Time Series Analysis: Conditional Volatility Models (I)

Mei-Yuan Chen
Department of Finance
National Chung Hsing University

February 25, 2013
Introduction

Characteristics of the return of an asset recognized in financial literature are as follows.

1. The volatility of an asset evolves over-time in a continuous manner.

2. Periods of large movements in prices alternate with periods during which prices hardly changes. Thus large returns (of either sign) are expected to follow large returns, and small returns (of either sign) to follow small returns. This characteristic feature is commonly referred as volatility clustering or volatility pooling. In other words, the current level of volatility tends to be positively correlated with its level during the immediately preceding periods.

3. The volatility of an asset does not diverge to infinite.

4. Asymmetric movement of the volatility exists, i.e., a large (small) change in prices is more likely followed be a large (small) price
图: Time Series Plot of Taiwan Stock Returns: Jan. 06, 1971 to April 20, 2010
Time series

1. White-noise process: constant mean, constant variance, and no autocorrelation (auto-covariance at nonzero lags vanish);

2. i.i.d. process: an example of white-noise process but

3. financial time series could be white noise but not i.i.d. (other type of dependence);

4. heteroskedastic white noise whose auto-covariance at nonzero lags vanish but the variance are time-dependent;

5. martingale difference series: $X_t$ is a martingale sequence if $E(X_t|\mathcal{F}_{t-1}) = 0$ for a filtration $\mathcal{F}_t$ so that 
   $\cdots \subset \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$. 

M.-Y. Chen  GARCH
Martingale Difference v.s. White-noise Series

1. Any martingale difference series with finite second moment is (possible heterosekedastic) white-noise series. $E(X_t|\mathcal{F}_{t-1}) = 0$ is equivalent to $X_t$ being orthogonal to all random variables $Y \in \mathcal{F}_{t-1}$, and this includes the variables $X_s \subset \mathcal{F}_s \subset \mathcal{F}_{t-1}$, for every $s < t$, so that $E(X_tX_s) = 0$ for every $s < t$.

2. Not every white-noise is a martingale difference series. Because $E(X_t|Y) = 0$ implies that $X_t$ is orthogonal to all measurable functions of $Y$, not just orthogonal to linear functions.
Models for Volatility

1. Historical Volatility: the sample variance (or standard deviation) calculated with historical returns.

2. Implied Volatility: based on an asset pricing model, the implied volatility of the spot asset is calculated with observed derivative asset prices. This implied volatility is the market’s forecast of the volatility of underlying asset returns over the lifetime of the derivative asset.

3. Exponential Weighted Moving Average Models:

4. Autoregressive Volatility Models: $\sigma_t^2 = \log(r_{t,\text{high}}/r_{t,\text{low}})$, where $r_{t,\text{high}}$ and $r_{t,\text{low}}$ are the highest and lowest returns at time period $t$. And then the autoregressive volatility model is
Let $r_t = \log(p_t) - \log(p_{t-1})$ be the log return of an asset and $p_t$ be the price level of the asset at $t$. Suppose the conditional mean and conditional variance of $r_t$ given $\mathcal{F}_{t-1}$ are

\[
E(r_t|\mathcal{F}_{t-1}) = \mu_t \\
\operatorname{var}(r_t|\mathcal{F}_{t-1}) = E[(r_t - \mu_t)^2|\mathcal{F}_{t-1}] = h_t,
\]

(1)

where $\mathcal{F}_{t-1}$ denotes the information set available at time $t - 1$. In general, the time series \{r_t\} can be represented as the sum of a predictable and an unpredictable part,

\[
(r_t|\mathcal{F}_{t-1}) = E(r_t|\mathcal{F}_{t-1}) + \epsilon_t.
\]

(2)
To allow $h_t$ for time-varying, $h_t \equiv h_t(F_{t-1})$ is specified and then $\epsilon_t$ is conditional heteroscedastic that can be expressed as

$$\epsilon_t = z_t \sqrt{h_t},$$

where $z_t$ is a white noise with mean zero and variance 1. The unconditional variance of $\epsilon_t$ is

$$\sigma^2 \equiv E(\epsilon_t^2) = E[E(\epsilon_t | F_{t-1})] = E[h_t],$$

which is usually assumed to be constant, i.e., $E(h_t)$ is constant. In literature, there are lot of volatility models have been suggested and they can be classified into linear and nonlinear representations.
Descriptive Statistics of Heteroskedasticity

Time-variation in volatility (heteroskedasticity) is a common feature of macroeconomic and financial data. The perhaps most straightforward way to gauge heteroskedasticity is to estimate a time-series of variances on “rolling samples.” For a zero-mean variable, $r_t - \mu_t$, this could mean

$$h_t = \left[ (r_{t-1} - \mu_{t-1})^2 + (r_{t-2} - \mu_{t-2})^2 + \cdots + (r_{t-q} - \mu_{t-q})^2 \right] / q$$

Notice that $h_t$ depends on lagged information, and could therefore be thought of as the prediction (made in $t - 1$) of the volatility in $t$. 

