Time Series Analysis: Conditional Volatility Models (III)

Mei-Yuan Chen
Department of Finance
National Chung Hsing University

February 25, 2013
Forecasting Volatility

The presence of time-varying volatility has some pronounced consequences for out-of-sample forecasting. For simplicity, we consider the following model

\[ r_t = \phi_1 r_{t-1} + \epsilon_t \]
\[ \epsilon_t = z_t \sqrt{h_t} \]
\[ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}. \]

The general case of ARMA\((k, l)\)-GARCH\((p, q)\) models is discussed in Baillie and Bollerslev (1992).
Forecasting the Conditional Mean in the Presence of Conditional Heteroskedasticity

Let $\hat{r}_{t+h|t}$ denote the $h$-step-ahead forecast of $r_t$ which minimizes the squared prediction errors (SPE)

$$\text{SPE}(h) \equiv E[e_{t+h|t}^2] = E[(r_{t+h} - \hat{r}_{t+h|t})^2],$$

where $e_{t+h|t}$ is the $h$-step-ahead forecast error. Baillie and Bollerslev (1992) show that the forecast that minimizes SPE is the same irrespective of whether the shocks $\epsilon_t$ are conditional homoscedastic or conditional heteroskedastic. Thus, the optimal $h$-step-ahead forecast of $r_{t+h}$ is its conditional expectation at time $t$, that is

$$\hat{r}_{t+h|t} = E[r_{t+h} | F_t].$$
For the AR(1) conditional mean, the optimal 1-step-ahead forecast is given by 
\( \hat{r}_{t+1|t} = \phi_1 r_t \) and the optimal \( h \)-step-ahead forecast is as 
\[
\hat{r}_{t+h|t} = \phi_1 \hat{r}_{t+h-1|t} \\
= \phi_1 [\phi_1 \hat{r}_{t+h-2|t}] \\
\vdots \\
= \phi_h^h r_t. 
\]

For the \( h \)-step-ahead prediction error, it follows that 
\[
e_{t+h|t} = r_{t+h} - \hat{r}_{t+h|t} = \phi_1 r_{t+h-1|t} + e_{t+h} - \phi_1^h r_t \\
= \phi_1 r_{t+h-1} + e_{t+h} - \phi_1^h r_t \\
= \ldots \\
= \phi_1^h r_t + \sum_{i=1}^{h} \phi_1^{h-i} e_{t+i} - \phi_1^h r_t \\
= \sum_{i=1}^{h} \phi_1^{h-i} e_{t+i}. 
\]
The conditional SPE of $e_{t+h|t}$ is given by

$$E[e_{t+h|t}^2|\mathcal{F}_t] = E \left[ \left( \sum_{i=1}^{h} \phi_1^{h-i} \epsilon_{t+i} \right)^2 | \mathcal{F}_t \right]$$

$$= \sum_{i=1}^{h} \phi_1^{2(h-i)} E[\epsilon_{t+i}^2 | \mathcal{F}_t]$$

$$= \sum_{i=1}^{h} \phi_1^{2(h-i)} E[h_{t+i} | \mathcal{F}_t].$$

In the case of homoscedastic errors, the conditional SPE for the optimal $h$-step-ahead forecast is constant, as $E[h_{t+i} | \mathcal{F}_t]$ is constant and equals to the unconditional variance of $\epsilon_t$, $\sigma^2$. 

M.-Y. Chen
However in the case of heteroskedastic errors, the conditional SPE is varying over time. Since

\[
E[e_{t+h|t}^2 | \mathcal{F}_t] = \sum_{i=1}^{h} \phi_1^{2(h-i)} \sigma^2 + \sum_{i=1}^{h} \phi_1^{2(h-i)} [E(h_{t+i} | \mathcal{F}_t) - \sigma^2],
\]

the conditional SPE in the case of heteroskedastic errors can be both larger and smaller than in the case of homoscedastic errors.
Recall that in the homoscedastic case, the SPE converges to the unconditional variance of $r_t$ as the forecast horizon $h$ increases, that is

$$\lim_{h \to \infty} E[e_{t+h|t}^2|\mathcal{F}_t] = \lim_{h \to \infty} \sum_{i=1}^{h} \phi_1^{2(h-i)} \sigma^2 = \frac{\sigma^2}{1 - \phi_1^2} \equiv \sigma_r^2.$$ 

