_{T-q+2}^2.\end{aligned}$$

$n$ -step-ahead forecast can be constructed similarly. For GARCH(1, 1) model in (9), the  $n$ -step-ahead forecast can be written as

$$\hat{\sigma}_{T+n}^2 = \frac{c(1 - (a + b)^{n-1})}{1 - a - b} + (a + b)^{n-1} \sigma_{T+1}^2 \rightarrow \frac{c}{1 - a - b},$$

as  $n$  goes to infinity.  $\frac{c}{1-a-b}$  is exactly the unconditional variance of  $\omega_t$ .

## 4 Extensions

There are many extensions to the GARCH model. In this section we discuss four of them, Integrated GARCH (IGARCH), GARCH in mean (GARCH-M), APGARCH, and Exponential GARCH (EGARCH).

### 4.1 IGARCH

When the ARMA representation in (7) of a GARCH model has a unit root in its AR polynomial, the GARCH model is integrated in  $\omega_t^2$ . The model is then called Integrated GARCH, or IGARCH.

The key feature of IGARCH lies in the implication that any shock in volatility is persistent. This is similar with ARIMA model, in which any shock in mean is persistent. Take

the example of IGARCH(1, 1) model, which can be written as,

$$\omega_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = c + b\sigma_{t-1}^2 + (1-b)\omega_{t-1}^2.$$

The shock in volatility is given by  $\eta_t = \omega_t^2 - \sigma_t^2$ . Then

$$\omega_t^2 = c + \omega_{t-1}^2 + \eta_t - b\eta_{t-1}.$$

To forecast volatility in the IGARCH(1, 1) framework, we first have

$$\sigma_{T+1}^2 = c + b\sigma_T^2 + (1-b)\omega_T^2.$$

Then we have

$$\begin{aligned} \hat{\sigma}_{T+2}^2 &= c + \sigma_{T+1}^2 \\ \hat{\sigma}_{T+3}^2 &= c + \hat{\sigma}_{T+2}^2 = 2c + \sigma_{T+1}^2 \\ &\dots \\ \hat{\sigma}_{T+n}^2 &= (n-1)c + \sigma_{T+1}^2 \end{aligned}$$

The case when  $c = 0$  is especially interesting. In this case, the volatility forecasts  $\hat{\sigma}_{T+n}^2 = \sigma_{T+1}^2$  for all  $n$ . This approach is indeed adopted by RiskMetrics for the calculation of VaR (Value at Risk).

## 4.2 GARCH-M

To model premium for holding risky assets, we may let the conditional mean depend on the conditional variance. This is the idea of GARCH in mean, or GARCH-M. A typical



GARCH-M may be written as

$$X_t = \mu_t + \omega_t, \quad \mu_t = \alpha' z_t + \beta \sigma_t^2, \quad \omega_t = \sigma_t \varepsilon_t, \quad (11)$$

where  $z_t$  is a vector of explanatory variables and the specification for  $\sigma_t^2$  is the same as in GARCH models.

### 4.3 APGARCH

To model leverage effects, which make volatility more sensitive to negative shocks, we may consider the Asymmetric Power GARCH of Ding, Granger, and Engle (1993). A typical APGARCH( $p, q$ ) can be written as

$$\sigma_t^\delta = c + \sum_{i=1}^q a_i (|\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i})^\delta + \sum_{i=1}^p b_i \sigma_{t-i}^\delta, \quad (12)$$

where  $\delta$ ,  $c$ ,  $(\gamma_i)$ ,  $(a_i)$ , and  $(b_i)$  are model parameters.

The impact of  $\varepsilon_{t-i}$  on  $\sigma_t^\delta$  is obviously asymmetric. Consider the term  $g(\varepsilon_{t-i}, \gamma_i) = |\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i}$ . We have

$$g(\varepsilon_{t-i}, \gamma_i) = \begin{cases} (1 + \gamma_i)|\varepsilon_{t-i}|, & \text{if } \varepsilon_{t-i} \geq 0 \\ (1 - \gamma_i)|\varepsilon_{t-i}|, & \text{if } \varepsilon_{t-i} < 0. \end{cases}$$

We expect  $\gamma_i < 0$ .

