Time Series Analysis: Conditional Volatility Models (II)

Mei-Yuan Chen
Department of Finance
National Chung Hsing University

February 28, 2013
In this section, several important issues related to the statistic testing for GARCH models are discussed.

1. Homoscedastic v.s. Linear GARCH
2. Linear GARCH v.s. Asymmetric Nonlinear GARCH
3. Homoscedastic v.s. Asymmetric Nonlinear GARCH
Testing for Linear GARCH

Engle (1982) developed a test for the conditional heteroskedasticity in the context of ARCH models based on the Lagrange Multiplier (LM) principle. Given an ARCH(q) model,

\[ h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2, \]

the null of constant \( h_t \) becomes \( H_0 : \alpha_1 = \cdots = \alpha_q = 0 \). The corresponding LM test can be computed as \( T R^2 \), where \( T \) is the sample size and the \( R^2 \) is obtained from auxiliary regression:

\[ \hat{\epsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\epsilon}_{t-1}^2 + \cdots + \alpha_q \hat{\epsilon}_{t-q}^2 + u_t, \]

where \( \hat{\epsilon}_t \) is the residual obtained from the regression of conditional mean under the null hypothesis.
The LM test-statistic has an asymptotic $\chi^2(q)$ distribution. Lee (1991) shows that the LM test against this GARCH\((p, q)\) alternative is the same as the LM test against the alternative of ARCH\((q)\) errors.
Testing for Nonlinear GARCH

With respect to the specification of nonlinear GARCH models, there are two possible routes one might follow. First, one can start with specifying and estimating a linear GARCH model and subsequently test the need for asymmetric or other nonlinear components in the model. Second, one can test the null hypothesis of conditional homoscedasticity against the alternative of asymmetric ARCH.
Engle and Ng (1993) discuss tests to check whether positive and negative shocks have a different impact on the conditional variance. Let $S_{t-1}^-$ denote a dummy variable which takes the value 1 when $\hat{\epsilon}_{t-1}$ is negative and 0 otherwise, where $\hat{\epsilon}_t$ are the residuals from estimating a model for the conditional mean of $r_t$ under the null of conditional homoscedasticity. The tests examine whether the squared residual $\hat{\epsilon}_t^2$ can be predicted by $S_{t-1}^-$, $S_{t-1}^-\hat{\epsilon}_{t-1}$, and/or $S_{t-1}^+\hat{\epsilon}_{t-1}$, where $S_{t-1}^+ \equiv 1 - S_{t-1}^-$. The test-statistics are computed as the $t$-ratio of the parameter $\phi_1$ in the regression

$$\hat{\epsilon}_t^2 = \phi_0 + \phi_1 \hat{w}_{t-1} + u_t,$$

where $\hat{w}_{t-1}$ is one of $S_{t-1}^-$, $S_{t-1}^-\hat{\epsilon}_{t-1}$, and $S_{t-1}^+\hat{\epsilon}_{t-1}$. 
Where $\hat{w}_{t-1} = S_{t-1}^-$, the test is called the Sign Bias (SB) test. In case $\hat{w}_{t-1} = S_{t-1}^- \hat{e}_{t-1}$ or $\hat{w}_{t-1} = S_{t-1}^+ \hat{e}_{t-1}$, the tests are called the Negative Size Bias (NSB) and Positive Size Bias (PSB) tests, respectively. As the SB, NSB and PSB statistics are $t$-ratios, they follow a standard normal distribution asymptotically.
The tests can be constructed jointly, by estimating the regression

\[ \hat{\epsilon}_t^2 = \phi_0 + \phi_1 S_{t-1}^- + \phi_2 S_{t-1}^- \hat{\epsilon}_{t-1} + \phi_3 S_{t-1}^+ \hat{\epsilon}_{t-1} + u_t. \]

The null hypothesis \( H_0 : \phi_1 = \phi_2 = \phi_3 = 0 \) can be evaluated by computing \( TR^2 \), where \( R^2 \) is the coefficient of determinant from the regression. The resulted test-statistic has an asymptotic \( \chi^2 \) distribution with 3 degrees of freedom.
Sentana (1995) discusses a test of homoscedasticity against the alternative of quadratic ARCH (QARCH). Consider the QARCH model in (??) and setting $\beta_1 = \cdots = \beta_p = 0$ as well as adding lagged shocks $\epsilon_{t-2}, \ldots, \epsilon_{t-q}$ and their squares, that is

$$h_t = \alpha_0 + \gamma_1 \epsilon_{t-1} + \gamma_2 \epsilon_{t-2} + \cdots + \gamma_q \epsilon_{t-q} + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_q \epsilon_{t-q}^2.$$