Moreover, the convergence is monotonic, in the sense that the $h$-step-ahead SPE is always smaller than the unconditional variance $\sigma_r^2$, while the $h$-step-ahead SPE is larger than the $(h - 1)$-step SPE for all finite horizons $h$. The convergence of the SPE to the unconditional variance of the time series also holds in the present case of heteroskedastic errors.
Forecasting the Conditional Variance

In the case of GARCH(1,1) model, the conditional expectation of $h_{t+s}$, i.e., the optimal $s$-step-ahead forecast of the conditional variance, can be computed recursively from

$$
\hat{h}_{t+s|t} = \alpha_0 + \alpha_1 \hat{\epsilon}_{t+s-1|t}^2 + \beta_1 \hat{h}_{t+s-1|t},
$$

where

1. $\hat{\epsilon}_{t+i|t}^2 = \hat{h}_{t+i|t}$ for $i > 0$ by definition,
2. $\hat{\epsilon}_{t+i|t}^2 = \epsilon_{t+i}^2$ and $\hat{h}_{t+i|t} = h_{t+i}$ for $i \leq 0$. 
Alternatively, by recursive substitution we obtain

\[
\hat{h}_{t+s|t} = \alpha_0 + \alpha_1 \hat{\epsilon}_{t+s-1|t}^2 + \beta_1 \hat{h}_{t+s-1|t}
\]

\[
= \alpha_0 + (\alpha_1 + \beta_1) \hat{h}_{t+s-1|t}
\]

\[
= \alpha_0 + (\alpha_1 + \beta_1) [\alpha_0 + (\alpha_1 + \beta_1) \hat{h}_{t+s-2|t}]
\]

\[
= [\alpha_0 + (\alpha_1 + \beta_1) \alpha_0]
\]

\[
+ (\alpha_1 + \beta_1)^2 [\alpha_0 + (\alpha_1 + \beta_1) \hat{h}_{t+s-3|t}]
\]

\[
= [\alpha_0 + (\alpha_1 + \beta_1) \alpha_0 + (\alpha_1 + \beta_1)^2 \alpha_0]
\]

\[
+ (\alpha_1 + \beta_1)^3 [\alpha_0 + (\alpha_1 + \beta_1) \hat{h}_{t+s-4|t}]
\]

\[
\vdots
\]

\[
= \alpha_0 \sum_{i=0}^{s-1} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{s-1} \hat{h}_{t+1|t}.
\]
Given the formula, $\sum_{i=1}^{s-1} r^i = (1 - r^{s-1})/(1 - r)$, the last equation becomes

$$
\hat{h}_{t+s|t} = \alpha_0 \frac{1 - (\alpha_1 + \beta_1)^{s-1}}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{s-1} \hat{h}_{t+1|t}
$$

$$
= \frac{\alpha_0}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{s-1} \left( \hat{h}_{t+1|t} - \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \right)
$$

$$
= \sigma^2 + (\alpha_1 + \beta_1)^{s-1} (\hat{h}_{t+1|t} - \sigma^2),
$$

where $\sigma^2 = \alpha_0/(1 - \alpha_1 - \beta_1)$ is the unconditional variance of $\epsilon_t$. 

M.-Y. Chen  
GARCH
The $s$-step-ahead forecasting error is $v_{t+s|t} = h_{t+s} - \hat{h}_{t+s|t}$. As in the GARCH(1,1) model, $h_{t+s} = \alpha_0 + \alpha_1 \epsilon_{t+s-1}^2 + \beta_1 h_{t+s-1}$, we obtain

$$v_{t+s|t} \equiv h_{t+s} - \hat{h}_{t+s|t}$$

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

$$- [\alpha_0 + \alpha_1 \hat{\epsilon}_{t+s-1|t}^2 + \beta_1 \hat{h}_{t+s-1|t}]$$

$$= \alpha_1 (\epsilon_{t+s-1}^2 - \hat{\epsilon}_{t+s-1|t}^2) + \beta_1 (h_{t+s-1} - \hat{h}_{t+s-1|t})$$

$$= \alpha_1 (\epsilon_{t+s-1}^2 - h_{t+s-1} + h_{t+s-1} - \hat{h}_{t+s-1|t})$$

$$+ \beta_1 (h_{t+s-1} - \hat{h}_{t+s-1|t})$$

$$= \alpha_1 \nu_{t+s-1} + (\alpha_1 + \beta_1) v_{t+s-1|t},$$

where we have used the fact that $\hat{\epsilon}_{t+i|t}^2 = \hat{h}_{t+i|t}$ for $i > 0$ and the definition $\nu_t = \epsilon_t^2 - h_t$. 

