LECTURE NOTES ON GARCH MODELS

EDUARDO ROSSI
University of Pavia

March, 2004

Abstract

In these notes we present a survey of the theory of univariate and multivariate GARCH models. ARCH, GARCH, EGARCH and other possible nonlinear extensions are examined. Conditions for stationarity (weak and strong) are presented. Inference and testing is presented in the quasi-maximum likelihood framework. Multivariate parameterizations are examined in details.

Contents

Chapter 1
UNIVARIATE ARCH MODELS

1.1 Empirical regularities

GARCH models have been developed to account for empirical regularities in financial data. Many financial time series have a number of characteristics in common.

1. Asset prices are generally non stationary. Returns are usually stationary. Some financial time series are fractionally integrated.

2. Return series usually show no or little autocorrelation.

3. Serial independence between the squared values of the series is often rejected pointing towards the existence of non-linear relationships between subsequent observations.

4. Volatility of the return series appears to be clustered.

5. Normality has to be rejected in favor of some thick-tailed distribution.

6. Some series exhibit so-called *leverage effect*, that is changes in stock prices tend to be negatively correlated with changes in volatility. A firm with debt and equity outstanding typically becomes more highly leveraged when the value of the firm falls.This raises equity returns volatility if returns are constant. Black, however, argued that the response of stock volatility to the direction of returns is too large to be explained by leverage alone.

7. Volatilities of different securities very often move together.

1.2 Why do we need ARCH models?

Wold's decomposition theorem establishes that any covariance stationary $\{y_t\}$ may be written as the sum of a linearly deterministic component and a linearly stochastic with a square-summable, one-sided moving average representation. We can write,

$$y_t = d_t + u_t$$

d_t is linearly deterministic and u_t is a linearly regular covariance stationary stochastic process, given by

$$u_t = B(L)\,\varepsilon_t$$

$$B(L) = \sum_{i=0}^{\infty} b_i L^i \qquad \sum_{i=0}^{\infty} b_i^2 < \infty \qquad b_0 = 1$$

$$\begin{aligned} E[\varepsilon_t] &= 0 \\ E[\varepsilon_t \varepsilon_\tau] &= \begin{cases} \sigma_\varepsilon^2 < \infty, & if \quad t = \tau \\ 0, & otherwise \end{cases} \end{aligned}$$

The uncorrelated innovation sequence need not to be Gaussian and therefore need not be independent. Non-independent innovations are characteristic of non-linear time series in general and conditionally heteroskedastic time series in particular.

Now suppose that y_t is a linear covariance stationary process with i.i.d. innovations as opposed to merely white noise. The unconditional mean and variance are

$$E[y_t] = 0$$

$$E\left[y_t^2\right] = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} b_i^2$$

which are both invariant in time. The conditional mean is time varying and is given by

$$E[y_t\,|\Phi_{t-1}] = \sum_{i=1}^{\infty} b_i \varepsilon_{t-i}$$

where the information set is $\Phi_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$. This model is unable to capture the conditional variance dynamics. In fact, the conditional variance of y_t is constant at

$$E\left[(y_t - E[y_t\,|\Phi_{t-1}])^2\,|\Phi_{t-1}\right] = \sigma_\varepsilon^2.$$

This restriction manifests itself in the properties of the k-step-ahead conditional prediction error variance. The k-step-ahead conditional prediction is

$$E[y_{t+k}\,|\Phi_t] = \sum_{i=0}^{\infty} b_{k+i} \varepsilon_{t-i}$$

and the associated prediction error is

$$y_{t+k} - E\left[y_{t+k}\left|\Phi_t\right.\right] = \sum_{i=0}^{k-1} b_i \varepsilon_{t+k-i}$$

which has a conditional prediction error variance

$$E\left[\left(y_{t+k} - E\left[y_{t+k}\left|\Phi_t\right.\right]\right)^2\left|\Phi_t\right.\right] = \sigma_\varepsilon^2 \sum_{i=0}^{k-1} b_i^2$$

As $k \to \infty$ the conditional prediction error variance converges to the unconditional variance $\sigma_\varepsilon^2 \sum_{i=0}^{\infty} b_i^2$. For any k, the conditional prediction error variance depends only on k and not on Φ_t. In conclusion, the simple "i.i.d. innovations model" is unable to take into account the relevant information which is available at time t.

1.3 The ARCH(q) Model

Let $\{\varepsilon_t(\theta)\}$ denote a discrete time stochastic process with conditional mean and variance parametrized by a the finite dimensional vector $\theta \epsilon \Theta \subseteq \Re^m$, where θ_0 denotes the true value. We assume, for the moment, that $\varepsilon_t(\theta_0)$ is a scalar.

$E_{t-1}[\cdot]$ denotes the conditional expectation when the conditioning set is composed by the past values of the process along with other information available at time $t-1$ (denoted by Φ_{t-1}):

$$E_{t-1}[\cdot] \equiv E\left[\cdot\left|\Phi_{t-1}\right.\right]$$

analogously for the conditional variance:

$$Var_{t-1}[\cdot] \equiv Var\left[\cdot\left|\Phi_{t-1}\right.\right]$$

Definition 1 *(Bollerslev, Engle and Nelson [5]) The process $\{\varepsilon_t(\theta_0)\}$ follows an ARCH model if*

$$E_{t-1}[\varepsilon_t(\theta_0)] = 0 \qquad t = 1, 2, \ldots \tag{1.1}$$

and the conditional variance

$$\sigma_t^2(\theta_0) \equiv Var_{t-1}[\varepsilon_t(\theta_0)] = E_{t-1}\left[\varepsilon_t^2(\theta_0)\right] \qquad t = 1, 2, \ldots \tag{1.2}$$

depends non trivially on the σ-field generated by the past observations: $\{\varepsilon_{t-1}(\theta_0), \varepsilon_{t-2}(\theta_0), \ldots\}$.

Let $\{y_t(\theta_0)\}$ denote the stochastic process of interest with conditional mean

$$\mu_t(\theta_0) \equiv E_{t-1}(y_t) \qquad t = 1, 2, \ldots \tag{1.3}$$

By the time convention, both $\mu_t(\theta_0)$ and $\sigma_t^2(\theta_0)$ are measurable with respect to the time $t-1$ information set.* Define the $\{\varepsilon_t(\theta_0)\}$ process by

$$\varepsilon_t(\theta_0) \equiv y_t - \mu_t(\theta_0). \tag{1.4}$$

It follows from eq.(1.1) and (1.2), that the standardized process

$$z_t(\theta_0) \equiv \varepsilon_t(\theta_0)\,\sigma_t^2(\theta_0)^{-1/2} \qquad t = 1, 2, \ldots \tag{1.5}$$

will have conditional mean zero ($E_{t-1}[z_t(\theta_0)] = 0$) and a time invariant conditional variance of unity.

We can think of $\varepsilon_t(\theta_0)$ as generated by

$$\varepsilon_t(\theta_0) = z_t(\theta_0)\,\sigma_t^2(\theta_0)^{1/2}$$

where $\varepsilon_t^2(\theta_0)$ is unbiased estimator of $\sigma_t^2(\theta_0)$. Let's suppose $z_t(\theta_0) \sim NID(0,1)$ and independent of $\sigma_t^2(\theta_0)$

$$E_{t-1}\left[\varepsilon_t^2\right] = E_{t-1}\left[\sigma_t^2\right] E_{t-1}\left[z_t^2\right] = E_{t-1}\left[\sigma_t^2\right]$$

because $z_t^2\,|\Phi_{t-1} \sim \chi^2_{(1)}$. The median of a $\chi^2_{(1)}$ is 0.455 so $\Pr\left\{\varepsilon_t^2 < \frac{1}{2}\sigma_t^2\right\} > \frac{1}{2}$.

If the conditional distribution of z_t is time invariant with a finite fourth moment, the fourth moment of ε_t is

$$E\left[\varepsilon_t^4\right] = E\left[z_t^4\right] E\left[\sigma_t^4\right] \geq E\left[z_t^4\right] E\left[\sigma_t^2\right]^2 = E\left[z_t^4\right] E\left[\varepsilon_t^2\right]^2$$

$$E\left[\varepsilon_t^4\right] \geq E\left[z_t^4\right] E\left[\varepsilon_t^2\right]^2$$

by Jensen's inequality†. The equality holds true for a constant conditional variance only. If $z_t \sim NID(0,1)$, then $E\left[z_t^4\right] = 3$, the unconditional distribution for ε_t is therefore leptokurtic

$$\begin{aligned} E\left[\varepsilon_t^4\right] &\geq 3E\left[\varepsilon_t^2\right]^2 \\ E\left[\varepsilon_t^4\right]/E\left[\varepsilon_t^2\right]^2 &\geq 3 \end{aligned}$$

---

*Andersen distinguishes between deterministic, conditionally heteroskedastic, conditionally stochastic and contemporaneously stochastic volatility process. The volatiltiy process is deterministic if the information set (σ-field), which we denote with Φ, is identical to the σ-field of all random vectors in the system up to and including time $t = 0$, the process is conditionally heteroskedastic if Φ contains information available and observable at time $t-1$, the process is conditionally stochastic if Φ contains up to period $t-1$ whereas the volatility process is contemporaneously stochastic if the information set Φ contains the random vectors up to period t.