M.-Y. Chen  
GARCH
Unfortunately, this method can produce quite abrupt changes in the estimate. An alternative is to apply an exponential moving average (EMA) estimator of volatility, which uses all data points since the beginning of the sample – but where recent observations carry larger weights. The weight for lag $s$ be $(1 - \lambda)\lambda^s$ where $0 < \lambda < 1$, so

$$h_t = (1 - \lambda)[(r_{t-1} - \mu_{t-1})^2 + \lambda(r_{t-2} - \mu_{t-2})^2 + \lambda^2(r_{t-q} - \mu_{t-q})^2 + \cdots].$$

which can be represented and calculated in a recursive fashion as

$$h_t = (1 - \lambda)(r_{t-1} - \mu_{t-1})^2 + \lambda h_{t-1},$$

with initial value which (before the sample) could be assumed to be zero or (perhaps better) the unconditional variance in a historical sample.
This method is commonly used by practitioners. For instance, the RISK Metrics uses this method with $\lambda = 0.94$ for use on daily data. Alternatively, $\lambda$ can be chosen to minimize some criterion function like $\sum_{t=1}^{T} [(r_t - \mu_t)^2 - h_t]^2$. 
Heteroskedastic Residuals in a Regression

Suppose we have a regression model

\[ y_t = \mathbf{x}_t' \beta_0 + e_t, \quad E(e_t) = 0, \text{cov}(\mathbf{x}_t, e_t) = 0. \]  \( (3) \)

In the standard case, classical assumptions assume that \( e_t \) is i.i.d. (independently and identically distributed), which rules out heteroskedasticity. In case the residuals actually are heteroskedastic, least squares (LS) is nevertheless a useful estimator: it is still consistent (we get the correct values as the sample becomes really large) – and it is reasonably efficient (in terms of the variance of the estimates). However, the standard expression for the standard errors (of the coefficients) is (except in a special case, see below) not correct.
There are two ways to handle this problem. First, we could use some other estimation method than LS that incorporates the structure of the heteroskedasticity. For instance, combining the regression model with an ARCH structure of the residuals – and estimate the whole thing with maximum likelihood (MLE) is one way. As a by-product we get the correct standard errors provided, of course, the assumed distribution is correct. Second, we could stick to OLS, but use another expression for the variance of the coefficients: a “heteroskedasticity consistent covariance matrix,” among which “White’s covariance matrix” is the most common.
To test for heteroskedasticity, we can use White’s test of heteroskedasticity. The null hypothesis is homoskedasticity, and the alternative hypothesis is the kind of heteroskedasticity which can be explained by the levels, squares, and cross products of the regressors – clearly a special form of heteroskedasticity. To implement White’s test, run a regression of squared fitted residuals from $y_t$ on $x_t$ on $z_t$ consisting of 1, $x_{ti}$ and $x_{ti} \times x_{tj}$ for

$$\hat{e}_t^2 = z'_t \gamma + u_t$$

(4)

and to test if all the slope coefficients (not the intercept) in $\gamma$ are zero. This can be done be using the statistic $T R^2$ which has the limiting distribution $\chi^2(\text{dim} - 1)$, where $R^2$ is from (4) and $\text{dim}$ is the dimension of $z_t$. 

M.-Y. Chen

GARCH
Linear Volatility Models: ARCH Models

The autoregressive conditional heteroskedasticity (ARCH) model, ARCH(\(q\)), suggested by Engle (1982) is defined as

\[
\begin{align*}
\begin{align*}
    r_t &= \mu_t + \epsilon_t, \\
    \epsilon_t &= z_t \sqrt{h_t}, \\
    z_t &\sim i.i.d. N(0, 1), \\
    h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i} = \alpha_0 + \alpha(L) \epsilon_{t-1}^2,
\end{align*}
\end{align*}
\]

where \(\mu_t\) may be formulated as a regression model with lagged dependent variables and exogenous variable contained in the information set \(\mathcal{F}_{t-1}\) or an ARMA model, and \(q\) is the order of the ARCH process, \(z_t\) and \(\epsilon_{t-1}\) are independent of each other, \(\epsilon_t\) is called to follow an ARCH(\(q\)) process.
The cluster effect which is the feature of financial data is captured here by $h_t$. The cluster effect indicates that large and small errors tend to cluster together (in continuous time periods). For this model to be well defined and the conditional variance $h_t$ to be positive, the following restrictions are needed, i.e., $\alpha > 0$, and $\alpha_i \geq 0$ for $i = 1, ..., q$. Besides, let $v_t \equiv \epsilon_t^2 - h_t$, then the ARCH(q) model in (6) can be rewritten as

$$\epsilon_t^2 = \alpha_0 + \alpha(L)\epsilon_{t-1}^2 + v_t.$$
Obviously, the model corresponds directly to an AR(q) model for the squared errors, $\epsilon^2_t$. The process is covariance stationary if and only if the sum of the positive autoregressive parameters is less than one, in which case the unconditional variance equals $\text{var}(\epsilon_t) = \alpha_0/(1 - \alpha_1 - \cdots - \alpha_q)$. It is obvious that the unconditional variance will be infinite if $\alpha_1 + \cdots + \alpha_q = 1$. 

M.-Y. Chen  
GARCH
For $q = 1$ and having $\mu_t = 0$, $r_t = \sqrt{h_t}z_t$ with $h_t = \alpha_0 + \alpha_1 r_{t-1}^2$ becomes the ARCH(1) model. It is clear that $h_t > 0$ with probability one provided $\alpha_1 > 0$ and $\alpha_1 \geq 0$.

As the unconditional variance $\text{var}(r_t)$ is

$$
\sigma_r^2 = \text{var}(r_t) = E(r_t^2), \quad \text{with } \mu_t = 0
$$

$$
= E\{E[r_t^2|\mathcal{F}_{t-1}]\} = E\{\alpha_0 + \alpha_1 r_{t-1}^2\}
$$

$$
= \alpha_0 + \alpha_1 E(r_{t-1}^2) = \alpha_0 + \alpha_1 \sigma_r^2
$$

$$
= \frac{\alpha_0}{1 - \alpha_1}.
$$

given $r_t$ is weakly (covariance) stationary and has finite variance.
Besides, as \( r_t = \epsilon_t = z_t \sqrt{h_t} \) given \( \mu_t = 0 \),

\[
\begin{align*}
r_t^2 &= h_t z_t^2 = h_t [1 + (z_t^2 - 1)] \\
&= h_t + h_t (z_t^2 - 1) = [\alpha_0 + \alpha_1 r_{t-1}^2] + [h_t z_t^2 - h_t] \\
&= [\alpha_0 + \alpha_1 r_{t-1}^2] + r_t^2 - h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \eta_t
\end{align*}
\]