M.-Y. Chen  GARCH
By continued recursive substitution, we have

\[ v_{t+s|t} = \alpha_1 v_{t+s-1} + (\alpha_1 + \beta_1)v_{t+s-1|t} \]
\[ = (\alpha_1 + \beta_1)[\alpha_1 v_{t+s-2} + (\alpha_1 + \beta_1)v_{t+s-2|t}] + \alpha_1 v_{t+s-1} \]
\[ = (\alpha_1 + \beta_1)^2[\alpha_1 v_{t+s-3} + (\alpha_1 + \beta_1)v_{t+s-3|t}] + [\alpha_1 v_{t+s-1} + (\alpha_1 + \beta_1) \ldots \]
\[ = \alpha_1 \sum_{i=1}^{s-1} (\alpha_1 + \beta_1)^{i-1} v_{t+s-i}. \]

As the \(\nu_t\)s are serially uncorrelated and can be written as \(\nu_t = h_t(z^2_t - 1)\), it follows that the conditional SPE of the \(s\)-step-ahead forecast \(\hat{h}_{t+s|t}\) is given by

\[ E[v^2_{t+s|t}|\mathcal{F}_t] = (\kappa - 1)\alpha_1^2 \sum_{i=1}^{s-1} (\alpha_1 + \beta_1)^{2(i-1)} E[h^2_{t+s-i}|\mathcal{F}_t], \]

where \(\kappa\) is the kurtosis of \(z_t\).
Forecasting in Nonlinear GARCH Models

For the GJR-GARCH(1,1) model, 

\[ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 [1 - I(\epsilon_{t-1} > 0)] + \gamma_1 \epsilon_{t-1}^2 I(\epsilon_{t-1} > 0) + \beta_1 h_{t-1}. \]

Assuming that the distribution of \( z_t \) is symmetric around 0, the 2-step-ahead forecast of \( h_{t+2} \) is given by

\[
\hat{h}_{t+2|t} = E\{ \alpha_0 + \alpha_1 \epsilon_{t+1}^2 [1 - I(\epsilon_{t+1} > 0)] + \gamma_1 \epsilon_{t+1}^2 I(\epsilon_{t+1} > 0) + \beta_1 h_{t+1} | \mathcal{F}_t \}
= \alpha_0 + [(\alpha_1 + \gamma_1)/2 + \beta_1] h_{t+1},
\]

which follows from observing that \( \epsilon_{t+1}^2 \) and the indicator function \( I(\epsilon_{t+1} > 0) \) are uncorrelated and \( E[I(\epsilon_{t+1} > 0)] = P(\epsilon_{t+1} > 0) = 0.5 \), and again using \( E(\epsilon_{t+1}^2 | \mathcal{F}_t) = h_{t+1} \).
In general, $s$-step-ahead forecasts can be computed either recursively as

$$
\hat{h}_{t+s|t} = \alpha_0 + \left[ (\alpha_1 + \gamma_1)/2 + \beta_1 \right] \hat{h}_{t+s-1|t},
$$

(1)

or directly from

$$
\hat{h}_{t+s|t} = \alpha_0 \sum_{i=0}^{s-1} \left[ (\alpha_1 + \gamma_1)/2 + \beta_1 \right]^i + \left[ (\alpha_1 + \gamma_1)/2 + \beta_1 \right]^{s-1} \hat{h}_{t+1|t}.
$$