†Jensen's inequality:

$$E\left[g(x)\right] \leq g\left(E[x]\right)$$

if $g(\cdot)$ is concave

$$E\left[g(x)\right] \geq g\left(E[x]\right)$$

if $g(\cdot)$ is convex.

The kurtosis can be expressed as a function of the variability of the conditional variance. In fact, if $\varepsilon_t \left|\Phi_{t-1} \sim N\left(0, \sigma_t^2\right)\right.$

$$\begin{aligned} E_{t-1}\left[\varepsilon_t^4\right] &= 3E_{t-1}\left[\varepsilon_t^2\right] \\ E\left[\varepsilon_t^4\right] &= 3E\left[E_{t-1}\left(\varepsilon_t^2\right)^2\right] \geq 3\left\{E\left[E_{t-1}\left(\varepsilon_t^2\right)\right]\right\}^2 = 3\left[E\left(\varepsilon_t^2\right)\right] \end{aligned}$$

$$\begin{aligned} E\left[\varepsilon_t^4\right] - 3\left[E\left(\varepsilon_t^2\right)\right]^2 &= 3E\left\{E_{t-1}\left[\varepsilon_t^2\right]^2\right\} - 3\left\{E\left[E_{t-1}\left(\varepsilon_t^2\right)\right]\right\}^2 \\ E\left[\varepsilon_t^4\right] &= 3\left[E\left(\varepsilon_t^2\right)\right]^2 + 3E\left\{E_{t-1}\left[\varepsilon_t^2\right]^2\right\} - 3\left\{E\left[E_{t-1}\left(\varepsilon_t^2\right)\right]\right\}^2 \\ k &= \frac{E\left[\varepsilon_t^4\right]}{\left[E\left(\varepsilon_t^2\right)\right]^2} = 3 + 3\frac{E\left\{E_{t-1}\left[\varepsilon_t^2\right]^2\right\} - \left\{E\left[E_{t-1}\left(\varepsilon_t^2\right)\right]\right\}^2}{\left[E\left(\varepsilon_t^2\right)\right]^2} \\ &= 3 + 3\frac{Var\left\{E_{t-1}\left[\varepsilon_t^2\right]\right\}}{\left[E\left(\varepsilon_t^2\right)\right]^2} = 3 + 3\frac{Var\left\{\sigma_t^2\right\}}{\left[E\left(\varepsilon_t^2\right)\right]^2} \end{aligned}$$

Another important property of the ARCH process is that the process is conditionally serially uncorrelated. Given that

$$E_{t-1}\left[\varepsilon_t\right] = 0$$

we have that with the Law of Iterated Expectations:

$$E_{t-h}\left[\varepsilon_t\right] = E_{t-h}\left[E_{t-1}\left(\varepsilon_t\right)\right] = E_{t-h}\left[0\right] = 0.$$

This orthogonality property implies that the $\{\varepsilon_t\}$ process is conditionally uncorrelated:

$$\begin{aligned} Cov_{t-h}\left[\varepsilon_t, \varepsilon_{t+k}\right] &= E_{t-h}\left[\varepsilon_t\varepsilon_{t+k}\right] - E_{t-h}\left[\varepsilon_t\right]E_{t-h}\left[\varepsilon_{t+k}\right] = \\ &= E_{t-h}\left[\varepsilon_t\varepsilon_{t+k}\right] = E_{t-h}\left[E_{t+k-1}\left(\varepsilon_t\varepsilon_{t+k}\right)\right] = \\ &= E\left[\varepsilon_t E_{t+k-1}\left[\varepsilon_{t+k}\right]\right] = 0 \end{aligned}$$

The ARCH model has showed to be particularly useful in modeling the temporal dependencies in asset returns.

The *ARCH* model introduced by Engle (Engle ([9])) is a linear function of past squared disturbances:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \tag{1.6}$$

In this model to assure a positive conditional variance the parameters have to satisfy the following constraints: $\omega > 0$ e $\alpha_1 \geq 0, \alpha_2 \geq 0, \ldots, \alpha_q \geq 0$. Defining

$$\sigma_t^2 \equiv \varepsilon_t^2 - v_t$$

where $E_{t-1}(v_t) = 0$ we can write (1.6) as an $AR(q)$ in ε_t^2:

$$\varepsilon_t^2 = \omega + \alpha(L)\varepsilon_t^2 + v_t$$

where $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_q L^q$ (where L is the lag operator, i.e. $x_{t-1} = Lx_t$). The process is weakly stationary if and only if $\sum_{i=1}^{q} \alpha_i < 1$; in this case the unconditional variance is given by

$$E\left(\varepsilon_t^2\right) = \omega / \left(1 - \alpha_1 - \ldots - \alpha_q\right). \tag{1.7}$$

The process is characterised by leptokurtosis in excess with respect to the normal distribution. In the case, for example, of *ARCH(1)* with $\varepsilon_t \left| \Phi_{t-1} \sim N\left(0, \sigma_t^2\right)\right.$, the kurtosis is equal to:

$$E\left(\varepsilon_t^4\right) / E\left(\varepsilon_t^2\right)^2 = 3\left(1 - \alpha_1^2\right) / \left(1 - 3\alpha_1^2\right) \tag{1.8}$$

with $3\alpha_1^2 < 1$, when $3\alpha_1^2 = 1$ we have

$$E\left(\varepsilon_t^4\right) / E\left(\varepsilon_t^2\right)^2 = \infty.$$

In both cases we obtain a kurtosis coefficient greater than 3, characteristic of the normal distribution. The result is readily obtained:

$$E\left(\varepsilon_t^4\right) = 3E\left(\sigma_t^4\right)$$

$$E\left(\varepsilon_t^4\right) = 3\left[\omega^2 + \alpha_1^2 E\left(\varepsilon_{t-1}^4\right) + 2\omega\alpha_1 E\left(\varepsilon_{t-1}^2\right)\right]$$

$$\begin{aligned} E\left(\varepsilon_t^4\right) &= \frac{3\left[\omega^2 + 2\omega\alpha_1 E\left(\varepsilon_{t-1}^2\right)\right]}{\left(1 - 3\alpha_1^2\right)} \\ &= \frac{3\left[\omega^2 + 2\omega\alpha_1\sigma^2\right]}{\left(1 - 3\alpha_1^2\right)} \end{aligned}$$

substituting $\sigma^2 = \omega / \left(1 - \alpha_1\right)$:

$$E\left(\varepsilon_t^4\right) = \frac{3\left[\omega^2\left(1 - \alpha_1\right) + 2\omega^2\alpha_1\right]}{\left(1 - 3\alpha_1^2\right)\left(1 - \alpha_1\right)} = \frac{3\omega^2\left(1 + \alpha_1\right)}{\left(1 - 3\alpha_1^2\right)\left(1 - \alpha_1\right)}$$

finally

$$E\left(\varepsilon_t^4\right) / E\left(\varepsilon_t^2\right)^2 = \frac{3\omega^2\left(1 + \alpha_1\right)\left(1 - \alpha_1\right)^2}{\left(1 - 3\alpha_1^2\right)\left(1 - \alpha_1\right)\omega^2} = \frac{3\left(1 - \alpha_1^2\right)}{1 - 3\alpha_1^2}. \tag{1.9}$$

1.3.1 The ARCH Regression Model

We have an ARCH regression model when the disturbances in a linear regression model follow an ARCH process:

$$y_t = x_t' b + \varepsilon_t$$

$$E_{t-1}(\varepsilon_t) = 0$$

$$E_{t-1}\left(\varepsilon_t^2\right) \equiv \sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2$$

$$\varepsilon_t \left| \Psi_{t-1} \sim N\left(0, \sigma_t^2\right)\right.$$

where x_t may include lagged dependent and exogenous variables.