Therefore, the null for conditional homoscedasticity is

$$H_0 : \alpha_i = \gamma_i = 0, \forall i = 1, \ldots, q.$$
A LM statistic to test the null can be computed as $TR^2$ from a regression of $\hat{\epsilon}_t^2$ on $\hat{\epsilon}_{t-1}, \ldots, \hat{\epsilon}_{t-q}$ and $\hat{\epsilon}_{t-1}^2, \ldots, \hat{\epsilon}_{t-q}^2$. Asymptotically, the LM-statistic is $\chi^2$ distributed with $2q$ degrees of freedom.
Hagerud (1997) suggests two statistics to test constant conditional variance against STARCH. The STARCH($q$)

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

where $F(\cdot)$ is either the logistic function or the exponential function. The null hypothesis of conditional homoscedasticity can again be specified as

$$H_0 : \alpha_1 = \cdots = \alpha_q = \gamma_1 = \cdots = \gamma_q = 0.$$
The testing problem is complicated in this case as the parameter $\theta$ in $F(\cdot)$ is not identified under the null. The solution is to approximate the transition function by a lower-order Taylor. In case of the Logistic STARCH (LSTARCH) model, this results in the auxiliary model

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i^* \epsilon_{t-i}^2 + \sum_{i=1}^{q} \gamma_i^* \epsilon_{t-i}^3.$$
An LM-statistic to test the equivalent null hypothesis

\[ H_0^* : \alpha_1^* = \cdots = \alpha_q^* = \gamma_1^* = \cdots = \gamma_q^* = 0 \]

can be computed as \( TR^2 \) from the regression of \( \hat{\epsilon}_t^2 \) on \( \hat{\epsilon}_{t-1}^2, \ldots, \hat{\epsilon}_{t-q}^2 \) and \( \hat{\epsilon}_{t-1}^3, \ldots, \hat{\epsilon}_{t-q}^3 \). Asymptotically, the statistic is \( \chi^2 \) distributed with \( 2q \) degrees of freedom. In case of the Exponential STARCH (ESTARCH) model, the auxiliary regression includes \( \hat{\epsilon}_{t-i}^4, i = 1, \ldots, q \) instead of \( \hat{\epsilon}_{t-i}^3, i = 1, \ldots, q \).
Testing for ARCH in the Presence of Misspecification

The small sample properties of the LM test for linear (G)ARCH have been investigated quite extensively. In particular, it has been found that rejection of the null hypothesis of homoscedasticity might be due to other sorts of model misspecification, such as neglected serial correlation (e.g., Engle, Hendry, and Trumble (1985), Bera, Higgins, and Lee (1992) and Sullivan and Giles (1995)), nonlinearity (e.g., Bera and Higgins (1997)) and omitted variables (e.g., Giles, Giles, and Wong (1993) and Lumsdaine and Ng (1999)) in the model for the conditional mean.
One suggestive solution might be estimating the conditional mean by nonparametric regression and then the fitted residuals are obtained. Using the significance test in nonparametric regression of $\hat{\epsilon}_t^2$ on $\hat{\epsilon}_{t-1}, \hat{\epsilon}_{t-1}^2, \hat{\epsilon}_{t-1}^3$ and $\hat{\epsilon}_{t-1}^4$. 
van Dijk, Franses, and Lucas (1999) show that the behaviour of the LM test for ARCH based on the regression:

\[ h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2, \]

in the presence of additive outliers (AO). If the AOs are neglected, the LM test suffers from the over-rejection of the null of homoscedasticity when it is in fact true. Besides, the test has difficulty detecting genuine GARCH effects, in the sense that the power of the test is reduced considerably. To overcome this problem, the robust LM test is suggested.
Suppose the conditional mean equation is \( r_t = \phi_1 r_{t-1} + \epsilon_t \). Lucas (1996) discusses how to use the Generalized M (GM) estimator to estimate the conditional mean. A GM estimator of \( \phi_1 \) is obtained from the moment condition:

\[
\sum_{t=1}^{T} \left( r_t - \tilde{\phi}_1 r_{t-1} \right) r_{t-1} \cdot w_{\tilde{\xi}}(\tilde{\xi}_t) = 0,
\]

where \( \tilde{\xi}_t \) denotes the standardized residual,

\[
\tilde{\xi}_t \equiv \frac{(r_t - \tilde{\phi}_1 r_{t-1})}{[\tilde{\sigma}_\epsilon w_x(r_{t-1})]},
\]