(2)
For the LSTGARCH(1,1) (logist smoothing transition GARCH(1,1)) model,

\[ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 [1 - F(\epsilon_{t-1})] + \gamma_1 \epsilon_{t-1}^2 F(\epsilon_{t-1}) + \beta_1 h_{t-1}, \]

where \( F(\epsilon_{t-1}) = [1 + \exp(\theta \epsilon_{t-1})]^{-1} \). As \( \epsilon_{t+i} \) and \( F(\epsilon_{t+i}) \) are uncorrelated, combined with the fact that \( F(\epsilon_{t+i}) \) is anti-symmetric around the expected value of \( \epsilon_{t+i} \), \( E(\epsilon_{t+i}) = 0 \), and thus \( E[F(\epsilon_{t+i})] = F[E(\epsilon_{t+i})] = 0.5 \). In general, a function \( G(x) \) is said to be anti-symmetric around \( a \) if \( G(x + a) - G(a) = -[G(-x + a) - G(a)] \) for all \( x \). If furthermore \( x \) is symmetrically distributed with mean \( a \) it holds that \( E[G(x)] = G[E(x)] = G(a) \). Then the \( s \)-step-ahead forecasts of LSTGARCH(1,1) is same as in (1) or in (2).
an Anti-symmetric Function around $\alpha$
For the VS-GARCH(1,1) (volatility-Switching GARCH(1,1)) model:

\[
    h_t = (\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1})[1 - I(\epsilon_{t-1} > 0)] \\
    + (\gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \delta_1 h_{t-1})I(\epsilon_{t-1} > 0)
\]  

(3)

and the ANST-GARCH(1,1) (asymmetric nonlinear smooth transition GARCH(1,1)) model:

\[
    h_t = (\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1})[1 - F(\epsilon_{t-1})] \\
    + (\gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \delta_1 h_{t-1})[1 - F(\epsilon_{t-1})].
\]  

(4)
the $s$-step-ahead forecasts can be computed either recursively from

$$\hat{h}_{t+s|t} = \alpha_0 + \left[ (\alpha_1 + \gamma_1)/2 + (\beta_1 + \delta_1)/2 \right] \hat{h}_{t+s-1|t},$$

or directly from

$$\hat{h}_{t+s|t} = \alpha_0 \sum_{i=0}^{s-1} \left[ (\alpha_1 + \gamma_1)/2 + (\beta_1 + \delta_1)/2 \right]^i$$

$$+ \left[ (\alpha_1 + \gamma_1)/2 + (\beta_1 + \delta_1)/2 \right]^{s-1} \hat{h}_{t+1|t}. $$
For the QGARCH(1,1) (quadratic GARCH(1,1)) model,

\[ h_t = \alpha_0 + \gamma_1 \epsilon_{t-1} + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}. \]

As the asymmetric term \( \gamma_1 \epsilon_{t-1} \) does not affect the forecasts for the conditional variance since the conditional expectation of \( \epsilon_{t+i} \) with \( i > 0 \) is zero by assumption. Hence, point forecasts for the conditional variance can be obtained either recursively as (1):

\[ \hat{h}_{t+s|t} = \alpha_0 + [(\alpha_1 + \gamma_1)/2 + \beta_1]\hat{h}_{t+s-1|t}, \]

or directly from (2):

\[ \hat{h}_{t+s|t} = \alpha_0 \sum_{i=0}^{s-1} [(\alpha_1 + \gamma_1)/2 + \beta_1]^i + [(\alpha_1 + \gamma_1)/2 + \beta_1]^{s-1}\hat{h}_{t+1|t}. \]
For the ESTGARCH (exponential smoothing transition GARCH) model, the exponential function
\[ F(\epsilon_{t+i}) = 1 - \exp(-\theta \epsilon_{t+i}^2) \]
is correlated with \( \epsilon_{t+i}^2 \), and it is not in the case that
\[ E[F(\epsilon_{t+i})] = F[E(\epsilon_{t+i})] \]. Therefore, it is not possible to derive a recursive or direct formula for the
\( s \)-step-ahead \( \hat{h}_{t+s|t} \) in this case. Instead, forecasts for future conditional variance have to be obtained by means of simulation. As to the Markov-Switching GARCH models, the analytical expression for multiple-step-ahead forecasts of the conditional variance can be obtained by exploiting the properties of the Markov-process, e.g., Hamilton and Lin (1996), Dueker (1997), and Klaassen (1999).
Evaluating Forecasts of Conditional Volatility