1.3.2 ARCH as a nonlinear model

The essential characteristic of the ARCH model is $Cov\left(\varepsilon_t^2, \varepsilon_{t-1}^2\right) \neq 0$, although $Cov\left(\varepsilon_t, \varepsilon_{t-1}\right) = 0$ for $j \neq 0$. We examine the relation of the ARCH model with the bilinear model. A time series $\{\varepsilon_t\}$ is said to follow a bilinear model if it satisfies

$$\varepsilon_t = \sum_{i=1}^{p} \phi_i \varepsilon_{t-i} + \sum_{j=1}^{r}\sum_{k=1}^{s} b_{jk}\varepsilon_{t-j}u_{t-k} + u_t$$

where u_t is a sequence of $i.i.d.(0, \sigma_u^2)$ variables. The first two conditional moments are

$$E_{t-1}(\varepsilon_t) = \sum_{i=1}^{p} \phi_i \varepsilon_{t-i} + \sum_{j=1}^{r}\sum_{k=1}^{s} b_{jk}\varepsilon_{t-j}u_{t-k} + u_t$$

$$Var_{t-1}(\varepsilon_t) = \sigma_u^2.$$

In contrast with the ARCH model in which the conditional variance is time varying, in the bilinear model the conditional variance is constant. Their unconditional moments, however, might be similar. The bilinear model

$$\varepsilon_t = b_{21}\varepsilon_{t-2}u_{t-1} + u_t$$

$$E(\varepsilon_t) = 0$$

$$Cov\left(\varepsilon_t^2, \varepsilon_{t-1}^2\right) = b_{21}^2\sigma_u^2$$

as this process is autocorrelated in squares, it will exhibit temporal clustering of large and small deviations like an ARCH process.

1.4 The GARCH(p,q) Model

In order to model in a parsimonious way the conditional heteroskedasticity, Bollerslev [2] proposed the Generalised *ARCH* model, i.e *GARCH(p,q)*:

$$\sigma_t^2 = \omega + \alpha(L)\,\varepsilon_t^2 + \beta(L)\,\sigma_t^2. \tag{1.10}$$

where $\alpha(L) = \alpha_1 L + \ldots + \alpha_q L^q$, $\beta(L) = \beta_1 L + \ldots + \beta_p L^p$. The *GARCH(1,1)* is the most popular model in the empirical literature[‡]:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{1.11}$$

To ensure that the conditional variance is well defined in a $GARCH(p,q)$ model all the coefficients in the corresponding linear $ARCH(\infty)$ should be positive. Rewriting the $GARCH(p,q)$ model as an $ARCH(\infty)$:

$$\begin{aligned}
\sigma_t^2 &= \left(1 - \sum_{i=1}^{p} \beta_i L_i\right)^{-1} \left[\omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2\right] \\
&= \omega^* + \sum_{k=0}^{\infty} \phi_k \varepsilon_{t-k-1}^2
\end{aligned} \tag{1.12}$$

$\sigma_t^2 \geq 0$ if $\omega^* \geq 0$ and all $\phi_k \geq 0$. The non-negativity of ω^* and ϕ_k is also a necessary condition for the non negativity of σ_t^2. In order to make ω^* e $\{\phi_k\}_{k=0}^{\infty}$ well defined, assume that :

i. the roots of the polynomial $\beta(x) = 1$ lie outside the unit circle.and that $\omega \geq 0$, this is a condition for ω^* to be finite and positive.

ii. $\alpha(x)$ e $1 - \beta(x)$ have no common roots.

These conditions are establishing nor that $\sigma_t^2 \leq \infty$ neither that $\{\sigma_t^2\}_{t=-\infty}^{\infty}$ is strictly stationary. For the simple *GARCH(1,1)* almost sure positivity of σ_t^2 requires, with the conditions (i) and (ii), that (Nelson and Cao [25]),

$$\begin{aligned}
\omega &\geq 0 \\
\beta_1 &\geq 0 \\
\alpha_1 &\geq 0
\end{aligned} \tag{1.13}$$

[‡]The GARCH model belongs to the class of deterministic conditional heteroskedasticity models in which the conditional variance is a function of variables that are in the information set available at time t.

For the *GARCH(1,q)* and *GARCH(2,q)* models these constraints can be relaxed, e.g. in the *GARCH(1,2)* model the necessary and sufficient conditions become:

$$\begin{aligned} \omega &\geq 0 \\ 0 &\leq \beta_1 < 1 \\ \beta_1\alpha_1 + \alpha_2 &\geq 0 \\ \alpha_1 &\geq 0 \end{aligned} \tag{1.14}$$

For the *GARCH(2,1)* model the conditions are:

$$\begin{aligned} \omega &\geq 0 \\ \alpha_1 &\geq 0 \\ \beta_1 &\geq 0 \\ \beta_1 + \beta_2 &< 1 \\ \beta_1^2 + 4\beta_2 &\geq 0 \end{aligned} \tag{1.15}$$

These constraints are less stringent than those proposed by Bollerslev [2]:

$$\begin{aligned} \omega &\geq 0 \\ \beta_i &\geq 0 \quad i = 1, \ldots, p \\ \alpha_j &\geq 0 \quad j = 1, \ldots, q \end{aligned} \tag{1.16}$$

These results cannot be adopted in the multivariate case, where, as we will see below, the requirement of positivity for $\{\sigma_t^2\}$ means the positive definiteness for the conditional variance-covariance matrix.

From the point of view of the maximum likelihood estimation of a *GARCH(p,q)* model we need to recursively calculate $\{\sigma_t^2\}_{t=0}^{\infty}$ starting from 0 applying the (1.10), assuming arbitrary values for the pre-sample period $\left\{\sigma_{-1}^2, \ldots, \sigma_{-p}^2, \varepsilon_{-1}^2, \ldots, \varepsilon_{-q}^2\right\}$. The conditions (1.16) guarantee that $\{\sigma_t^2\}_{t=0}^{\infty}$ is not negative given arbitrary non negative values for $\left\{\sigma_{-1}^2, \ldots, \sigma_{-p}^2, \varepsilon_{-1}^2, \ldots, \varepsilon_{-q}^2\right\}$. On the contrary, the conditions which guarantee that $\omega^* \geq 0$ and $\phi_k \geq 0$ (1.14) for the *GARCH(1,2)* model and the conditions (1.15) for the *GARCH(2,1)* model) do not. This problem can be solved choosing the starting values that mantain non negative $\{\sigma_t^2\}_{t=0}^{\infty}$ with probability 1, given non negative ω^* and $\{\phi_k\}_{k=0}^{\infty}$. Nelson and Cao suggest to arbitrarly pick a $\varepsilon^2 \geq 0$ and set $\varepsilon_t^2 = \varepsilon^2$ for t from -1 to ∞. and $\sigma_t^2 = \sigma^2$ for $1 - p \leq t \leq 0$ where

$$\begin{aligned} \sigma^2 &= \left(1 - \sum_{i=1}^{p} \beta_i\right)^{-1} \left[\omega + \varepsilon^2 \sum_{j=1}^{q} \alpha_j\right] \\ &= \omega^* + \varepsilon^2 \sum_{k=0}^{\infty} \phi_k \end{aligned}$$

So doing we have a sequence $\{\sigma_t^2\} \geq 0$ for all $t \geq 0$ with probability 1, as

$$\sigma_t^2 = \omega^* + \sum_{k=0}^{t-1} \phi_k \varepsilon_{t-k-1}^2 + \sum_{k=t}^{\infty} \phi_k \varepsilon^2$$

Supposing that $\sum_{i=1}^{p} \beta_i + \sum_{j=1}^{q} \alpha_j < 1$ we can set σ^2 e ε^2 equal to their common unconditional mean:

$$\sigma^2 \equiv \varepsilon^2 \equiv \omega / \left(1 - \sum_{i=1}^{p} \beta_i - \sum_{j=1}^{q} \alpha_j\right).$$

1.4.1 The Yule-Walker equations for the squared process

In the GARCH(p,q) model the process $\{\varepsilon_t^2\}$ has an ARMA(m,p) representations, where $m = \max(p, q)$

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{m} \left(\alpha_j + \beta_j\right) \varepsilon_{t-j}^2 + \left(\upsilon_t - \sum_{i=1}^{p} \beta_i \upsilon_{t-i}\right)$$

where $E_{t-1}[\upsilon_t] = 0$, $\upsilon_t \in [-\sigma_t^2, \infty[$ we can apply the classical results of ARMA model. We can study the autocovariance function, that is:

$$\gamma^2(k) = cov\left(\varepsilon_t^2, \varepsilon_{t-k}^2\right)$$

$$\gamma^2(k) = cov\left[\omega + \sum_{j=1}^{m} \left(\alpha_j + \beta_j\right) \varepsilon_{t-j}^2 + \left(\upsilon_t - \sum_{i=1}^{p} \beta_i \upsilon_{t-i}\right), \epsilon_{t-k}^2\right]$$

$$\gamma^2(k) = \left[\sum_{j=1}^{m} \left(\alpha_j + \beta_j\right) cov\left(\varepsilon_{t-j}^2, \varepsilon_{t-k}^2\right)\right] + cov\left[\upsilon_t - \sum_{i=1}^{p} \beta_i \upsilon_{t-i}, \varepsilon_{t-k}^2\right] \tag{1.17}$$

When k is big enough, the last term on the right of expression (1.17) is null. The sequence of autocovariances satisfy a linear difference equation of order $Max(p, q)$, for $k \geq p + 1$

$$\gamma^2(k) = \left[\sum_{j=1}^{m} \left(\alpha_j + \beta_j\right) \gamma^2(k - j)\right]$$

This system can be used to identify the lag order m and p, that is the p and q order if $q \geq p$, the order p if $q < p$.