with \( \tilde{\sigma}_\epsilon \) a measure of scale of the residuals \( \tilde{\epsilon}_t \equiv r_t - \tilde{\phi}_1 r_{t-1} \) and \( w_x \) and \( w_{\tilde{\xi}} \) are weight functions bounded between 0 and 1.
As AO at $t = \tau$ shows up as aberrant value of $r_\tau$ and/or 
$(r_{\tau+1} - \phi_1 r_\tau)/\sigma_\epsilon$, whereas the latter can also be caused by an 
AO at time $\tau + 1$ of course. The functions $w_x$ and $w_\xi$ should 
be chosen such that the $\tau + 1$st observation receives a 
relatively small weight if either the regressor $r_\tau$ or the 
standardized residual $(r_{\tau+1} - \tilde{\phi}_1 r_\tau)/\tilde{\sigma}_\epsilon$ becomes large, such 
that the outlier does not influence the estimate of $\phi_1$ and $\sigma_\epsilon$. 
The weight function \( w_\xi(\tilde{\xi}_t) \) usually is specified in term of a function \( \psi(\tilde{\xi}_t) \) as \( w_\xi(\tilde{\xi}_t) = \psi(\tilde{\xi}_t)/\tilde{\xi}_t \) for \( \tilde{\xi}_t \neq 0 \) and \( w_\xi(0) = 1 \). Common choices for the \( \psi(\cdot) \) function are the Huber and Tukey bisquare functions. The Huber \( \psi \) function is given by

\[
\psi(\tilde{\xi}_t) = \begin{cases} 
-c & \text{if } \tilde{\xi}_t \leq -c \\
\tilde{\xi}_t & \text{if } -c < \tilde{\xi}_t \leq c \\
c & \text{if } \tilde{\xi}_t > c,
\end{cases}
\]

or \( \psi(\tilde{\xi}) = \text{med}(−c, c, \tilde{\xi}) \), where med denotes the median and \( c > 0 \).
Usually $c$ is taken equal to 1.345 to produce an estimator that has an efficiency 95% compared to the OLS estimator if $\epsilon_t$ is normally distributed. The weights $w_\xi(\tilde{\xi}_t)$ implied by the Huber function have the attractive property that $w_\xi(\tilde{\xi}_t) = 1$ if $-c \leq \tilde{\xi}_t < c$. Only observations for which the standardized residual is outside this region receive less weight. A disadvantage is that these weights decline to zero very slowly. Subjective judgement is thus required to decide whether a weight is small or not.
The Tukey bisquare function is given by

\[
\psi(\tilde{\xi}_t) = \begin{cases} 
\tilde{\xi}_t(1 - (\tilde{\xi}_t/c)^2)^2 & \text{if } |\tilde{\xi}_t| \leq c \\
0 & \text{if } |\tilde{\xi}_t| > c.
\end{cases}
\]

Usually $c$ is set equal to 4.685, again to achieve 95% efficiency for normally distributed $\epsilon_t$. 
A third possibility is the polynomial $\psi$ function as proposed in Lucas, van Dijk and Kloek (1996), given by

$$
\psi(\tilde{\xi}_t) = \begin{cases} 
\tilde{\xi}_t & \text{if } |\tilde{\xi}_t| \leq c_1 \\
\text{sgn}(\tilde{\xi}_t)g(|\tilde{\xi}_t|) & \text{if } c_1 < |\tilde{\xi}_t| \leq c_2 \\
0 & \text{if } |\tilde{\xi}_t| > c_2.
\end{cases}
$$

Usually, $c_1 = 2.576$ and $c_2 = 3.291$ are taken.
The weight function $w_x(\cdot)$ for the regressor is commonly specified as

$$w_x(r_{t-1}) = \psi(d(r_{t-1})^\alpha)/d(r_{t-1})^\alpha,$$

where $d(r_{t-1})$ is the Mahalanobis distance of $r_{t-1}$, that is $d(r_{t-1}) = |r_{t-1} - m_r|/\sigma_r$, with $m_r$ and $\sigma_r$ measures of location and scale of $r_{t-1}$, respectively. These measures can be estimated robustly by the median $m_r = \text{med}(r_{t-1})$ and the median absolute deviation (MAD) $\sigma_r = 1.483 \cdot \text{med}(|r_{t-1} - m_r|)$, respectively. Finally, following Simpson, Ruppert, and Carroll (1992), the constant $\alpha$ usually is set equal to 2 to obtain robustness of standard errors.
Notice that the weight $w_\xi(\cdot)$ depend on the unknown parameter $\phi_1$ and therefore are not fixed a priori but are determined endogenously. Consequently, the moment condition is nonlinear in $\phi_1$ and $\sigma_\epsilon$, and estimation of these parameters requires an iterative procedure. In fact, interpreting $w_\xi(\cdot)$ as a function of $(\phi_1, \sigma_\epsilon)$, $w_\xi(\phi_1, \sigma_\epsilon)$, and denote the estimates of $\phi_1$ and $\sigma_\epsilon$ at the $n$th iteration by $\phi_1^{(n)}$ and $\hat{\sigma}_\epsilon^{(n)}$, respectively.
It follows the moment condition that $\hat{\phi}_1^{(n+1)}$ might be computed as the weighted least squares estimate