The APGARCH model includes several interesting special cases,

- (a) GARCH, when  $\delta = 2$  and  $\gamma_i = 0$  for all  $i$
- (b) NGARCH of Higgins and Bera (1992), when  $\gamma_i = 0$  for all  $i$ .
- (c) GJR-GARCH of Glosten, Jagannathan, and Runkle (1993), when  $\delta = 2$  and  $-1 \leq \gamma_i \leq 0$ ,

(d) TGARCH of Zakoian (1994) when  $\delta = 1$  and  $-1 \leq \gamma_i \leq 0$ ,

(e) Log-GARCH, when  $\delta \rightarrow 0$  and  $\gamma_i = 0$  for all  $i$ .

#### 4.4 EGARCH

To model leverage effects, we may also consider Exponential GARCH, or EGARCH, proposed by Nelson (1990). An EGARCH( $p, q$ ) model can be written as

$$h_t = \log \sigma_t^2, \quad h_t = c + \sum_{i=1}^p a_i (|\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i}) + \sum_{i=1}^q b_i h_{t-i}. \quad (13)$$

As in APGARCH, we expect  $\gamma_i < 0$ . When  $\varepsilon_{t-i} > 0$  (there is good news), the impact of  $\varepsilon_{t-i}$  on  $h_t$  is  $(1 + \gamma_i)|\varepsilon_{t-i}|$ . If  $\varepsilon_{t-i} < 0$  (bad news), the impact is  $(1 - \gamma_i)|\varepsilon_{t-i}|$ .

## 5 Stochastic Volatility Models

In all ARCH and GARCH models, the evolution of the conditional variance  $\sigma_t^2$  is deterministic, conditional on the information available up to time  $t - 1$ .

SV (Stochastic Volatility) models relax this constraint and posit that the volatility itself is random. A typical SV model may be defined as

$$\omega_t = \sigma \varepsilon_t, \quad \beta(L) \log(\sigma_t^2) = c + v_t, \quad (14)$$

where  $c$  is a constant,  $\beta(z) = 1 - b_1 z - \dots - b_q z^q$ , and  $(v_t)$  is iid  $N(0, \sigma_v^2)$ .

The SV model can be estimated using quasi-likelihood methods via Kalman filtering or MCMC (Monte Carlo Markov Chain). Some applications show that SV models provide better performance in terms of model fitting. But their performance in out-of-sample volatility forecasts is less convincing.

## Appendix

### 5.1 Ljung-Box Test

The Ljung-Box test is a test of whether any of a group of autocorrelations of a time series are different from zero. It is a joint test based on a number of lags and is therefore a portmanteau test.

The Ljung-Box test statistic is defined as,

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k},$$

where  $n$  is the sample size,  $\hat{\rho}_k$  is the sample autocorrelation at lag  $k$ , and  $h$  is the number of lags being tested.  $Q$  is asymptotically distributed as the chi-square distribution with  $h$  degrees of freedom. The LjungCBox test is commonly used in model diagnostics after estimating time series models.

### 5.2 Jarque-Bera Test

The Jarque-Bera test can be used to test the null hypothesis that the data are from a normal distribution, based on the sample kurtosis and skewness. The test statistic  $JB$  is defined as

$$JB = \frac{n}{6} (S^2 + (K - 3)^2/4),$$

where  $n$  is the number of observations,  $S$  the sample skewness, and  $K$  the sample kurtosis.

$JB$  is distributed as  $\chi_2^2$ . The null hypothesis is a joint hypothesis of both the skewness and excess kurtosis being 0, since samples from a normal distribution have an expected skewness of 0 and an expected excess kurtosis of 0.