1.4.2 The GARCH Regression Model

Let $w_t = \left(1, \varepsilon_{t-1}^2, \cdots, \varepsilon_{t-q}^2, \sigma_{t-1}^2, \cdots, \sigma_{t-p}^2\right)'$, $\gamma = \left(\omega, \alpha_1, \cdots, \alpha_q, \beta_1, \cdots, \beta_p\right)$ and $\theta \in \Theta$, where $\theta = (b', \gamma')$ and Θ is a compact subspace of a Euclidean space such that ε_t possesses finite second moments. We may write the GARCH regression model as:

$$\varepsilon_t = y_t - x_t' b$$

$$\varepsilon_t \left| \Psi_{t-1} \sim N\left(0, \sigma_t^2\right)\right.$$

$$\sigma_t^2 = w_t' \gamma$$

1.4.3 Stationarity

The process $\{\varepsilon_t\}$ which follows a GARCH(p,q) model is a martingale difference sequence. In order to study second-order stationarity it's sufficient to consider that:

$$Var\left[\varepsilon_t\right] = Var\left[E_{t-1}\left(\varepsilon_t\right)\right] + E\left[Var_{t-1}\left(\varepsilon_t\right)\right] = E\left[\sigma_t^2\right]$$

and show that is asymptotically constant in time (it does not depend upon time).

Proposition 2 *A process $\{\varepsilon_t\}$ which satisfies a GARCH(p,q) model with positive coefficient $\omega \geq 0$, $\alpha_i \geq 0$ $i = 1, \ldots, q$, $\beta_i \geq 0$ $i = 1, \ldots, p$ is covariance stationary if and only if:*

$$\alpha\left(1\right) + \beta(1) < 1$$

This is a sufficient but non necessary conditions for strict stationarity. Because ARCH processes are thick tailed, the conditions for covariance stationarity are often more stringent than the conditions for strict stationarity.

Example 3 *A GARCH(1,1) model can be written as*

$$\sigma_t^2 = \omega\left[1 + \sum_{k=1}^{\infty}\prod_{i=1}^{k}\left(\beta_1 + \alpha_1 z_{t-i}^2\right)\right]$$

In fact,

$$\sigma_{t+1}^2 = \omega + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2 = \omega + \sigma_t^2\left(\alpha_1 z_t^2 + \beta_1\right)$$

$$\sigma_t^2 = \omega\left[1 + \sum_{k=1}^{\infty}\prod_{i=1}^{k}\left(\alpha_1 z_{t-i}^2 + \beta_1\right)\right]$$

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 \left(\alpha_1 z_{t-1}^2 + \beta_1\right)$$

$$\sigma_{t-1}^2 = \omega + \sigma_{t-2}^2 \left(\alpha_1 z_{t-2}^2 + \beta_1\right)$$

$$\begin{aligned}
\sigma_t^2 &= \omega + \left[\omega + \sigma_{t-2}^2 \left(\alpha_1 z_{t-2}^2 + \beta_1\right)\right] \left(\alpha_1 z_{t-1}^2 + \beta_1\right) \\
&= \omega + \omega \left(\alpha_1 z_{t-1}^2 + \beta_1\right) + \sigma_{t-2}^2 \left(\alpha_1 z_{t-2}^2 + \beta_1\right) \left(\alpha_1 z_{t-1}^2 + \beta_1\right) \\
&= \omega + \omega \left(\alpha_1 z_{t-1}^2 + \beta_1\right) + \omega \left(\alpha_1 z_{t-1}^2 + \beta_1\right) \left(\alpha_1 z_{t-2}^2 + \beta_1\right) \\
&\quad + \sigma_{t-3}^2 \left(\alpha_1 z_{t-3}^2 + \beta_1\right) \left(\alpha_1 z_{t-2}^2 + \beta_1\right) \left(\alpha_1 z_{t-1}^2 + \beta_1\right)
\end{aligned}$$

Nelson [23] shows that when $\omega > 0$, $\sigma_t^2 < \infty$ a.s.and $\{\varepsilon_t, \sigma_t^2\}$ is strictly stationary if and only if $E\left[\ln\left(\beta_1 + \alpha_1 z_t^2\right)\right] < 0$

$$E\left[\ln\left(\beta_1 + \alpha_1 z_t^2\right)\right] \leq \ln\left[E\left(\beta_1 + \alpha_1 z_t^2\right)\right] = \ln\left(\alpha_1 + \beta_1\right)$$

when $\alpha_1 + \beta_1 = 1$ the model is strictly stationary. $E\left[\ln\left(\beta_1 + \alpha_1 z_t^2\right)\right] < 0$ is a weaker requirement than $\alpha_1 + \beta_1 < 1$.

Example 4 *ARCH(1), with $\alpha_1 = 1$, $\beta_1 = 0$, $z_t \sim N(0,1)$*

$$E\left[\ln\left(z_t^2\right)\right] \leq \ln\left[E\left(z_t^2\right)\right] = \ln(1)$$

It's strictly but not covariance stationary. The ARCH(q) is covariance stationary if and only if the sum of the positive parameters is less than one.

1.4.4 Forecasting volatility

A GARCH(p,q) can be represented as an ARMA process, given that $\varepsilon_t^2 = \sigma_t^2 + \upsilon_t$, where $E_{t-1}\left[\upsilon_t\right] = 0$, $\upsilon_t \in [-\sigma_t^2, \infty[$:

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{\max(p,q)} \left(\alpha_j + \beta_j\right) \varepsilon_{t-j}^2 + \left(\upsilon_t - \sum_{i=1}^{p} \beta_i \upsilon_{t-i}\right)$$

$\varepsilon_t^2 \sim$ARMA(m,p) with $m = \max(p, q)$. Forecasting with a GARCH(p,q) (Engle and Bollerslev [11]):

$$\sigma_{t+k}^2 = \omega + \sum_{i=1}^{n} \left[\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2\right] + \sum_{i=k}^{m} \left[\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2\right]$$

where $n = \min\{m, k-1\}$ and by definition summation from 1 to 0 and from $k > m$ to m both are equal to zero. Thus

$$E_t\left[\sigma_{t+k}^2\right] = \omega + \sum_{i=1}^{n} \left[\left(\alpha_i + \beta_i\right) E_t\left(\sigma_{t+k-i}^2\right)\right] + \sum_{i=k}^{m} \left[\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2\right].$$

In particular for a GARCH(1,1) and $k > 2$:

$$\begin{aligned} E_t\left[\sigma_{t+k}^2\right] &= \sum_{i=0}^{k-2} \left(\alpha_1+\beta_1\right)^i \omega + \left(\alpha_1+\beta_1\right)^{k-1} \sigma_{t+1}^2 \\ &= \omega \frac{\left[1-\left(\alpha_1+\beta_1\right)^{k-1}\right]}{\left[1-\left(\alpha_1+\beta_1\right)\right]} + \left(\alpha_1+\beta_1\right)^{k-1} \sigma_{t+1}^2 \\ &= \sigma^2 \left[1-\left(\alpha_1+\beta_1\right)^{k-1}\right] + \left(\alpha_1+\beta_1\right)^{k-1} \sigma_{t+1}^2 \\ &= \sigma^2 + \left(\alpha_1+\beta_1\right)^{k-1} \left[\sigma_{t+1}^2 - \sigma^2\right] \end{aligned}$$

When the process is covariance stationary, it follows that $E_t\left[\sigma_{t+k}^2\right]$ converges to σ^2 as $k \to \infty$.

1.4.5 The IGARCH(p,q) model

Definition 5 *The GARCH(p,q) process characterised by the first two conditional moments:*

$$E_{t-1}\left[\varepsilon_t\right] = 0$$

$$\sigma_t^2 \equiv E_{t-1}\left[\varepsilon_t^2\right] = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

where $\omega \geq 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$ for all i and the polynomial

$$1 - \alpha\left(x\right) - \beta(x) = 0$$

has $d > 0$ unit root(s) and $\max\left\{p, q\right\} - d$ root(s) outside the unit circle is said to be:

i) Integrated in variance of order d if $\omega = 0$

ii) Integrated in variance of order d with trend if $\omega > 0$.