$$
\hat{\phi}_1^{(n+1)} = \frac{\sum_{t=1}^{T} w_\xi(\hat{\phi}_1^{(n)}, \hat{\sigma}_\epsilon^{(n)}) r_{t-1} r_t}{\sum_{t=1}^{T} w_\xi(\hat{\phi}_1^{(n)}, \hat{\sigma}_\epsilon^{(n)}) r_{t-1}^2},
$$

where the estimate of $\sigma_\epsilon$ can be updated at each iteration using a robust estimator of scale, such as the median absolute deviation (MAD) given by $\sigma_\epsilon = 1.483 \cdot \text{med}(|\epsilon_t - \text{med}(\epsilon_t)|)$. 
Let $\hat{\xi}_t$ be the standardized residuals and the weight function $w_\xi(\hat{\xi}_t)$. The weighted and standardized residuals $w_\xi(\hat{\xi}_t)\hat{\xi}_t = \psi(\hat{\xi}_t)$ can be constructed. A robust equivalent to the LM test for ARCH($q$) is obtained as $TR^2$, where $R^2$ is from the regression of $\psi(\hat{\xi}_t)^2$ on a constant and $\psi(\hat{\xi}_{t-1})^2, \ldots, \psi(\hat{\xi}_{t-q})^2$. Under conventional assumptions, the outlier-robust LM test has a $\chi^2(q)$ distribution asymptotically. A similar procedure can be followed to obtain outlier-robust tests against the alternative of nonlinear ARCH. For example, a robust test against LSTARCH($q$) can be computed as $TR^2$ with $R^2$ from a regression of $\psi(\hat{\xi}_t)^2$ on $\psi(\hat{\xi}_{t-1})^2, \ldots, \psi(\hat{\xi}_{t-q})^2$ and $\psi(\hat{\xi}_{t-1})^3, \ldots, \psi(\hat{\xi}_{t-q})^3$. 

M.-Y. Chen GARCH
Recall that \( r_t = \log(p_t) - \log(p_{t-1}) \) be the log return of an asset at time \( t \), where \( p_t \) is the price level of the asset at \( t \). Suppose the conditional mean and conditional variance of \( r_t \) given \( \mathcal{F}_{t-1} \) are

\[
E(r_t|\mathcal{F}_{t-1}) = \mu_t = G(x_t; \varphi) \\
\text{var}(r_t|\mathcal{F}_{t-1}) = E[(r_t - \mu_t)^2|\mathcal{F}_{t-1}] = h_t,
\]

where \( G(x_t; \varphi) \) is the conditional mean function with parameters \( \varphi \) for exogenous regressors \( x_t \).
The function $G(\cdot)$ is assumed at least twice continuously differentiable. Then

$$r_t = G(x_t; \varphi) + \epsilon_t.$$

As to the conditional variance, the model for time-varying variance is assumed to have parameters $\psi = (\alpha_0, \alpha_i, \beta_i)'$. Denote $\theta \equiv (\varphi', \psi')'$ and their true values as $\theta_0 = (\varphi'_0, \psi'_0)'$. Given the assumed pdf $f(z_t)$ for $z_t$, the parameters in $\theta$ can conveniently be estimated by maximum likelihood (ML).
The conditional log likelihood for the \( t \)th observation is

\[ l_t(\theta) = \ln f(\epsilon_t / \sqrt{h_t}) - \ln \sqrt{h_t}. \]

For example, if \( z_t \) is normally distributed,

\[ l_t(\theta) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sqrt{h_t} - \frac{\epsilon_t^2}{2h_t}. \]

The maximum likelihood estimator (MLE) for \( \theta \), denoted as \( \hat{\theta}_{\text{ML}} \), is found by maximizing the log likelihood for the full sample, which is simply the sum of the conditional log likelihoods.
The MLE solves the first-order condition

\[
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} \overset{\text{set}}{=} 0.
\]