is a form of AR(1) model in \( r_t^2 \), where \( \eta_t = r_t^2 - h_t \) is a mean zero innovation uncorrelated with its past, albeit heteroskedastic.
Likewise

\[ h_t = \alpha_0 + \alpha_1 r_{t-1}^2 \]

\[ = \alpha_0 + \alpha_1 (\eta_{t-1} + h_{t-1}) = \alpha_0 + \alpha_1 h_{t-1} + \alpha_1 \eta_{t-1}, \]

is a form of AR(1) model in \( h_t \). In addition,

\[
\text{corr}(r_t^2, r_{t-s}^2) = \frac{\text{cov}(r_t^2, r_{t-s}^2)}{\text{var}(r_t^2)} = \text{corr}(h_t, h_{t-s})
\]

\[ = \alpha_1^s > 0. \]
Exercise

Can an ARCH($q$) model be expressed as an AR($q$) model in $h_t$ and $r_t^2$ given $\mu_t = 0$?
Suppose $z_t$ is standard normal and process and $r_t$ is weakly stationary and possesses finite fourth moments. Then, as

\[
E(z_t^4) = 3[E(z_t^2)]^2 = 3
\]
\[
E(r_t^4) = E(z_t^4 h_t^2) = E(z_t^4)E(h_t^2) = 3E(h_t^2)
\]
\[
= 3E[(\alpha_0 + \alpha_1 r_{t-1}^2)^2]
\]
\[
= 3E[\alpha_0^2 + \alpha_1^2 r_{t-1}^4 + 2\alpha_0 \alpha_1 r_{t-1}^2]
\]
\[
= 3[\alpha_0^2 + \alpha_1^2 E(r_{t-1}^4) + 2\alpha_0 \alpha_1 \text{var}(r_{t-1})].
\]

\[
E(r_t^4) = \begin{cases} 
\frac{3(\alpha_0 + 2\alpha_0 \alpha_1 \text{var}(r_t))}{1 - 3\alpha_1^2} & \alpha^2 < 1/3 \\
\infty & \text{else}
\end{cases}
\]
Hence, the excess kurtosis is

\[
\kappa_4(r_t) = \frac{E(r_t^4)}{\text{var}(r_t)^2} - 3
\]

\[
= \frac{3\alpha_0^2 + 6\alpha_0 \alpha_1 \text{var}(r_t)}{1 - 3\alpha_1^2} \times \frac{(1 - \alpha_1)^2}{\alpha_0^2} - 3
\]

\[
= \frac{3\alpha_0^2 + 6\alpha_0 \alpha_1 [\alpha_0/(1 - \alpha_1)]}{1 - 3\alpha_1^2} \times \frac{(1 - \alpha_1)^2}{\alpha_0^2} - 3
\]

\[
= \frac{6\alpha_1^2}{1 - 3\alpha_1^2} \geq 0,
\]

given \(\alpha_1^2 < 1/3\). It is obvious that \(E(r_t^4) = \infty\) if \(\alpha_1 \geq \sqrt{1/3}\) and becomes finite if \(\alpha_1 < \sqrt{1/3}\). The existence of moments is important for the interpretation of the sample correlogram of \(r_t\) and \(r_t^2\), and inference.
However, suppose that $z_t$ is not Gaussian and has the fourth moment $E(z_t^4) < \infty$. Then,

$$E(r_t^4) < \infty \iff \alpha_1^2 \frac{1}{E(z_t^4)}$$

$$\kappa_4(r_t) = \frac{E(z_t^4) - 3 + 2E(z_t^4)\alpha_1^2}{1 - E(z_t^4)\alpha_1^2}.$$  

By Cauchy-Schwarz inequality, we have $E(z_t^4) \geq 1$ and so $\kappa(r_t) \geq -2$. In principle there is no restriction on $\alpha_1$ so long as $E(z_t^4)$ is close to one. If $z_t$ is not i.i.d., then restriction is even weaker.
Although the ARCH(1) model implies heavy tails and volatility clustering, it does not in practice generate enough of either. ARCH(q) for $q$ big does a bit better but at a price in terms of parsimony. There are also many inequality restrictions to impose, that can be violated otherwise.
As \( \kappa_4(r_t) = E[r_t - E(r_t)]^4/\var(r_t)^2 \) and if \( r_t = \sqrt{h_t}z_t \), where \( h_t \) is \( \mathcal{F}_{t-1} \)-measurable and \( z_t \) is independent of \( \mathcal{F}_{t-1} \) with mean zero and variance 1. Then, as \( E(r_t) = 0 \),

\[
\kappa_4(r_t) = \frac{E(r_t^4)}{\var(r_t)^2} = \frac{E(r_t^4)}{[E(r_t^2)]^2} = \frac{E(h_t^2z_t^4)}{[E(r_t^2)]^2} = \frac{E(h_t^2)E(z_t^4)}{[E(r_t^2)]^2}
\]

\[
= \kappa_4(z_t) \left[ \frac{E\{E(r_t^2|\mathcal{F}_{t-1})\]^2\}}{\{E[E(r_t^2|\mathcal{F}_{t-1})]\]^2} \right]
\]

\[
= \kappa_4(z_t) \left[ \frac{\{E[E(r_t^2|\mathcal{F}_{t-1})]\]^2 + E\{E(r_t^2|\mathcal{F}_{t-1})\]^2 - \{E[E(r_t^2|\mathcal{F}_{t-1})]\]^2}{\{E[E(r_t^2|\mathcal{F}_{t-1})]\]^2} \right]
\]

\[
= \kappa_4(z_t) \left[ 1 + \frac{\var[E(r_t^2|\mathcal{F}_{t-1})]}{[E(r_t^2)]^2} \right] \geq \kappa_4(z_t) = 3.
\]
GARCH Models