The Integrated GARCH(p,q) models, both with or without trend, are therefore part of a wider class of models with a property called "persistent variance" in which the current information remains important for the forecasts of the conditional variances for all horizon.

So we have the Integrated GARCH(p,q) model when (necessary condition)

$$\alpha\left(1\right) + \beta(1) = 1$$

To illustrate consider the IGARCH(1,1) which is characterised by

$$\alpha_1 + \beta_1 = 1$$

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \left(1-\alpha_1\right) \sigma_{t-1}^2 \\ \sigma_t^2 &= \omega + \sigma_{t-1}^2 + \alpha_1 \left(\varepsilon_{t-1}^2 - \sigma_{t-1}^2\right) \qquad 0 < \alpha_1 \leq 1 \end{aligned}$$

For this particular model the conditional variance k steps in the future is:

$$E_t\left[\sigma_{t+k}^2\right] = \left(k-1\right) \omega + \sigma_{t+1}^2$$

1.4.6 Persistence

In many studies of the time series behavior of asset volatility the question has been how long shocks to conditional variance persist. If volatility shocks persist indefinitely, they may move the whole term structure of risk premia. There are many notions of convergence in the probability theory (almost sure, in probability, in L^p), so whether a shock is transitory or persistent may depend on the definition of convergence. In linear models it typically makes no difference which of the standard definitions we use, since the definitions usually agree. In GARCH models the situation is more complicated. In the IGARCH(1,1):

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

where $\alpha_1 + \beta_1 = 1$. Given that $\varepsilon_t^2 = z_t^2 \sigma_t^2$, we can rewrite the IGARCH(1,1) process as

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 \left[(1-\alpha_1) + \alpha_1 z_{t-1}^2 \right] \qquad 0 < \alpha_1 \leq 1.$$

When $\omega = 0$, σ_t^2 is a martingale. Based on the nature of persistence in linear models, it seems that IGARCH(1,1) with $\omega > 0$ and $\omega = 0$ are analogous to random walks with and without drift, respectively, and are therefore natural models of "persistent" shocks. This turns out to be misleading, however: in IGARCH(1,1) with $\omega = 0$, σ_t^2 collapses to zero almost surely, and in IGARCH(1,1) with $\omega > 0$, σ_t^2 is strictly stationary and ergodic and therefore does not behave like a random walk, since random walks diverge almost surely.

Two notions of persistence.

1. Suppose σ_t^2 is strictly stationary and ergodic. Let $F(\sigma_t^2)$ be the unconditional cdf for σ_t^2, and $F_s(\sigma_t^2)$ the conditional cdf for σ_t^2, given information at time $s < t$. For any s $F(\sigma_t^2) - F_s(\sigma_t^2) \to 0$ at all continuity points as $t \to \infty$. There is no persistence when $\{\sigma_t^2\}$ is stationary and ergodic.

2. Persistence is defined in terms of forecast moments. For some $\eta > 0$, the shocks to σ_t^2 fail to persist if and only if for every s, $E_s\left(\sigma_t^{2\eta}\right)$ converges, as $t \to \infty$, to a finite limit independent of time s information set.

Whether or not shocks to $\{\sigma_t^2\}$ "persist" depends very much on which definition is adopted. The conditional moment may diverge to infinity for some η, but converge to a well-behaved limit independent of initial conditions for other η, even when the $\{\sigma_t^2\}$ is stationary and ergodic.

Example 6 *GARCH(1,1)*

$$\sigma_{t+1}^2 = \omega + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2 = \omega + \sigma_t^2 \left(\alpha_1 z_t^2 + \beta_1\right)$$

$$E_{t-3}\left(\sigma_t^2\right) = \omega\left[\sum_{k=0}^{t-(t-3)-1}\left(\alpha_1+\beta_1\right)^k\right] + \sigma_{t-3}^2\left(\alpha_1+\beta_1\right)\left(\alpha_1+\beta_1\right)\left(\alpha_1+\beta_1\right)$$

The volatility forecast for time t, conditioning on information set at time s:

$$E_s\left(\sigma_t^2\right) = \omega\left[\sum_{k=0}^{t-s-1}\left(\alpha_1+\beta_1\right)^k\right] + \sigma_{t-s}^2\left(\alpha_1+\beta_1\right)^{t-s} E_s\left(\sigma_t^2\right)$$

converges to the unconditional variance of $\omega/\left(1-\alpha_1-\beta_1\right)$ as $t \to \infty$ if and only if $\alpha_1+\beta_1 < 1$. In the IGARCH(1,1) model with $\dot{\omega} > 0$ and $\alpha_1+\beta_1 = 1$ $E_s\left(\sigma_t^2\right) \to \infty$ a.s. as $t \to \infty$. Nevertheless, IGARCH models are strictly stationary and ergodic.

1.4.7 The Component Model

A permanent and transitory component model of stock returns volatility (Engle and Lee, 1993).

The finding of a unit root in the volatility process indicates that there is a stochastic trend as well as a transitory component in stock return volatility. The decomposition of the conditional variance of asset returns in a permanent and transitory component is a way to investigate the long-run and the short-run movement of volatility in the stock market.

The GARCH(1,1) model can also be written as

$$\begin{aligned}\sigma_t^2 &= \left(1-\alpha_1-\beta_1\right)\sigma^2 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \\ &= \sigma^2 + \alpha_1\left(\varepsilon_{t-1}^2-\sigma^2\right) + \beta_1\left(\sigma_{t-1}^2-\sigma^2\right)\end{aligned}$$

The last two terms have expected value zero. This model is extended to allow the possibility that volatility is not constant in the long run. Let q_t be the *permanent component* of the conditional variance, the *component model* for the conditional variance is defined as

$$\begin{aligned}\sigma_t^2 &= q_t + \alpha_1\left(\varepsilon_{t-1}^2-q_{t-1}\right) + \beta_1\left(\sigma_{t-1}^2-q_{t-1}\right) \\ &= q_t - \left(\alpha_1+\beta_1\right)q_{t-1} + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2\end{aligned} \tag{1.18}$$

$$\left(1-\beta_1 L\right)\sigma_t^2 = \left[1-\left(\alpha_1+\beta_1\right)L\right]q_t + \alpha_1\varepsilon_{t-1}^2$$

$$q_t = \omega + q_{t-1} + \phi\left(\varepsilon_{t-1}^2-\sigma_{t-1}^2\right)$$

The constant volatility σ^2 has been replaced by the time-varying trend, q_t, and its past value. The forecasting error, $\varepsilon_{t-1}^2 - \sigma_{t-1}^2$, serves as a driving force for the time-dependent movement of the trend. The difference between the conditional variance and its trend, $\sigma_{t-1}^2 - q_{t-1}$, is the *transitory component* of the conditional variance.

The multistep forecast of the trend is just the current trend plus a constant drift:

$$q_{t+k} = \omega + q_{t+k-1} + \phi\left(\varepsilon_{t+k-1}^2 - \sigma_{t+k-1}^2\right)$$

$$E_{t-1}\left[q_{t+k}\right] = \omega + E_{t-1}\left[q_{t+k-1}\right] + \phi E_{t-1}\left[\varepsilon_{t+k-1}^2 - \sigma_{t+k-1}^2\right]$$

but $E_{t-1}\left(\varepsilon_{t+k-1}^2\right) = E_{t-1}\left(\sigma_{t+k-1}^2\right)$ such that $E_{t-1}\left[\varepsilon_{t+k-1}^2 - \sigma_{t+k-1}^2\right] = 0$.

$$\begin{aligned} E_{t-1}\left[q_{t+k}\right] &= \omega + \omega + E_{t-1}\left[q_{t+k-2}\right] + \phi E_{t-1}\left[\varepsilon_{t+k-2}^2 - \sigma_{t+k-2}^2\right] \\ &= \ldots \\ &= k\omega + q_t \end{aligned} \tag{1.19}$$