Denote \( s_t(\theta) \equiv \frac{\partial l_t(\theta)}{\theta} \) as the scores. As \( \epsilon_t = z_t \sqrt{h_t} \), the score can be decomposed as \( s_t(\theta) = \left( \frac{\partial l_t(\theta)}{\partial \varphi}, \frac{\partial l_t(\theta)}{\partial \psi} \right) \), where

\[
\frac{\partial l_t(\theta)}{\partial \varphi} = \frac{\epsilon_t}{h_t} \frac{\partial G(x_t; \varphi)}{\partial \varphi} + \frac{1}{2h_t} \left( \frac{\epsilon_t^2}{h_t} - 1 \right) \frac{\partial h_t}{\partial \varphi},
\]

\[
\frac{\partial l_t(\theta)}{\partial \psi} = \frac{1}{2h_t} \left( \frac{\epsilon_t^2}{h_t} - 1 \right) \frac{\partial h_t}{\partial \psi}.
\]
If the conditional distribution \( f(\cdot) \) is correctly specified, the resulting estimates are consistent and asymptotically normal. That is,

\[
\sqrt{T}(\hat{\theta}_{ML} - \theta_0) \rightarrow^d N(0, A_0^{-1}),
\]

where \( A_0^{-1} \) is the inverse of the information matrix evaluated at the true parameter vector \( \theta_0 \),

\[
A_0 = \frac{1}{T} \sum_{t=1}^{T} E \left( \frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} \right) = \frac{1}{T} \sum_{t=1}^{T} E(H_t(\theta_0)).
\]
The negative definite of the matrix of second-order partial derivatives of the log likelihood with respect to the parameters, \( H_t(\theta) \equiv -\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \), is called the Hessian matrix. The matrix \( A_0 \) can be consistently estimated by its sample analogue

\[
A_T(\hat{\theta}_{ML}) = -\frac{1}{T} \sum_{t=1}^{T} \left( \frac{\partial^2 l_t(\hat{\theta}_{ML})}{\partial \theta \partial \theta'} \right).
\]
As argued previously, conditional normality of $\epsilon_t$ is often not a very realistic assumption for high-frequency financial time series, as the resulting model fails to capture the kurtosis in the data. Instead, one sometimes assumes that $z_t$ follows a Student-$t$ distribution or any other distributions, say stable distributions. The parameters in the GARCH models can then be estimated by maximizing the log likelihood corresponding with the particular distribution. As one can never be sure that the specified distribution of $z_t$ is the correct one, an alternative approach is to ignore the problem and base the likelihood on the normal distribution. This method usually is referred to as quasi maximum likelihood estimator (QMLE). In general, the resulting estimates still are consistent and asymptotically normal.
However, the asymptotic variance-covariance of the QMLE has to be adjusted as $A_0^{-1}B_0A_0^{-1}$, where $B_0$ is the expected value of the outer product of the gradient matrix

$$B_0 = \frac{1}{T} \sum_{t=1}^{T} E \left( \frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'} \right) = \frac{1}{T} \sum_{t=1}^{T} E[s_t(\theta_0)s_t(\theta_0)'].$$

The asymptotic covariance matrix can be estimated consistently by using the sample analogues for both $A_0$, given above, and $B_0$, given by

$$B_T(\hat{\theta}_{\text{QMLE}}) = \frac{1}{T} \left( \frac{\partial l_t(\hat{\theta}_{\text{QMLE}})}{\partial \theta} \frac{\partial l_t(\hat{\theta}_{\text{QMLE}})}{\partial \theta'} \right) = \frac{1}{T} \sum_{t=1}^{T} s_t(\hat{\theta}_{\text{QMLE}})s_t(\hat{\theta}_{\text{QMLE}})'.$$
Robust Estimation

Several approaches to handle outliers in GARCH models have been investigated. Sakata and White (1998) consider outlier-robust estimation for GARCH models, using the technique mentioned previously. Hotta and Tsay (1998) derive test-statistics to detect outliers in a GARCH model, distinguishing between outliers which do and which do not affect the conditional volatility. Franses and Ghijsels (1999) apply the outlier detection method of Chen and Liu (1993) to GARCH models. For simplicity, we consider a GARCH (1,1) model

\[
\epsilon_t = z_t \sqrt{h_t}
\]

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

where \(\alpha_0 > 0, \alpha_1 > 0, \beta_1 > 0\) and \(\alpha_1 + \beta_1 < 1\) such that the model is covariance-stationary.
As the GARCH(1,1) model can be represented as an ARMA(1,1) model for $\epsilon_t$,

$$
\epsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\epsilon_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1},
$$

where $\nu_t = \epsilon_t^2 - h_t$. In additional, (1) can be rewritten as

$$
[1 - (\alpha_1 + \beta_1)L]\epsilon_t^2 = \alpha_0 + (\beta_1 L)\nu_t
$$

and then

$$
\nu_t = \frac{-\alpha_0}{1 - \beta_1 L} + \frac{1 - (\alpha_1 + \beta_1)L}{1 - \beta_1 L} \epsilon_t^2,
$$