In the empirical applications of ARCH(q) models a long lag length and a large number of parameters are often called for. In order to improve this shortcomings, Bollerslev (1986) proposed the generalized ARCH, or GARCH(p,q) model, which specified the conditional variance to be a function of lagged squared errors and past conditional variance. The GARCH(p, q) process is given as

\[
\begin{align*}
    r_t &= \mu_t + \epsilon_t, \\
    \epsilon_t &= z_t \sqrt{h_t}, \quad z_t \sim \text{i.i.d. } N(0, 1), \\
    h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \\
    \alpha_0 &> 0, \quad \alpha_i \geq 0, \quad i = 1, \cdots, q, \\
    \beta_i &\geq 0, \quad i = 1, \cdots, p.
\end{align*}
\] (7)
GARCH(1, 1) Models

A GARCH (1,1) is expressed as

\[ h_t = \alpha_0 + \alpha_1 \epsilon^2_{t-1} + \beta_1 h_{t-1} \] (8)

Denote \( e_t = \epsilon^2_t - h_t \) or \( h_t = \epsilon^2_t - e_t \) and then the equation (8) becomes

\[ \epsilon^2_t - e_t = \alpha_0 + \alpha_1 \epsilon^2_{t-1} + \beta_1 (\epsilon^2_{t-1} - e_{t-1}) \]

\[ \epsilon^2_t = \alpha_0 + (\alpha_1 + \beta_1) \epsilon^2_{t-1} + e_t - \beta_1 e_{t-1} \]

is an ARMA(1, 1) process for the squared errors.
Given the GARCH(1,1) model in (8), we have

\[ h_{t-1} = \alpha_0 + \alpha_1 \epsilon_{t-2}^2 + \beta_1 h_{t-2} \]

\[ h_{t-2} = \alpha_0 + \alpha_1 \epsilon_{t-3}^2 + \beta_1 h_{t-3} \]

Then, (8) becomes

\[ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 (\alpha_0 + \alpha_1 \epsilon_{t-2}^2 + \beta_1 h_{t-2}) \]

\[ h_t = \alpha_0 + \alpha_0 \beta_1 + \alpha_1 \epsilon_{t-1}^2 + \alpha_1 \beta_1 \epsilon_{t-2}^2 + \beta_1^2 h_{t-2} \]

\[ h_t = \alpha_0 + \alpha_0 \beta_1 + \alpha_1 \epsilon_{t-1}^2 + \alpha_1 \beta_1 \epsilon_{t-2}^2 + \beta_1^2 (\alpha_0 + \alpha_1 \epsilon_{t-3}^2 + \beta_1 h_{t-3}) \]

\[ h_t = \alpha_0 (1 + \beta_1 + \beta_1^2) + \alpha_1 \epsilon_{t-1}^2 + \alpha_1 \beta_1 \epsilon_{t-2}^2 + \beta_1^2 \alpha_1 \epsilon_{t-3}^2 + \beta_1^3 h_{t-3} \]

\[ h_t = \alpha_0 (1 + \beta_1 + \beta_1^2) + \alpha_1 (1 + \beta_1 L + \beta_1^2 L^2) \epsilon_{t-1}^2 + \beta_1^3 h_{t-3} \]
An infinite number of successive substitutions of this kind would yield

\[ h_t = \alpha_0 (1 + \beta_1 + \beta_1^2 + \cdots) + \alpha_1 (1 + \beta_1 L + \beta_1^2 L^2 + \cdots) \epsilon^2_{t-1} + \beta_1^\infty h_{t-\infty}. \]

As \( \beta_1^\infty \to 0 \), the GARCH(1,1) model can be written as

\[ h_t = \alpha_0 (1 + \beta_1 + \beta_1^2 + \cdots) + \alpha_1 (1 + \beta_1 L + \beta_1^2 L^2 + \cdots) \epsilon^2_{t-1} = \gamma_0 + \gamma_1 \epsilon^2_{t-1} + \gamma_2 \epsilon^2_{t-2} + \cdots + \gamma_\infty \epsilon^2_{t-\infty} \]

where \( \gamma_j = \alpha_1 \beta_1^j, j = 1, 2, \ldots \). Therefore, a GARCH(1, 1) models can be expressed as an ARCH(\( \infty \)) model.
Recall that the autocovariance function of an ARMA(1, 1) process \( r_t = \phi r_{t-1} + z_t + \pi z_{t-1} \) with \( 0 < |\phi| < 1 \) is

\[
\gamma_{r_t}(k) = \sigma^2 \phi^k \left( \frac{(1 + \pi \phi)(1 + \pi/\phi)}{1 - \phi^2} \right)
\]

\[
\gamma_{r_t}(0) = \sigma^2 \left( \frac{(1 + \pi \phi)(1 + \pi/\phi)}{1 - \phi^2} - \frac{\pi}{\phi} \right)
\]

As an GARCH(1,1) model, \( h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1} \),

\[
E(h_t) = E(r_t^2) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1},
\]

given \( E(h_t) = E(r_t^2) \) is independent of \( t \) when \( r_t \) is stationary.
As an GARCH(1,1) model can represented as an ARMA(1, 1) model in a form of $r_t^2$, the autocovariances of $r_t^2$ is

$$
\gamma_{r_t^2}(k) = \tau^2 (\alpha_1 + \beta_1)^k \left( \frac{[1 - \beta_1 (\beta_1 + \alpha_1)][1 - \beta_1/(\alpha_1 + \beta_1)]}{1 - (\alpha_1 + \beta_1)^2} \right), \quad k > 0,
$$

$$
\gamma_{r_t^2}(0) = \tau^2 \left( \frac{[1 - \beta_1 (\beta_1 + \alpha_1)][1 - \beta_1/(\alpha_1 + \beta_1)]}{1 - (\alpha_1 + \beta_1)^2} + \frac{\beta_1}{\alpha_1 + \beta_1} \right),
$$

where $\tau^2$ is the variance of $\eta_t = r_t^2 - E(r_t^2|\mathcal{F}_{t-1})$. 