From (1.18)

$$\sigma_{t+1}^2 - q_{t+1} = \alpha_1\left(\varepsilon_t^2 - q_t\right) + \beta_1\left(\sigma_t^2 - q_t\right)$$

$$\begin{aligned} E_{t-1}\left(\sigma_{t+1}^2\right) - E_{t-1}\left(q_{t+1}\right) &= \alpha_1 E_{t-1}\left(\varepsilon_t^2 - q_t\right) + \beta_1 E_{t-1}\left(\sigma_t^2 - q_t\right) \\ &= \left(\alpha_1 + \beta_1\right)\left(\sigma_t^2 - q_t\right) \end{aligned}$$

$$\sigma_{t+2}^2 - q_{t+2} = \alpha_1\left(\varepsilon_{t+1}^2 - q_{t+1}\right) + \beta_1\left(\sigma_{t+1}^2 - q_{t+1}\right)$$

$$\sigma_{t+3}^2 - q_{t+3} = \alpha_1\left(\varepsilon_{t+2}^2 - q_{t+2}\right) + \beta_1\left(\sigma_{t+2}^2 - q_{t+2}\right)$$

$$\begin{aligned} E_t\left(\sigma_{t+3}^2 - q_{t+3}\right) &= \alpha_1 E_t\left(\varepsilon_{t+2}^2 - q_{t+2}\right) + \beta_1 E_t\left(\sigma_{t+2}^2 - q_{t+2}\right) \\ &= \alpha_1 E_t\left(\varepsilon_{t+2}^2\right) - \alpha_1 E_t\left(q_{t+2}\right) + \beta_1 E_t\left(\sigma_{t+2}^2\right) - \beta_1 E_t\left(q_{t+2}\right) \\ &= \left(\alpha_1 + \beta_1\right)\left[E_t\left(\sigma_{t+2}^2\right) - E_t\left(q_{t+2}\right)\right] \\ &= \left(\alpha_1 + \beta_1\right) E_t\left[\alpha_1\left(\varepsilon_{t+1}^2 - q_{t+1}\right) + \beta_1\left(\sigma_{t+1}^2 - q_{t+1}\right)\right] \end{aligned}$$

$$\begin{aligned} E_{t-1}\left(\sigma_{t+3}^2 - q_{t+3}\right) &= \left(\alpha_1 + \beta_1\right) E_{t-1}\left[\alpha_1\left(\varepsilon_{t+1}^2 - q_{t+1}\right) + \beta_1\left(\sigma_{t+1}^2 - q_{t+1}\right)\right] \\ &= \left(\alpha_1 + \beta_1\right)\left[\left(\alpha_1 + \beta_1\right) E_{t-1}\left(\sigma_{t+1}^2\right) - \left(\alpha_1 + \beta_1\right) E_{t-1}\left(q_{t+1}\right)\right] \\ &= \left(\alpha_1 + \beta_1\right)\left[\left(\alpha_1 + \beta_1\right)\left(E_{t-1}\left(\sigma_{t+1}^2\right) - E_{t-1}\left(q_{t+1}\right)\right)\right] \\ &= \left(\alpha_1 + \beta_1\right)^3\left(\sigma_t^2 - q_t\right) \end{aligned}$$

$$\begin{aligned} E_{t-1}\left(\sigma_{t+k}^2\right) - E_{t-1}\left(q_{t+k}\right) &= \left(\alpha_1+\beta_1\right)\left(E_{t-1}\left(\sigma_{t+k-1}^2\right) - E_{t-1}\left(q_{t+k-1}\right)\right) \\ &= \left(\alpha_1+\beta_1\right)^k\left(\sigma_t^2 - q_t\right) \end{aligned}$$

The forecast $E_{t-1}\left(\sigma_{t+k}^2\right) - E_{t-1}\left(q_{t+k}\right)$ will eventually converge to zero as the forecasting horizon extends into the remote future

$$E_{t-1}\left(\sigma_{t+k}^2\right) - E_{t-1}\left(q_{t+k}\right) = 0 \text{ as } k \to \infty \tag{1.20}$$

Therefore there will be no difference between the conditional variance and the trend in the long run. This is the motivation for q_t being called the permanent component of the conditional variance. Combining (1.20) and (1.19), the long run forecast of the conditional variance is just the current expectation of the trend plus a constant drift,

$$E_{t-1}\left(\sigma_{t+k}^2\right) = k\omega + q_t \text{ as } k \to \infty.$$

The component model can be extended to include non-unit-root process. The general component model becomes

$$\sigma_t^2 = q_t + \alpha_1\left(\varepsilon_{t-1}^2 - q_{t-1}\right) + \beta_1\left(\sigma_{t-1}^2 - q_{t-1}\right) \tag{1.21}$$

$$q_t = \omega + \rho q_{t-1} + \phi\left(\varepsilon_{t-1}^2 - \sigma_{t-1}^2\right) \tag{1.22}$$

q_t stil represents the component of the conditional variance with the longer memory, as long as $\rho > (\alpha_1 + \beta_1)$. The multistep forecast of the conditional variance and the trend are

$$E_{t-1}\left(\sigma_{t+k}^2\right) - E_{t-1}\left(q_{t+k}\right) = \left(\alpha_1+\beta_1\right)^k\left(\sigma_t^2 - q_t\right) \tag{1.23}$$

$$q_{t+k} = \omega + \rho q_{t+k-1} + \phi\left(\varepsilon_{t+k-1}^2 - \sigma_{t+k-1}^2\right)$$

$$\begin{aligned} E_{t-1}\left[q_{t+k}\right] &= \omega + \rho E_{t-1}\left[q_{t+k-1}\right] + \phi E_{t-1}\left[\varepsilon_{t+k-1}^2 - \sigma_{t+k-1}^2\right] \\ &= \omega + \rho\left[\omega + \rho E_{t-1}\left[q_{t+k-2}\right]\right] \\ &= \ldots \\ &= \left(1 + \rho + \ldots + \rho^{k-1}\right)\omega + \rho^k q_t \end{aligned}$$

$$E_{t-1}\left[q_{t+k}\right] = \frac{\left(1-\rho^k\right)}{(1-\rho)}\omega + \rho^k q_t \tag{1.24}$$

for $\rho < 1$ and $(\alpha_1 + \beta_1) < 1$. If $\rho > (\alpha_1 + \beta_1)$, the transitory component in (1.23) decays faster than the trend in (1.24) so that the trend will dominate the forecast of the conditional variance as the forecasting horizon extends. The conditional variance will eventually converge to a constant since the trend itself is stationary,

$$E_{t-1}\left(\sigma_{t+k}^2\right) = E_{t-1}\left(q_{t+k}\right) = \omega / \left(1-\rho\right) \text{ as } k \to \infty.$$

By rewriting (1.21) as

$$\sigma_t^2 = \left(1 - \alpha_1 L - \beta_1 L\right) q_t + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

and (1.22) as

$$\left(1-\rho L\right) q_t = \omega + \phi\left(\varepsilon_{t-1}^2 - \sigma_{t-1}^2\right) \tag{1.25}$$

and multiplying by $(1-\rho L)$ the general component model reduces to

$$\left(1-\rho L\right)\sigma_t^2 = \left(1-\rho L\right)\left[\left(1-\alpha_1 L - \beta_1 L\right) q_t + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2\right] \tag{1.26}$$

substituting (1.25) into (1.26)

$$\left(1-\rho L\right)\sigma_t^2 = \left(1-\alpha_1 L - \beta_1 L\right)\left[\omega + \phi\left(\varepsilon_{t-1}^2 - \sigma_{t-1}^2\right)\right] + \left(1-\rho L\right)\left(\alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2\right)$$

$$\left(1-\rho L\right)\sigma_t^2 = \left(1-\alpha_1-\beta_1\right)\omega + \left(1-\alpha_1 L - \beta_1 L\right)\phi\left(\varepsilon_{t-1}^2 - \sigma_{t-1}^2\right) + \left(1-\rho L\right)\left(\alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2\right)$$

$$\begin{aligned}\left(1-\rho L\right)\sigma_t^2 \;&=\; \left(1-\alpha_1-\beta_1\right)\omega + \left(\phi+\alpha_1\right)\varepsilon_{t-1}^2 + \left(-\rho\alpha_1 - \left(\alpha_1+\beta_1\right)\phi\right)\varepsilon_{t-2}^2 \\ &\quad + \left(\rho - \phi + \beta_1\right)\sigma_{t-1}^2 + \left(\phi\left(\alpha_1+\beta_1\right) - \beta_1\rho\right)\sigma_{t-2}^2\end{aligned}$$

A GARCH(2,2) process represents the underlying data generating process for the conditional variance defined in the component model. When $\rho = \phi = 0$, then the component model will reduce to the GARCH(1,1). So the GARCH(1,1) only describes a single dynamic component of the conditional variance.

1.5 Asymmetric Models

1.5.1 The EGARCH(p,q) Model

The simple structure of (1.10) imposes important limitations on GARCH models.