As discussed previously, it is quite difficult to select a suitable nonlinear GARCH model on the basis of specification tests only. The out-of-sample forecasting ability of various GARCH models is an alternative approach to judge the adequacy of different models. Suppose a GARCH model has been estimated using a sample of $T$ observations, whereas observations at $t = T + 1, \ldots, T + m - s - 1$ are held back for evaluation of $s$-step-ahead forecasts for the conditional variance.
Most studies use statistical criteria such as the mean squared prediction error (MSPE), which for a set of $m$ $s$-step-ahead forecasts is computed as

$$\text{MSPE} = \frac{1}{m} \sum_{j=0}^{m-1} \left( \hat{h}_{T+s+j|T+j} - h_{T+s+j} \right)^2$$

or the $R^2$ from the regression

$$h_{T+s+j} = a + b \hat{h}_{T+s+j|T+j} + e_{T+s+j}, \quad j = 0, \ldots, m - 1.$$  

To make these forecast evaluation criteria operational, the unobservable $h_{T+s+j}$ is usually replaced by the squared shock $\epsilon_{T+s+j}$.
A common finding from forecast competitions is that all GARCH models provide seemingly poor volatility forecasts and explain only very little of the variability of asset returns, in the sense that the MSPE is very large while the $R^2$ is very small, typically below 0.1. In addition, the forecasts from GARCH appears to be biased, as it commonly found that $\hat{a} \neq 0$. Anderson and Bollerslev (1998) and Christodoulakis and Satchell (1998) demonstrated that this poor forecasting performance is caused by the fact that the unobservable true volatility $h_{T+s+j}$ is approximated with the squared shock $\epsilon^2_{T+s+j}$. As shown by Anderson and Bollerslev (1998), for a GARCH(1,1) model with a finite unconditional fourth moment the population $R^2$ for $s = 1$ and $h_{T+s+j}$ replaced with $\epsilon^2_{T+s+j}$ is equal to

\[
R^2 = \frac{\alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1}.
\]
As the condition for a finite unconditional fourth moment in the GARCH(1,1) model is given by \( \kappa \alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1 < 1 \), it follows that the population \( R^2 \) is bounded from above by \( 1/\kappa \). Where \( z_t \) is normally distributed, the \( R^2 \) cannot be larger than \( 1/3 \), while the upper bound is even smaller if, for example, \( z_t \) is assumed to be Student-\( t \) distributed.
Christodoulakis and Satchell (1998) explain the occurrence of apparent bias in GARCH volatility forecasts by noting that

\[
\ln(\epsilon^2_{T+s+j}) = \ln(h_{T+s+j}) + \ln(z^2_{T+s+j}),
\]
or

\[
\ln(\epsilon^2_{T+s+j}) - \ln(\hat{h}_{T+s+j}|T+j) = (\ln(h_{T+s+j}) - \ln(\hat{h}_{T+s+j}|T+j)) + \ln(z^2_{T+s+j}).
\]

As \(\ln(x) \approx -(1 - x)\) for small \(x\), the left-hand-side of above equation is approximately equal to the observed bias \(\epsilon^2_{T+s+j} - \hat{h}_{T+s+j}|T+j\). If the GARCH forecasts are unbiased, the first term on the right-hand-side of above equation is equal to zero. Hence, the expected observed bias is equal to \(E[\ln(z^2_{T+s+j})]\), which in the case of normally distributed \(z_t\) is equal to -1.27.
Anderson and Bollerslev (1998) suggest that a (partial) solution to the above-mentioned problems might be to estimate the unobserved volatility with data which is sampled more frequently than the time series of interest. For example, if $r_t$ is a time series of weekly returns, the corresponding daily returns – if available – might be used to obtain a more accurate measure of the weekly volatility.
Applications of GARCH Models to Estimate VaR
Multivariate Conditional Variance Models

The multivariate GARCH\((p,q)\) regression model can be written as

\[
\mathbf{y}_t = \mathbf{\beta} \mathbf{z}'_t + \mathbf{\epsilon}_t
\]

\[
\mathbf{\epsilon}_t \mid \Psi_{t-1} \sim N(0, \mathbf{H}_t)
\]

where \(\mathbf{\epsilon}_t\) and \(\mathbf{y}_t\) are \(n \times 1\) vectors. \(n\) is the number of random variables. \(\mathbf{z}_t\) is a \(1 \times k\) vector where \(k\) is the number of independent variables. \(\mathbf{\beta}\) is an \(n \times k\) matrix. \(\mathbf{H}_t\) is an \(n \times n\) matrix. \(\Psi_{t-1}\) is the information set available at time \(t - 1\). Three specifications of the conditional variance-covariance matrix, \(\mathbf{H}_t\), are presented as follows.
First, allowing each element of $H_t$ to depend on $q$ lagged values of the squares and cross-products of $\varepsilon_t$, as well as $p$ lagged values of the elements of $H_t$, and a $J \times 1$ vector of weakly exogenous variables $x_t$, the “vec representation” is