M.-Y. Chen  
GARCH
For $p = 0$, the GARCH($p$,$q$) process reduces to an ARCH($q$) process, and for $p = q = 0$, $\epsilon_t$ is simply white noise. The process is covariance stationary if and only if the sum of the positive $\alpha_i$ and $\beta_i$ is less than one, in which case the unconditional variance equals

$$\text{var}(\epsilon_t) = \alpha_0/(1 - \sum_{i=1}^q \alpha_i - \sum_{i=1}^p \beta_i).$$

Obviously, the unconditional variance will be infinite if

$$\alpha_1 + \cdots + \alpha_q + \beta_1 + \cdots + \beta_p = 1.$$
ARCH-in-mean(or ARCH-M) Models

From the asset price theorem, the returns of an asset depend on the level of risk as the asset takes. Thus the conditional variance of an asset can influence its conditional mean. Therefore, Engle, Lilien, and Robins (1987) extended the basic ARCH framework to impose the conditional variance into the mean function which is called ARCH-M model. Thus, the conditional variance in ARCH-M model, \( h_t \), is regarded as explanatory variable in the conditional mean equation specified as

\[
\begin{align*}
  r_t &= \mu_t + \gamma h_t + \epsilon_t, \\
  \epsilon_t &= z_t \sqrt{h_t}, \\
  z_t &\sim i.i.d. \ N(0,1), \\
  h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2,
\end{align*}
\]
IGARCH Models

A common finding in much of the empirical studies using high-frequency financial data concerns the apparent persistence implied by the estimates of the conditional variance equation. In the linear GARCH\((p,q)\) model the persistence is manifested by the presence of an approximate unit root in the autoregressive polynomial; i.e., \(\alpha_1 + \cdots + \alpha_q + \beta_1 + \cdots + \beta_p = 1\). Engle and Bollerslev (1986) refer to this model as Integrated in variance or IGARCH which is

\[
h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j};
\]

\[
\alpha_1 + \cdots + \alpha_q + \beta_1 + \cdots + \beta_p = 1.
\]

Furthermore, the unconditional variance for IGARCH\((p,q)\) model does not exists.
The linear GARCH(p,q) model successfully captures thick tailed, and volatility clustering, but is not well suited to capture the "leverage effect". The empirical evidence for stock returns reflects that there exists a negative correlation between current returns and future volatility. It implies a stock price decrease tends to increase subsequent volatility by more than would a stock price increase of the same magnitude. This phenomenon is reflecting the asymmetry in the conditional variance equation. Hence, many other parametric formulations explained the leverage effect have been considered in the literature.
Engle and Ng (1993) compare several alternative specifications of this leverage effect, and conclude that the parameterization of Glosten, Jagannathan, and Runkle (1993) is the most promising. Their GARCH-L(1,1) specification is

\[ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \xi d_{t-1} \epsilon_{t-1}^2 + \beta_1 h_{t-1}, \]

where \( d_{t-1} \) is a dummy variable that is equal to zero if \( \epsilon_{t-1} > 0 \) and equal to unity if \( \epsilon_{t-1} \leq 0 \). This allows the impact of the squared errors on conditional volatility to be different according to the sign of the lagged error terms. The leverage effect mentioned above predicts that \( \xi > 0 \).
Beta-t-EGARCH Models

The first-order Beta-t-(E) GARCH model suggested by Harvey and Chakravarty (2008), Cambridge Working Papers in Economics 0840, Faculty of Economics, University of Cambridge, is

\[ r_t = \sqrt{h_t} \epsilon_t, \quad \epsilon_t \sim i.i.d.t(\nu) \]

\[ \log(h_t) = \delta + \phi_1 \log(h_{t-1}) + \kappa_1 u_{t-1} + \kappa_1^* \text{sign}(-r_t)(u_{t-1} + 1) \]

where $u_t$ is the score or first derivative of the log-likelihood with respect to $\log(h_t)$. The score $u_t$ is zero-mean, i.i.d. and Beta distributed. Estimation and simulation of the 1st order Beta-t-EGARCH model is provided in the R package `betategarch`.
Stochastic Volatility Models

In the GARCH model, the conditional volatility of $r_t$ is driven by the same shocks as its conditional mean. Furthermore, conditional upon the history of $r_t$ as summarized in the information set $\mathcal{F}_{t-1}$, current volatility $h_t$ is deterministic. An alternative class of volatility models assumes that $h_t$ is subject to an additional contemporaneous shock. The basic stochastic volatility (SV) model, introduced by Taylor (1989), is given by

$$
\epsilon_t = z_t \sqrt{h_t} \quad (10)
$$

$$
\ln(h_t) = \gamma_0 + \gamma_1 \ln(h_{t-1}) + \gamma_2 \eta_t, \quad (11)
$$

with $z_t \sim i.i.d. N(0, 1)$, $\eta_t \sim i.i.d. N(0, 1)$, and $z_t$ and $\eta_t$ are uncorrelated.
Nonlinear GARCH Models

For stock returns, it appears to be the case that volatile periods often are initiated by a large negative shock, which suggests that positive and negative shocks may have an asymmetric impact on the conditional volatility of subsequent observations. This was recognized by Black (1976), who suggested that a possible explanation for this finding might be the way firms are financed. When the value of (the stock of) a firm falls, the debt-to-equity ratio increases, which in turns leads to an increase in the volatility of the returns on equity. As the debt-to-equity ratio is also known as “leverage” of the firm, this phenomenon is commonly referred to as the “leverage effect”. 
Exponential GARCH Models