- The negative correlation between stock returns and changes in returns volatility, i.e. volatility tends to rise in response to "bad news", (excess returns lower than expected) and to fall in response to "good news" (excess returns higher than expected). GARCH models, however, assume that only the magnitude and not the positivity or negativity of unanticipated excess returns determines feature σ_t^2. If the distribution of z_t is symmetric, the change in variance tomorrow is conditionally uncorrelated with excess returns today (Nelson [24]). If we write σ_t^2 as a function of lagged σ_t^2 and lagged z_t^2, where $\varepsilon_t^2 = z_t^2\sigma_t^2$

$$\sigma_t^2 = \omega + \sum_{j=1}^{q}\alpha_j z_{t-j}^2\sigma_{t-j}^2 + \sum_{i=1}^{p}\beta_i\sigma_{t-i}^2$$

  it is evident that the conditional variance is invariant to changes in sign of the $z_t's$. Moreover, the innovations $z_{t-j}^2\sigma_{t-j}^2$ are not *i.i.d.*

- Another limitation of GARCH models results from the nonnegativity constraints on ω^* and ϕ_k in (1.12), which are imposed to ensure that σ_t^2 remains nonnegative for all t with probability one. These constraints imply that increasing z_t^2 in any period increases σ_{t+m}^2 for all $m \geq 1$, ruling out random oscillatory behavior in the σ_t^2 process.

- The GARCH models are not able to explain the observed covariance between ε_t^2 and ε_{t-j}. This is possible only if the conditional variance is expressed as an asymmetric function of ε_{t-j}.

- In GARCH(1,1) model, shocks may persist in one norm and die out in another, so the conditional moments of GARCH(1,1) may explode even when the process is strictly stationary and ergodic.

- GARCH models essentially specify the behavior of the square of the data. In this case a few large observations can dominate the sample.

The asymmetric models provide an explanation for the so called *leverage effect*, i.e. an unexpected price drop increases volatility more than an analogous unexpected price increase. The *EGARCH(p,q)* model *(Exponential GARCH(p,q))* put forward by Nelson [24] provides a first explanation for the σ_t^2 depends on both size and the sign of lagged residuals. This is the first example of asymmetric model:

$$\ln\left(\sigma_t^2\right) = \omega + \sum_{i=1}^{p}\beta_i \ln\left(\sigma_{t-i}^2\right) + \sum_{i=1}^{q}\alpha_i\left[\phi z_{t-i} + \psi\left(|z_{t-i}| - E\,|z_{t-i}|\right)\right] \tag{1.27}$$

$\alpha_1 \equiv 1$, $E|z_t| = (2/\pi)^{1/2}$given that $z_t \sim NID(0,1)$, where the parameters ω, β_i, α_i are not restricted to be nonnegative. Let define

$$g(z_t) \equiv \phi z_t + \psi\left[|z_t| - E|z_t|\right]$$

by construction $\{g(z_t)\}_{t=-\infty}^{\infty}$ is a zero-mean, i.i.d. random sequence. The components of $g(z_t)$ are ϕz_t and $\psi\left[|z_t| - E|z_t|\right]$, each with mean zero. If the distribution of z_t is symmetric, the components are orthogonal, though they are not independent. Over the range $0 < z_t < \infty$, $g(z_t)$ is linear in z_t with slope $\phi + \psi$, and over the range $-\infty < z_t \leq 0$, $g(z_t)$ is linear with slope $\phi - \psi$. Thus, $g(z_t)$ allows for the conditional variance process $\{\sigma_t^2\}$ to respond asymmetrically to rises and falls in stock price. The term $\psi\left[|z_t| - E|z_t|\right]$ represents a magnitude effect. If $\psi > 0$ and $\phi = 0$, the innovation in $\ln(\sigma_{t+1}^2)$ is positive (negative) when the magnitude of z_t is larger (smaller) than its expected value. If $\psi = 0$ and $\phi < 0$, the innovation in conditional variance is now positive (negative) when returns innovations are negative (positive).

A negative shock to the returns which would increase the debt to equity ratio and therefore increase uncertainty of future returns could be accounted for when $\alpha_i > 0$ and $\phi < 0$.

In the EGARCH model $\ln\left(\sigma_{t+1}^2\right)$ is homoskedastic conditional on σ_t^2, and the partial correlation between z_t and $\ln\left(\sigma_{t+1}^2\right)$ is constant conditional on σ_t^2.

An alternative possible specification of the news impact curve is the following (Bollerslev, Engle, Nelson (1994))

$$g(z_t, \sigma_t^2) = \sigma_t^{-2\theta_0} \frac{\theta_1}{1 + \theta_2 |z_t|} + \sigma_t^{-2\gamma_0}\left[\frac{\gamma_1 |z_t|^\rho}{1 + \gamma_2 |z_t|^\rho} - E_t\left(\frac{\gamma_1 |z_t|^\rho}{1 + \gamma_2 |z_t|^\rho}\right)\right]$$

The parameters γ_0 and θ_0 parameters allow both the conditional variance of $\ln\left(\sigma_{t+1}^2\right)$ and its conditional correlation with z_t to vary with the level of σ_t^2.

If $\theta_1 < 0$ then $Corr_t(\ln\left(\sigma_{t+1}^2\right), z_t) < 0$: *leverage effect*.

The EGARCH model constraints $\theta_0 = \gamma_0 = 0$, so that the conditional correlation is constant, as is the conditional variance of $\ln(\sigma_t^2)$.

The ρ, γ_2, and θ_2 parameters give the model flexibility in how much weight to assign to the tail observations: e.g., $\gamma_2 > 0$, $\theta_2 > 0$, the model downweights large $|z_t|$'s.

A number of authors, e.g., Nelson ([24]), have found that standardized residuals from estimated GARCH models are leptokurtic relative to the normal, see also Engle and Gonzalez-Rivera ([18]). Nelson [24] assumes that z_t has a GED distribution (also called the exponential power family). The density of a GED random variable normalized to have mean of zero and a variance of one is given by:

$$f(z; \upsilon) = \frac{\upsilon \exp\left[-\left(\frac{1}{2}\right)|z/\lambda|^\upsilon\right]}{\lambda 2^{(1+1/\upsilon)}\Gamma(1/\upsilon)} \qquad -\infty < z < \infty, 0 < \upsilon \leq \infty$$

where $\Gamma(\cdot)$ is the gamma function, and

$$\lambda \equiv \left[2^{(-2/\upsilon)}\Gamma(1/\upsilon)/\Gamma(3/\upsilon)\right]^{1/2}$$

υ is a tail thickness parameter. When $\upsilon = 2$, z has a standard normal distribution. For $\upsilon < 2$, the distribution of z has thicker tails than the normal (e.g. when $\upsilon = 1$, z has a double exponential distribution) and for $\upsilon > 2$, the distribution of z has thinner tails than the normal (e.g., for $\upsilon = \infty$, z is uniformly distributed on the interval $\left[-3^{1/2}, 3^{1/2}\right]$). With this density, we obtain that $E|z_t| = \dfrac{\lambda 2^{1/\upsilon}\Gamma(2/\upsilon)}{\Gamma(1/\upsilon)}$ (Hamilton, [21]).

More general than the GED we have the Generalized t Distribution, which takes the form:

$$f\left(\varepsilon_t\sigma_t^{-1};\eta,\zeta\right) = \frac{\eta}{2\sigma_t b\zeta^{1/\eta}B(1/\eta,\zeta)\left[1+|\varepsilon_t|^\eta/\left(\zeta b^\eta\sigma_t^\eta\right)\right]^{\zeta+1/\eta}}$$

where $B(1/\eta,\zeta) \equiv \Gamma(1/\eta)\Gamma(\zeta)\Gamma(1/\eta+\zeta)$ denotes the beta function,

$$b \equiv \left[\Gamma(\zeta)\Gamma(1/\eta)/\Gamma(3/\eta)\Gamma(\zeta-2/\eta)\right]^{1/2}$$

and $\zeta\eta > 2$, $\eta > 0$ and $\zeta > 0$. The factor b makes $Var\left(\varepsilon_t\sigma_t^{-1}\right) = 1$. The Generalized t nests both the Student's t distribution and the GED. The GED is obtained for $\zeta = \infty$. The GED has only one shape parameter η, which is apparently insufficient to fit both the central part and the tails of the conditional districbution.

Stationarity

In order to simply state the stationarity conditions, we write the EGARCH(p,q) model as:

$$\begin{aligned}
\left[1-\sum_{i=1}^{p}\beta_i L^i\right]\ln\left(\sigma_t^2\right) &= \omega + \sum_{i=1}^{q}\alpha_i L^i\left[\phi z_t + \psi\left(|z_t| - E|z_t|\right)\right] \\
\ln\left(\sigma_t^2\right) &= \left[1-\sum_{i=1}^{p}\beta_i\right]^{-1}\omega + \left[1-\sum_{i=1}^{p}\beta_i L^i\right]^{-1}\left[\sum_{i=1}^{q}\alpha_i L^i\right]g(z_t) \\
\ln\left(\sigma_t^2\right) &= \omega^* + \sum_{i=1}^{\infty}\varphi_i g(z_{t-i}).
\end{aligned}$$

In the EGARCH(p,q) model $\ln\left(\sigma_t^2\right)$ is a linear process, and its stationarity (covariance or strict) and ergodicity are easily checked.