$$
\begin{align*}
    h_t &= \text{vec}(H_t); \\
    \tilde{x}_t &= \text{vec}(x_t x_t') ; \\
    \eta_t &= \text{vec}(\varepsilon_t \varepsilon_t') ; \\
    h_t &= C_0 + C_1 \tilde{x}_t + \sum_{i=1}^{q} A_i \eta_{t-i} + \sum_{i=1}^{p} G_i h_{t-i} .
\end{align*}
$$

(5)

If we take no account of exogenous influences, the conditional variance equation becomes

$$
\begin{align*}
    h_t &= C_0 + \sum_{i=1}^{q} A_i \eta_{t-i} + \sum_{i=1}^{p} G_i h_{t-i} .
\end{align*}
$$

(6)
For example, for $n = 2$ and $p = q = 1$, (6) becomes

$$h_t = \begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} C_{01,t} \\ C_{02,t} \\ C_{03,t} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix}.$$  \hspace{1cm} (7)
In the above formulation, we omit the redundant variables such as \(h_{21,t}\) and the coefficients to \(\varepsilon_{2,t-1}\varepsilon_{1,t-1}\) and \(h_{21,t-1}\). After eliminating the redundant terms, a total of \(\left(\frac{n(n+1)}{2}\right)^2\) unique parameters exist in each of the \(A_i\) and \(G_i\) matrices.
Second, in the \textit{vec} model, if the matrices $A_i$ and $G_i$ are assumed to be diagonal, a “diagonal representation” of $H_t$ will be obtained. For example, in the bivariate case, the diagonal model (no volatility spillovers) for GARCH(1,1) is specified as

$$h_t = \begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{21,t} \end{bmatrix} = \begin{bmatrix} C_{01,t} \\ C_{02,t} \\ C_{03,t} \end{bmatrix} + \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix}$$

$$+ \begin{bmatrix} g_{11} & 0 & 0 \\ 0 & g_{22} & 0 \\ 0 & 0 & g_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix},$$

(8)
or

\[ h_{11,t} = C_{01,t} + a_{11} \varepsilon_{1,t-1}^2 + g_{11} h_{11,t-1}; \]
\[ h_{12,t} = C_{02,t} + a_{22} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + g_{22} h_{12,t-1}; \]
\[ h_{22,t} = C_{03,t} + a_{33} \varepsilon_{2,t-1}^2 + g_{33} h_{22,t-1}. \]

This means that a conditional variance depends only on its own lagged squared residuals and its lagged values. In the bivariate GARCH(1,1) model, there are only three parameters in each of the \( A_1 \) and \( G_1 \) matrices, and in the general \( n \)-variate diagonal model there are \(((n(n + 1))/2)\) free parameters in each matrix. No matter which parameterization of \( H_t \) is used, \( H_t \) is required to be positive definite for all values of \( \varepsilon_t \) and \( x_t \) in the sample space.
Baba, Engle, Kraft and Kroner (1990) suggest the following parameterization, known as the “BEKK representation”, which is almost guaranteed to be positive definite:

\[
H_t = C_0^* C_0^* + \sum_{k=1}^{K} C_{1k}^* x_t x_t' C_{1k}^* \nonumber \\
+ \sum_{k=1}^{K} \sum_{i=1}^{q} A_{ik}^* \varepsilon_{t-i} \varepsilon_{t-i}' A_{ik}^* + \sum_{k=1}^{K} \sum_{i=1}^{p} G_{ik}^* H_{t-i} G_{ik}^* \\
= C_0 + \sum_{k=1}^{K} C_{1k}^* x_t x_t' C_{1k}^* \nonumber \\
+ \sum_{k=1}^{K} \sum_{i=1}^{q} A_{ik}^* \varepsilon_{t-i} \varepsilon_{t-i}' A_{ik}^* + \sum_{k=1}^{K} \sum_{i=1}^{p} G_{ik}^* H_{t-i} G_{ik}^* \quad (9)
\]