The earliest variant of the GARCH model which allows for asymmetric effects is the Exponential GARCH (EGARCH), introduced by Nelson (1991). The EGARCH($p, q$) model is given by

$$\ln(h_t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i [\phi z_{t-i} + \gamma(|z_{t-i}| - E(|z_{t-i}|))] + \sum_{j=1}^{p} \beta_j \ln(h_{t-j}).$$

(12)
Denote $z_{t,q} = [z_t, z_{t-1}, \ldots, z_{t-q}]'$ and then
\[ g(z_t) = \sum_{i=0}^{q-1} \alpha_i [\phi z_{t-i} + \gamma(|z_{t-i}| - E(|z_{t-i}|))]. \]
The function $g(z_t)$ is piecewise linear in $z_t$, as it can be written as
\[
g(z_t) = \sum_{i=0}^{q-1} \left( \alpha_i \phi + \gamma \right) z_{t-i} I(z_{t-i} \geq 0) \\
+ \left( \alpha_i \phi - \gamma \right) z_{t-i} I(z_{t-i} < 0) - \gamma E(z_{t-i}).
\]
GJR-GARCH Models

The model introduced by Glosten, Jagannathan, and Runkle (1993) offers an alternative method to allow for asymmetric effects of positive and negative shocks to volatility. The GJR-GARCH\((p, q)\) model can be written as

\[
h_t = \alpha_0 + \sum_{i=1}^{q} \{ \alpha_i \epsilon_{t-i}^2 [1 - I(\epsilon_{t-i} > 0)] \\
+ \gamma_i \epsilon_{t-i}^2 I(\epsilon_{t-i} > 0) \} + \sum_{j=1}^{p} \beta_j h_{t-j}. \quad (13)
\]
The Logistic Smooth Transition GARCH (LSTGARCH\((p, q)\)) model is given by

\[
h_t = \alpha_0 + \sum_{i=1}^{q} \{ \alpha_i \epsilon_{t-i}^2 [1 - F(\epsilon_{t-i})] \\
+ \gamma_i \epsilon_{t-i}^2 F(\epsilon_{t-i}) \} + \sum_{j=1}^{p} \beta_j h_{t-j},
\]

(14)

where the function \( F(\epsilon_{t-i}) \) is the logistic function

\[
F(\epsilon_{t-1}) = \frac{1}{1 + \exp(-\theta \epsilon_{t-i})}, \quad \theta > 0.
\]

(15)
As the function $F(\epsilon_{t-i})$ in (15) changes monotonically from 0 to 1 as $\epsilon_{t-i}$ increases, the impact of $\epsilon_{t-i}^2$ on $h_t$ changes also smoothly. When the parameter $\theta$ in (15) becomes large, the logistic function approaches a step function which equals 0 for negative $\epsilon_{t-i}$ and 1 for positive $\epsilon_{t-i}$. In that case, the LSTGARCH($p, q$) reduces to the GJR-GARCH($p, q$).
The smooth transition model can also be used to describe asymmetric effects of large and small shocks on conditional volatility by using the exponential function

\[ F(\epsilon_{t-i}) = 1 - \exp(-\theta \epsilon^2_{t-i}), \quad \theta > 0. \]  

(16)

The function \( F(\epsilon_{t-i}) \) in (16) changes from 1 for large negative values of \( \epsilon_{t-i} \) to 0 for \( \epsilon_{t-i} = 0 \) and increases back again to 1 for large positive values of \( \epsilon_{t-i} \). Thus, the effective parameter of \( \epsilon^2_{t-i} \) in the exponential smooth transition GARCH (ESTGARCH) model changes from \( \gamma_i \) to \( \alpha_i \) and back to \( \gamma_i \) again.
R Package: RSTAR
Volatility-Switching GARCH Models

The LSTGARCH and GJR-GARCH models assume that the asymmetric behaviour of $h_t$ depends only on the sign of the past shocks $\epsilon_{t-i}$. In applications, it is typically found that a negative shock increase the conditional variance more than a positive shock of the same size. On the other hand, the ESTGARCH model assumes that the asymmetry is caused by the size of the shock. Rabemananjara and Zakoïan (1993) point out that negative shocks increase future conditional volatility more than positive shocks only if the shock is larger in absolute value. For small shocks they observe the opposite kind of asymmetry, in that small positive shocks increase the conditional volatility more than small negative shocks.
Fornari and Mele (1996, 1997) discuss a model which allows for such complicated asymmetry behaviour. The model is obtained by allowing all parameters in the conditional variance equation to depend on the sign of the shock $\epsilon_{t-i}$. The Volatility Switching GARCH (VS-GARCH) model is given by

$$h_t = \left( \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \right)[1 - I(\epsilon_{t-i} > 0)] + \left( \gamma_0 + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \delta_j h_{t-j} \right)I(\epsilon_{t-i} > 0).$$

(17)

Clearly, it is a generalization of the GJR-GARCH models.
Asymmetric Nonlinear Smooth Transition GARCH Models

Anderson, Nam, and Vahid (1999) modify the VS-GARCH model by allowing the transition from one regime to the other to be smooth. The resulting Asymmetric Nonlinear Smooth Transition GARCH (ANST-GARCH) model is given by

\[ h_t = (\alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j})[1 - F(\epsilon_{t-i})] \]

\[ + (\gamma_0 + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \delta_j h_{t-j})F(\epsilon_{t-i}). \]  (18)
Quadratic GARCH Models

Sentana (1995) introduces the Quadratic GARCH (QGARCH) model to cope with asymmetric effects of shocks on volatility. The QGARCH\((p, q)\) model is specified as

\[
h_t = \alpha_0 + \gamma \epsilon_{t-1} + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}
\]  

(19)
A general Markov-Switching GARCH (MSW-GARCH) model is given by

\[
    h_t = (\alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}) I(s_t = 1)
\]

\[
    + (\gamma_0 + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \delta_j h_{t-j}) I(s_t = 2)
\]  