where $L$ is the lag operator.
Define the lag polynomial $\pi(L)$ as

$$
\pi(L) = \frac{1 - (\alpha_1 + \beta_1)L}{1 - \beta_1 L} = (1 - \beta_1 L)^{-1}[1 - (\alpha_1 + \beta_1)L]
= (1 + \beta_1 L + \beta_1^2 L^2 + \beta_1^3 L^3 + \cdots)[1 - (\alpha_1 + \beta_1)L]
= 1 - \alpha_1 L - \alpha_1 \beta_1 L^2 - \alpha_1 \beta_1^2 L^3 - \cdots
= \sum_{k=0} \pi_k.
$$
Suppose that instead of the true series $\epsilon_t$ one observes the series $e_t$ which is defined by

$$e_t^2 = \epsilon_t^2 + \eta I[t = \tau],$$

where $I[t = \tau]$ is the indicator function defined as $I[t = \tau] = 1$ if $t = \tau$ and zero otherwise, and $\eta$ is a nonzero constants indicating the magnitude of the outlier. Applying the GARCH(1,1) model to the observed series $e_t^2$, it is straightforward to show that the corresponding residuals $u_t$ are given by

$$u_t = \frac{-\alpha_0}{1 - \beta_1 L} + \frac{1 - (\alpha_1 + \beta_1)L}{1 - \beta_1 L} e_t^2$$

$$= \frac{-\alpha_0}{1 - \beta_1 L} + \pi(L)e_t^2$$

$$= \frac{-\alpha_0}{1 - \beta_1 L} + \pi(L)(e_t^2 + \eta I[t = \tau])$$

$$= \frac{-\alpha_0}{1 - \beta_1 L} + \pi(L)e_t^2 + \pi(L)\eta I[t = \tau]$$

$$= \nu_t + \pi(L)\eta I[t = \tau].$$
The last line can be interpreted as a regression for $u_t$, that is

$$u_t = \eta x_t + \nu_t,$$

with

$$x_t = \begin{cases} 0, & \text{for } t < \tau, \\ 1, & \text{for } t = \tau, \\ -\pi_k, & \text{for } k = 1, 2, \ldots. \end{cases}$$

The magnitude $\eta$ of the outlier at time $t = \tau$ then can be estimated as

$$\hat{\eta}(\tau) = \left( \sum_{t=1}^{T} x_t^2 \right)^{-1} \left( \sum_{t=1}^{T} x_t u_t \right).$$
For fixed $\tau$, the $t$-statistic of $\hat{\eta}(\tau)$, denoted as $t_{\hat{\eta}(\tau)}$, has an asymptotic standard normal distribution. In practice, the timing of possible outliers is of course unknown. In that case, an intuitive plausible test-statistic is the maximum of the absolute values of the $t$-statistic over the entire sample, that is

$$t_{\max}(\hat{\eta}) = \max_{1 \leq \tau \leq T} |t_{\hat{\eta}(\tau)}|.$$ 

The distribution of $t_{\max}(\hat{\eta})$ is nonstandard and deserves to be investigated.
outlier detection method for GARCH(1,1)

1. Estimate a GARCH(1,1) model for the observed series $e_t$ and obtain estimates of the conditional variance $\hat{h}_t$ and $\hat{\nu}_t \equiv e_t^2 - \hat{h}_t$.

2. Obtain estimates $\hat{\eta}(\tau)$ for all possible $\tau = 1, \ldots, T$ and compute the test-statistic $t_{\text{max}}(\hat{\eta})$. If the value of $t_{\text{max}}(\hat{\eta})$ exceeds the pre-specified critical value $C$ and outlier is detected at the observation for which the $t$-statistic of $\hat{\eta}$ is maximized (in absolute value), say $\hat{\tau}$.

3. Replace $e_{\hat{\tau}}^2$ with $e_{\hat{\tau}}^{*2} \equiv e_{\hat{\tau}}^2 - \hat{\eta}(\hat{\tau})$ and define the outlier corrected series $e_t^*$ as $e_t^* = e_t$ for $t \neq \hat{\tau}$ and

$$e_{\hat{\tau}}^{*} = \text{sgn}(e_{\hat{\tau}}) \sqrt{e_{\hat{\tau}}^{*2}}.$$

4. Return to step (1) to estimate a GARCH(1,1) model for the series $e_t^*$. 
Diagnostic Checking

One of the assumptions which is made in GARCH models is that the innovations $z_t = \epsilon_t h_t^{-1/2}$ are independent and identically distributed. Hence, if the model is correctly specified, the standardized residuals $\hat{z}_t = \hat{\epsilon}_t \hat{h}_t^{-1/2}$ should possess the classical properties of well behaved regression errors, such as constant variance, lack of serial correlation, and so on. One particular interest is to test whether the standardized residuals still contain signs of conditional heterokedasticity.
Lundbergh and Teräsvirta (1998) suggest the LM test for remaining ARCH\(^{(m)}\) in \(\hat{z}_t\) as \(TR^2\), where \(R^2\) is from the auxiliary regression

\[
\hat{z}_t^2 = \phi_0 + \phi_1 \hat{z}_{t-1} + \cdots + \phi_m \hat{z}_{t-m} + \lambda' \hat{x}_t + u_t,
\]

where the vector \(\hat{x}_t\) consists of the partial derivatives of the conditional variance \(h_t\) with respect to the parameters in the original GARCH model, evaluated under the null hypothesis, that is, \(\hat{x}_t \equiv \hat{h}_t^{-1} \partial \hat{h}_t / \partial \alpha\).
For example, in the case of a GARCH(1,1) model