where \(C_0^*, A_{ik}^*, \) and \(G_{1k}^*\) are \(n \times n\) parameter matrices with \(C_0^*\) triangular; \(G_{1k}^*\) are \(J \times n\) parameter matrices; and the summation limit \(K\) determined the generality of the process.
Due to the triangular $n \times n$ matrix $C_0^*$, $C_0 = [c_{ij}]$ is an $n \times n$ symmetric matrix. It is clear that (9) will be positive definite under very weak conditions. Besides, this presentation is sufficiently general because it includes all positive definite diagonal representations and nearly all positive definite vec representations. For example, for $K = 1$ and no exogenous influences, the conditional variance of GARCH(p,q) model is

$$H_t = C_0 + \sum_{i=1}^{q} A_{i1}^* \varepsilon_{t-i} \varepsilon_{t-i}' A_{i1}^* + \sum_{i=1}^{p} G_{i1}^* H_{t-i} G_{i1}^*.$$
Moreover, for the bivariate GARCH(1,1) model with $K = 1$ and no exogenous influences, the conditional variance-covariance matrix, $H_t$, is

$$H_t = C_0 + A_{11}^* \varepsilon_{t-1} \varepsilon_{t-1}' + G_{11}^* H_{t-1} G_{11}^*$$

$$= \begin{bmatrix} c_{11}^* & c_{12}^* \\ c_{21}^* & c_{22}^* \end{bmatrix} + \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1} \varepsilon_{2,t-1} \\ \varepsilon_{2,t-1} \varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix} + \begin{bmatrix} g_{11}^* & g_{12}^* \\ g_{21}^* & g_{22}^* \end{bmatrix}' H_{t-1} \begin{bmatrix} g_{11}^* & g_{12}^* \\ g_{21}^* & g_{22}^* \end{bmatrix}.$$ 

Comparing it with the vec representation and excluding constants, there are $((n(n + 1))/2)^2$ parameters in the vec representation while there are $n^2$ parameters to be estimated in the BEKK representation.
Impulse Respond Function for Multivariate GARCH Model

As mentioned in Lin (1997), the impulse response function for conditional volatility is defined as the impact of a small perturbation of the $i$th innovation on the future predicted volatility. Because the classes of multivariate GARCH models can be written as a function of $\varepsilon_t \varepsilon'_t$, the response of future volatility to one unit shock in $\varepsilon_t$ will depend on the impact filtered through $\varepsilon_t \varepsilon'_t$. 
Hence, the impulse response function is equal to the derivative of the conditional variance with respect to the vector of $dg(\varepsilon_t \varepsilon'_t)$, where $dg(\varepsilon_t \varepsilon'_t)$ is an $n \times 1$ vector containing diagonal elements of $\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t$. For example, the impulse response function for the vec representation of $H_t$ is defined as

$$R_{s,n} = \frac{\partial \text{vech}(H_{t+s}|t)}{\partial dg(\varepsilon_t \varepsilon'_t)},$$

where $R_{s,n}$ is an $N \times n$ matrix and $N = (n + 1)n/2$. As mentioned previously, there are various possible formulations when $H_t$ is parameterized as a function of past information. Various forms of $H_t$ and their definitions of the impulse response function are summarized as follows.
In vec representation of vector multivariate GARCH(p,q) model,

\[
\text{vech}(H_t) = C_0 + \sum_{i=1}^{q} A_i \text{vech}(\varepsilon_{t-i} \varepsilon_{t-i}') + \sum_{i=1}^{p} G_i \text{vech}(H_{t-i}),
\]

where \(A_i\) and \(G_i\) are \(n(n + 1)/2 \times n(n + 1)/2\) parameter matrices and \(C_0\) is an \(n(n + 1)/2 \times 1\) parameter vector. The impulse response function becomes

\[
R_{s,n} = \frac{\partial \text{vech}(H_{t+s|t})}{\partial dg(\varepsilon_t \varepsilon_t')},
\]

in which the number of parameters equals to \([n(n + 1)]^2(p + q)/4 + n(n + 1)/2\).
In BEKK representation of the generalized multivariate GARCH\((p,q)\) model,

\[
H_t = C_0^{*'}C_0^{*} + \sum_{i=1}^{q} A_i^{*'}\varepsilon_{t-i}\varepsilon'_{t-i}A_i^{*} + \sum_{i=1}^{p} G_i^{*'}H_{t-i}G_i^{*},
\]