(20)

where \(s_t\) is a two-state Markov chain with certain transition probability matrix. The markov-switching ARCH models for conditional variance is suggested by Hamilton and Susmel (1994) and Cai (1994). On the other hand, the markov-switching GARCH models for volatility is suggested in Garry (1996).
ARCH Models with Conditionally Non-normal Disturbances

In applying ARCH or GARCH model to empirical study, many researchers criticize the assumption of conditionally normal distribution in the ARCH or GARCH process since the fat-tail distributed data is usually observed. Bollerslev (1987) suggested using the student-$t$ distribution with an estimated kurtosis regulated by the degrees of freedom parameter could fit the data more better. That is a $t$-GARCH model.
Thus, the specification of the $t$-GARCH model is

$$
\begin{align*}
 r_t &= \mu_t + \epsilon_t; \\
 \epsilon_t | \mathcal{F}_{t-1} &\sim f_\nu(\epsilon_t | \mathcal{F}_{t-1}) \\
 &= \Gamma\left(\frac{\nu + 1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)^{-1}((\nu - 2)h_t | t-1)^{-\frac{1}{2}} \\
 &\quad \times \left(1 + \epsilon_t^2 h_t^{-1}(\nu - 2)^{-1}\right)^{-\frac{(\nu + 1)}{2}}, \nu > 2; \\
 h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}; \\
 t &= 1, \ldots, T.
\end{align*}
$$

where $\mathcal{F}_{t-1}$ is the information set available up through time $t - 1$, $f_\nu(\epsilon_t | \mathcal{F}_{t-1})$ is the student-$t$ distribution function for $\epsilon_t$ with degrees of freedom $\nu$, $h_t$ is the conditional variance for $\epsilon_t$, and $T$ is the sample size.
Besides, $f_\nu(\epsilon_t|\mathcal{F}_{t-1})$ is symmetric around zero and the variance and the fourth moment are

$$\text{Var}(\epsilon_t|\mathcal{F}_{t-1}) = h_{t|t-1};$$
$$\mathbb{E}(\epsilon_t^4|\mathcal{F}_{t-1}) = 3(\nu - 2)(\nu - 4)^{-1}h_{t|t-1}^2, \quad \nu > 4.$$ 

Furthermore, if $\nu \to \infty$ the student-$t$ distribution approximates a normal distribution with variance $h_{t|t-1}$, but when $\nu$ is not large enough, the student-$t$ distribution has fatter tails than the corresponding normal distribution.
Here, the degree of freedom $\nu$ is treated as an unknown parameter, which need to be estimated with the other unknown parameters in the model. Let $\theta$ be all the unknown parameters in the model. Then the log-likelihood function for the whole sample is

$$L_T(\theta) = \sum_{t=1}^{T} \log f_{\nu}(\epsilon_t | \theta, \mathcal{F}_{t-1}).$$

The standard inference procedures regarding $\theta$ are available.
$t$-distributed Errors

When the innovation $z_t$ are assumed to be normally distributed, the conditional distribution of $\epsilon_t$ is normal with mean zero and variance $h_t$. The unconditional distribution of a series $\epsilon_t$ for which the conditional variance follows a GARCH model is non-normal. In particular, the kurtosis of $\epsilon_t$ is larger than the normal value of 3, the unconditional distribution has fatter tails than the normal distribution.
Long Memory Stochastic Volatility (LMSV) Models

The stochastic volatility model is specified by

\[ \epsilon_t = z_t \sqrt{h_t}, \quad h_t = \sigma \exp(\nu_t/2), \]

where \( \{\nu_t\} \) is independent of \( z_t \), \( \{z_t\} \) is independent and identically distributed with mean zero and variance one, and \( \{\nu_t\} \) is an ARMA. The long memory stochastic volatility (LMSV) model is suggested by Breidt, Crato, and Lima (1998) for \( \{\nu_t\} \) being a stationary long-memory process.
When \( \{v_t\} \) is Gaussian, \( \{\epsilon_t\} \) is both covariance and strictly stationary. Denote \( \gamma(\cdot) \) as the autocovariance function of \( \{v_t\} \). The covariance structure of \( \epsilon_t \) is obtained from properties of the lognormal distribution:

\[
\begin{align*}
\mathbb{E}(\epsilon_t) &= 0, \\
\text{var}(\epsilon_t) &= \exp(\gamma(0)/2) \cdot \sigma^2 \\
\text{cov}(\epsilon_t, \epsilon_{t+h}) &= 0 \text{ for } h \neq 0.
\end{align*}
\]

So that \( \{\epsilon_t\} \) is a **white noise sequence**. Since

\[
\frac{\mathbb{E}(\epsilon_t^4)}{[\mathbb{E}(\epsilon_t^2)]^2} - 3 = 3\{\exp[\gamma(0)] - 1\} > 3
\]

when the driving noise \( \{z_t\} \) is Gaussian, \( \{\epsilon_t\} \) displays **excess kurtosis**.
The process \( \{ \epsilon_t^2 \} \) is also both covariance and strictly stationary. Thus

\[
\mathbb{E}(\epsilon_t^2) = \exp \left[ \frac{\gamma(0)}{2} \right] \cdot \sigma^2,
\]

\[
\text{var}(\epsilon_t^2) = \sigma^4 \left\{ [1 + \text{var}(z_t^2)] \exp[2\gamma(0)] - \exp \gamma(0) \right\},
\]

\[
\text{cov}(\epsilon_t^2, \epsilon_{t+h}^2) = \sigma^4 \left\{ \exp[\gamma(0) + \gamma(h)] - \exp[\gamma(0)] \right\}, \quad \text{for } h \neq 0.
\]
The series is simple to analyze after it is transformed to the stationary process