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

it follows that

\[ \frac{\partial h_t}{\partial \alpha} = (1, \epsilon_{t-1}^2, h_{t-1})' + \beta_1 \frac{\partial h_{t-1}}{\partial \alpha}. \]

Then

\[ \hat{x}_t = \left( \frac{\sum_{i=1}^{t-1} \hat{\beta}_1^{i-1}}{\hat{h}_t}, \frac{\sum_{i=1}^{t-1} \hat{\beta}_1^{i-1} \hat{\epsilon}_{t-i}^2}{\hat{h}_t}, \frac{\sum_{i=1}^{t-1} \hat{\beta}_1^{i-1} \hat{h}_{t-i}}{\hat{h}_t} \right)' . \]

The test-statistic for the null \( H_0 : \phi_1 = \cdots = \phi_m = 0 \) is asymptotically \( \chi^2 \) distributed with \( m \) degree of freedom.
Testing for Higher-order GARCH

The LM-statistic to test a \( \text{GARCH}(p, q) \) specification against either a \( \text{GARCH}(p + r, q) \) or \( \text{GARCH}(p, q + s) \) alternative has been discussed by Bollerslev (1986). The test statistic is \( TR^2 \) where \( R^2 \) is from the regression

\[
\hat{z}_t^2 = \phi_0 + \phi_1 \hat{\epsilon}_{t-q-1}^2 + \cdots + \phi_r \hat{\epsilon}_{t-q-r}^2 + \lambda' \hat{x}_t + u_t,
\]

or

\[
\hat{z}_t = \phi_0 + \phi_1 \hat{z}_{t-p-1} + \cdots + \phi_s \hat{z}_{t-p-s} + \lambda' \hat{x}_t + u_t.
\]
Testing Parameter Constancy

Consider a standard Gaussian GARCH(\(p,q\)) model:

\[ \epsilon_t \mid \Phi_{t-1} \sim N(0, h_t), \]

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

where \( p \geq 0, q \geq 0, \alpha_0, \alpha_i \geq 0, \beta_i \geq 0 \) for all \( i \). Let \([T \delta]\) denote the integer part of \( T \delta \). The quasi log-likelihood function under the alternative of one-time parameter shift in the variance equation is

\[ L_T(r_1, r_2, \delta \mid .) = T^{-1} \sum_{t=1}^{[T \delta]} \ln f_{1t} + T^{-1} \sum_{t=[T \delta] + 1}^{T} \ln f_{2t}, \quad (2) \]

where \( \delta \) is the break-point parameter within the interval \((0, 1)\).
\[
\begin{align*}
\ln f_{1t} &= \frac{1}{2}[-\ln h_{1t} - h_{1t}^{-1}\epsilon_t^2], \quad h_{1t} = r_1' Z_t, \\
\ln f_{2t} &= \frac{1}{2}[-\ln h_{2t} - h_{2t}^{-1}\epsilon_t^2], \quad h_{2t} = r_2' Z_t, \\
r_1 &= (\alpha_{10}, \alpha_{11}, \ldots, \alpha_{1q}, \beta_{11}, \ldots, \beta_{1p})' \\
r_2 &= (\alpha_{20}, \alpha_{21}, \ldots, \alpha_{2q}, \beta_{21}, \ldots, \beta_{2p})' \\
Z_t &= (1, \epsilon_{t-1}^2, \ldots, \epsilon_{t-q}^2, h_{t-1}, \ldots, h_{t-p})'.
\end{align*}
\]
Let $k = p + q + 1$ be the number of parameters in the variance equation and $r = (r'_1, r'_2)'$. Based on (5), the score is given by

$$(\partial L_T / \partial r)_{2k \times 1} = [(\partial L_T / \partial r_1)', (\partial L_T / \partial r_2)']',$$

where

$$\partial L_T / \partial r_1$$

$$= T^{-1} \sum_{t=1}^{[T\delta]} \frac{1}{2} \left[ -h_{1t}^{-1} (\partial h_{1t} / \partial r_1) - (0 - \epsilon_t^2 / h_{1t}^2)(\partial h_{1t} / \partial r_1) \right]$$