where \(A_i\) and \(G_i\) are \(n \times n\) parameter matrices for all \(i\), and \(C^{*}\) is an \(n \times n\) triangular parameter matrix. The corresponding impulse response function is

\[
R_{s,n} = \frac{\partial \text{vech}(H_{t+s|t})}{\partial \text{dg}(\varepsilon_t\varepsilon_t')},
\]

with number of parameters \((p + q)n^2 + n(n + 1)/2\).
In the constant correlation multivariate GARCH \((p,q)\) model,

\[
h_t = C + \sum_{i=1}^{q} A_i u_{t-i} + \sum_{i=1}^{p} G_i h_{t-i}
\]

\[
h_{ij,t} = \rho_{ij} \left( h_{ii,t} h_{jj,t} \right)^{\frac{1}{2}},
\]

where \(H_t = [h_{ij,t}]\), \(h_t = (h_{11,t}, \cdots, h_{nn,t})'\), \(u_t = dg(\varepsilon_t \varepsilon_t')\), \(A_i\) and \(G_i\) are \(n \times n\) parameter matrices for all \(i\), and \(C\) is an \(n \times 1\) parameter vector. The corresponding impulse response function is \(R_{s,n} = \frac{\partial h_{t+s|t}}{\partial u_t}\) with number of parameters \((p + q)n^2 + n(n + 1)/2\).
Testing Spillover Effects

Suppose the multivariate GARCH(1,1) regression model without exogenous variables is estimated. The model is

\[ y_t = c + \varepsilon_t \]

and

\[ \varepsilon_t \mid \Psi_{t-1} \sim N(0, H_t), \]

where \( y_t, c \) and \( \varepsilon_t \) are \( n \times 1 \) vectors and \( n \) is the number of random variables. \( H_t \) is an \( n \times n \) matrix and \( \Psi_{t-1} \) is the information set available at time \( t - 1 \). In this section, the bivariate, four-variate and eight-variate GARCH(1,1) models are considered respectively.
Let the stock returns in Taiwan be $\varepsilon_{1,t}$, the stock returns in New York be $\varepsilon_{2,t}$, $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$, and $H_t = [h_{ij,t}]$ for $i = 1, 2$, $j = 1, 2$. $H_t$ is a symmetric matrix. The BEKK bivariate GARCH(1,1) model for the Taiwan and the New York returns is

$$
\varepsilon_t | \Psi_{t-1} \sim N(0, H_t),
$$

$$
\begin{align*}
h_{11,t} &= c_{11} + a_{11}^* \varepsilon_{1,t-1}^{*2} + 2a_{11}^* a_{21}^* \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{21}^* \varepsilon_{2,t-1}^{*2} \\
 &\quad + g_{11}^* h_{11,t-1} + 2g_{11}^* g_{21}^* h_{12,t-1} + g_{21}^* h_{22,t-1}, \\
h_{12,t} &= h_{21,t} \\
&= c_{12} + a_{11}^* a_{12}^* \varepsilon_{1,t-1}^{*2} + (a_{21}^* a_{12}^* + a_{11}^* a_{22}^*) \varepsilon_{1,t-1} \varepsilon_{2,t-1} \\
 &\quad + a_{21}^* a_{22}^* \varepsilon_{2,t-1}^{*2} \\
 &\quad + g_{11}^* g_{12}^* h_{11,t-1} + g_{21}^* g_{12}^* h_{21,t-1} \\
 &\quad + g_{11}^* g_{22}^* h_{12,t-1} + g_{21}^* g_{22}^* h_{22,t-1}, \\
h_{22,t} &= c_{22} + a_{12}^* \varepsilon_{1,t-1}^{*2} + 2a_{12}^* a_{22}^* \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{22}^* \varepsilon_{2,t-1}^{*2} \\
 &\quad + g_{12}^* h_{11,t-1} + 2g_{12}^* g_{22}^* h_{12,t-1} + g_{22}^* h_{22,t-1},
\end{align*}
$$
The likelihood ratio test statistic is used to test the volatility spillover effect between Taiwan and New York. The null hypothesis of no volatility spillovers from New York to Taiwan is

\[ H_0 : a_{21}^* = g_{21}^* = 0, \]

and the null hypothesis of no volatility spillovers from Taiwan to New York is

\[ H_0 : a_{12}^* = g_{12}^* = 0. \]
R Code for Multivariate GARCH: ccgarch