\[ x_t = \log(\epsilon_t^2) = \log(h_t z_t^2) = \log(h_t) + \log(z_t^2) \]

\[ = \log[\sigma^2(\exp(v_t/2))^2] + \log(z_t^2) \]

\[ = \log(\sigma^2) + \log[(\exp(v_t/2))^2] + \log(z_t^2) \]

\[ = \log(\sigma^2) + 2 \log[\exp(v_t/2)] + \log(z_t^2) \]

\[ = \log(\sigma^2) + \text{IE}[\log(z_t^2)] + v_t + \{\log(z_t^2) - \text{IE}[\log(z_t^2)]\} \]

\[ = \mu + v_t + \xi_t, \]

where \( \{\xi_t\} \) is i.i.d. with mean zero and variance \( \sigma^2_\xi \). For example, if \( z_t \) is standard normal, then \( \log(z_t^2) \) is distributed as the log of a \( \chi_1^2 \) random variable, \( \text{IE}[\log(z_t^2)] = -1.27 \) and \( \sigma^2_\xi = \pi/2 \) (Wishart, 1947).
Therefore, \( \{x_t\} \) is a long-memory Gaussian signal plus an i.i.d. non-Gaussian noise, with \( \mathbb{E}(x_x) = \mu \), and

\[
\gamma_x(k) = \text{cov}(x_t, x_{t+k}) = \gamma(k) + \sigma^2_\xi \cdot I_{\{k=0\}},
\]

where

\[
I_{\{k=0\}} = 1, \quad \text{if} \quad k = 0 \\
= 0, \quad \text{otherwise}.
\]

It turns out that the autocovariance function of the process \( \{\log(\epsilon^2_t)\} \) is the same as that of a specified fractional integrated EGARCH model, see Breidt, et al (1998).
The LMSV model considered in this paper is to extend the original LMSV model of Breidt, et al (1998) by replacing $h_t = \sigma \exp(x_t/2)$ with a more general specification $\sigma_t = K(x_t)$, i.e., :

$$r_t = \sigma_t \epsilon_t, \quad \sigma_t = K(x_t),$$

where $K(\cdot)$ is a positive Borel function,

$$x_t = \sum_{i=1}^{\infty} \alpha_i \eta_{t-i}$$

is a long-memory linear process with regular varying coefficients $\alpha_i = i^{-\alpha} L(i), 1/2 < \alpha < 1$, and i.i.d. zero-mean-unit-variance innovations $\{\eta_t\}$ which is independent of $\{\epsilon_t\}$. $\mathbb{E}(x_t^2) = 1$ is also assumed. This model is called GLMSV in this paper.
Under the GLMSV model assumption, this paper aims to estimate and test the high-order cross moments (or higher-order cross covariance)

\[ \beta(p, q, k) = \text{cov}(r^p_t, r^{q}_{t+k}) \]

by using the nature sample estimate

\[ \hat{\beta}(p, q, k) = n^{-1} \sum_{t=1}^{n} (r^p_t - \bar{\mu}_p)(r^q_{t+k} - \bar{\mu}_q), \]

where \( n \) is the sample size and \( \bar{\mu}_s = n^{-1} \sum_{t=1}^{n} r^s_t, s = p, q. \)
The major implication of $\beta(p, q, k)$ is that it can be used to reflect various kind of forecast efficiency.

1. $\beta(1, q, k) = 0$ for all $q$ and $k$ speaks of the usual market efficiency;

2. Non-zero $\beta(p, q, k)$ with restricted $q$ and fixed $k$ is related to the efficiency of volatility forecast, skewness forecast and kurtosis forecast for $p = 2, 3, \text{and} 4$, respectively.

The limiting distribution of the normalized $\hat{\beta}(p, q, k)$ is derived in Theorem 1. Since the result of (1) depends on the values of $\alpha$ and $d$, consistent estimates of $\alpha$ and $d$ are suggested in empirical inference on $\beta(p, q, k)$. 

M.-Y. Chen  
GARCH
Simulated and Estimate ARMA-GARCH Time series using R

In the packages, TSA and fGarch, a time series with ARMA(p, q) conditional mean and GARCH(p, q) conditional variance can be simulated. Some example R codes, like TSA_garch_sim.R and fGarch_sim.R, are to simulated various time series process with conditional mean and variance.
R Packages for Volatility Models

1. For volatility modeling, the standard GARCH(1,1) model can be estimated with the `garch()` function in the `tseries` package.

2. `Rmetrics` contains the `fGarch` package which has additional models.

3. The `rugarch` package can be used to model a variety of univariate GARCH models with extensions such as ARFIMA, in-mean, external regressors and various other specifications; with methods for fit, forecast, simulation, inference and plotting are provided too.

4. The `egarch` package provides functions for simulating and fitting EGARCH models.

5. The `betategarch` package can estimate and simulate the Beta-t-EGARCH model by Harvey.

6. The `bayesGARCH` package can perform Bayesian estimation of a GARCH(1,1) model with Student’s t innovations.
R package: rugarch

The rugarch package provides for a comprehensive set of methods for modelling univariate GARCH processes, including fitting, filtering, forecasting, simulation as well as diagnostic tools including plots and various tests. Additional methods such as rolling estimation, bootstrap forecasting and simulated parameter density to evaluate model uncertainty provide a rich environment for the modelling of these processes. See the R code rugarch.R and rugarch1.R.
R package: egarch

1. **egarch**: This function fits an Exponential Generalized Autoregressive Conditional Heteroscedastic Model (EGARCH\((p, q)\)-Model) with normal or ged distributed innovations to a given univariate time series.

2. **egarchSim**: This function simulates a univariate EGARCH\((p, q)\) model with normal or ged distributed innovations.

See the R code eGARCH.R. The package **rugarch** also provides EGARCH specification, fitting, inferences and forecasting.