$$= T^{-1} \sum_{t=1}^{[T\delta]} \frac{1}{2} \left( \partial h_{1t} / \partial r_1 \right) \left[ \epsilon_t^2 h_{1t}^{-2} - h_{1t}^{-1} \right]$$

$$= T^{-1} \sum_{t=1}^{[T\delta]} \frac{1}{2} h_{1t}^{-1} \left( \partial h_{1t} / \partial r_1 \right) \left[ \epsilon_t^2 / h_{1t} - 1 \right]$$
\[
\partial L_T/\partial r_2 \\
= T^{-1} \sum_{t=[T\delta]+1}^{T} \frac{1}{2} [ -h_{2t}^{-1} (\partial h_{2t}/\partial r_2) - (0 - \epsilon_t^2/h_{2t}^2)(\partial h_{2t}/\partial r_2) ] \\
= T^{-1} \sum_{t=[T\delta]+1}^{T} \frac{1}{2} (\partial h_{2t}/\partial r_2) [\epsilon_t^2 h_{2t}^{-2} - h_{2t}^{-1}] \\
= T^{-1} \sum_{t=[T\delta]+1}^{T} \frac{1}{2} h_{2t}^{-1} (\partial h_{2t}/\partial r_2) [\epsilon_t^2/h_{2t} - 1].
\]
We are interested in testing the null hypothesis of no structural change \( H_0 : r_1 = r_2 \) versus \( H_0 : r_1 \neq r_2 \), where \( r_1 \) is the pre-break parameter vector of order \( k \times 1 \) in the variance equation and \( r_2 \) is the post-break parameters of the same order \( k \). The LM test requires estimating the null model only. The null hypothesis under testing is \( H_0 : r_1 = r_2 = r_0 \). We use the restricted quasi-maximum likelihood estimators (QMLE) \( \tilde{r} \) to construct the prototype LM statistic.
By Aitcheson and Silvey (1958) we have

\[ \text{LM}_T(\delta) \]

\[ = T[\partial L_T(\tilde{r}_T, \delta)/\partial r] A_0^{-1}(\delta) R'[RC_0(\delta) R']^{-1} RA_0^{-1}(\delta) [\partial L_T(\tilde{r}_T, \delta)/\partial r] \]

\[ = T\tilde{\lambda}_T'[RA_0^{-1}(\delta) R'][RC_0(\delta) R']^{-1} [RA_0^{-1}(\delta) R']\tilde{\lambda}_T, \tag{3} \]

where $\tilde{r}_T$ is the quasi-maximum likelihood estimators for $r$ under the null model, $R = [I_k - I_k]$, $\tilde{\lambda}_T$ is a $k \times 1$ Lagrange multiplier, $A_0(\delta) \equiv \lim_{T \to \infty} E[\partial^2 L_T(r_0, r_0, \delta)/\partial r \partial r']$, $B_0(\delta) = \lim_{T \to \infty} \text{var}[T^{1/2} \partial L_T(r_0, r_0, \delta)/\partial r]$ and $C_0(\delta) = A_0^{-1} B_0 A_0^{-1}$. 
Given the process

\[ \{v_t \equiv h_{0t}^{-1} \left( \frac{\partial h_{0T}}{\partial r} \right) \left( \frac{\epsilon_t^2}{h_{0t}} - 1 \right) \}, \quad h_{0t} \equiv r_{0}^{'}Z_{t}, \]

obeys the functional central limit theorem, that is

\[ T^{-1/2}V_0^{-1/2}S_{[T\delta]} \Rightarrow W_k(\delta), \]

where \( V_0 = \lim_{T \to \infty} T^{-1} E\left[ \left( \sum_{t=1}^{T} v_t \right) \left( \sum_{t=1}^{T} v_t \right) \right], \quad S_{[T\delta]} = \sum_{t=1}^{[T\delta]} v_t, \) and \( W_k(\delta) \) is a standard \( k \)-dimensional Wiener process, the LM statistic for detecting one-time parameter shift is

\[ LM_T = \max_{\delta \in \Pi} LM_T(\delta) \Rightarrow \sup_{\delta \in \Pi} B_k^2(\delta), \]

where

\[ \sup_{\delta \in \Pi} B_k^2(\delta) \equiv \frac{[W_k(\delta) - \delta W_k(1)]'[W_k(\delta) - \delta W_k(1)]}{\delta(1 - \delta)} \]

known as the square of a tied-down Bessel process of order \( k \).
It follows from the continuous mapping theorem (Billingsley, 1968) that \( \sup_{\delta \in \Pi} LM(\delta) \Rightarrow \sup_{\delta \in \Pi} B_k^2(\delta) \), where \( \Pi \) is a prescribed subset of \([0, 1]\) to prevent the limiting distribution from diverging.