Given $\phi \neq 0$ or $\psi \neq 0$, then

$$\left|\ln\left(\sigma_t^2\right) - \omega^*\right| < \infty \quad \text{a.s.} \quad \text{when } \sum_{i=1}^{\infty}\varphi_i^2 < \infty$$

follows from the independence and finite variance of the $g(z_t)$ and from Billingsley (1986, Theorem 22.6). From this we have that

$$\begin{aligned} \left|\ln\left(\frac{\sigma_t^2}{\exp(\omega^*)}\right)\right| &< \infty \quad \text{a.s.} \\ \left|\frac{\sigma_t^2}{\exp(\omega^*)}\right| &< \infty \quad \text{a.s.} \end{aligned}$$

$\{\exp(-\omega^*)\,\sigma_t^2\}$, $\{\exp(-\omega^*/2)\,\varepsilon_t\}$, where $\varepsilon_t = z_t\sigma_t$, z_t is i.i.d., are ergodic and strictly stationarity. For all t $E[\ln(\sigma_t^2) - \omega^*] = 0$ and the variance $Var[\ln(\sigma_t^2) - \omega^*] = Var(g(z_t))\sum_{i=1}^{\infty}\varphi_i^2$. Since $Var(g(z_t))$ is finite and the distribution of $(\ln(\sigma_t^2) - \omega^*)$ is independent of t, the first two moments of $(\ln(\sigma_t^2) - \omega^*)$ are finite and time invariance, so $(\ln(\sigma_t^2) - \omega^*)$ is covariance stationary if $\sum_{i=1}^{\infty}\varphi_i^2 < \infty$. If $\sum_{i=1}^{\infty}\varphi_i^2 = \infty$, then $|\ln(\sigma_t^2) - \omega^*| = \infty$ almost surely.

Since $\ln(\sigma_t^2)$ is written in ARMA(p,q) form, when $\left[1 - \sum_{i=1}^{p}\beta_i x^i\right]$ and $\left[\sum_{i=1}^{q}\alpha_i x^i\right]$ have no common roots, conditions for strict stationarity of $\ln(\sigma_t^2)$ are equivalent to all the roots of $\left[1 - \sum_{i=1}^{p}\beta_i x^i\right]$ lying outside the unit circle.

The strict stationarity of $\{\exp(-\omega^*)\,\sigma_t^2\}$, $\{\exp(-\omega^*/2)\,\varepsilon_t\}$ need not imply covariance stationarity, since $\{\exp(-\omega^*)\,\sigma_t^2\}$, $\{\exp(-\omega^*/2)\,\varepsilon_t\}$ may fail to have finite unconditional means and variances. For some distribution of $\{z_t\}$ (e.g., the Student t with finite degrees of freedom[§]), $\{\exp(-\omega^*)\,\sigma_t^2\}$ and $\{\exp(-\omega^*/2)\,\varepsilon_t\}$ typically have no finite unconditional moments. If the distribution of z_t is GED and is thinner-tailed than the double exponential, and if $\sum_{i=1}^{\infty}\varphi_i < \infty$, then $\{\exp(-\omega^*)\,\sigma_t^2\}$ and $\{\exp(-\omega^*/2)\,\varepsilon_t\}$ are not only strictly stationary and ergodic, but have arbitrary finite moments, which in turn implies that they are covariance stationary.

1.5.2 Other Asymmetric Models

There is a long tradition in finance that models stock return volatility as negatively correlated with stock returns. The explanation for this phenomenon is based on leverage. A drop in the value of the stock (negative return) increases financial leverage, which makes the stock riskier and increases its volatility. The news have asymmetric effects on volatility. In the aymmetric volatility models good news and bad news have different predictability for future volatility.

[§]The Student t distribution is:

$$f[z;\eta] = [\pi(\eta-2)]^{-1/2}\,\Gamma\left[\frac{1}{2}(\eta+1)\right]\Gamma\left(\frac{\eta}{2}\right)^{-1}\left[1 + z(\eta-2)^{-1}\right]^{-(\eta+1)/2}$$

as η (degree of freedom) goes to infinity the t distribution converges to the normal. When $4 < \eta < \infty$, the kurtosis coefficient is $k = \frac{3(\eta-2)}{(\eta-4)} > 3$.

The *Non linear ARCH(1)* model (Engle - Bollerslev [11]):

$$\sigma_t^\gamma = \omega + \sum_{i=1}^{q} \alpha_i \left|\varepsilon_{t-i}\right|^\gamma + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^\gamma$$

$$\sigma_t^\gamma = \omega + \sum_{i=1}^{q} \alpha_i \left|\varepsilon_{t-i} - k\right|^\gamma + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^\gamma$$

for $k \neq 0$, the innovations in σ_t^γ will depend on the size as well as the sign of lagged residuals, thereby allowing for the leverage effect in stock return volatility.

The Glosten - Jagannathan - Runkle model[19]:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{i=1}^{q} \left(\alpha_i \varepsilon_{t-1}^2 + \gamma_i S_{t-i}^- \varepsilon_{t-i}^2\right)$$

where

$$S_t^- = \begin{cases} 1 & if \quad \varepsilon_t < 0 \\ 0 & if \quad \varepsilon_t \geq 0 \end{cases}$$

The *Asymmetric GARCH(p,q)* model (Engle, [10]):

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \left(\varepsilon_{t-i} + \gamma\right)^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

The *QGARCH* by Sentana (Sentana, [28]):

$$\sigma_t^2 = \sigma^2 + \Psi' x_{t-q} + x_{t-q}' A x_{t-q} + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

when $x_{t-q} = (\varepsilon_{t-1}, \ldots, \varepsilon_{t-q})'$. The linear term $(\Psi' x_{t-q})$ allows for asymmetry. The off-diagonal elements of A accounts for interaction effects of lagged values of x_t on the conditional variance.

The QGARCH nests several asymmetric models. The augmented GARCH assumes $\Psi = 0$ (Bera and Lee, 1990). The *ARCH(q)* model corresponds to $\Psi = 0$, $\beta_i = 0$ and A diagonal. The asymmetric GARCH model assumes A to be diagonal. The linear standard deviation model (Robinson, 1991) corresponds to $\beta_i = 0$, $\sigma^2 = \rho^2$, $\Psi = 2\rho\phi$ and $A = \phi\phi'$

$$\sigma_t^2 = \left(\rho + \phi' x_{t-q}\right)^2 .$$

The Conditional Standard Deviation Model (Taylor, [29])

$$\sigma_t = \omega + + \sum_{i=1}^{q} \alpha_i \left|\varepsilon_{t-i}\right| + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

the conditional standard deviation is a distributed lag of absolute residuals.

1.6 The News Impact Curve

The news have asymmetric effects on volatility. In the aymmetric volatility models good news and bad news have different predictability for future volatility. The news impact curve characterizes the impact of past return shocks on the return volatility which is implicit in a volatility model.

Holding constant the information dated $t-2$ and earlier, we can examine the implied relation between ε_{t-1} and σ_t^2, with $\sigma_{t-i}^2 = \sigma^2$ $i = 1, \ldots, p$. This curve is called, with all lagged conditional variances evaluated at the level of the unconditional variance of the stock return, the news impact curve because it relates past return shocks (news) to current volatility. This curve measures how new information is incorporated into volatility estimates.

For the GARCH model the News Impact Curve (NIC) is centered on $\epsilon_{t-1} = 0$. In the case of EGARCH model the curve has its minimum at $\epsilon_{t-1} = 0$ and is exponentially increasing in both directions but with different paramters.

GARCH(1,1):

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

The news impact curve has the following expression:

$$\sigma_t^2 = A + \alpha\epsilon_{t-1}^2$$

$$A \equiv \omega + \beta\sigma^2$$

EGARCH(1,1):

$$\ln\left(\sigma_t^2\right) = \omega + \beta\ln\left(\sigma_{t-1}^2\right) + \phi z_{t-1} + \psi\left(|z_{t-1}| - E\,|z_{t-1}|\right)$$

where $z_t = \epsilon_t/\sigma_t$. The news impact curve is

$$\sigma_t^2 = \begin{cases} A\exp\left[\dfrac{\phi+\psi}{\sigma}\epsilon_{t-1}\right] & for \quad \epsilon_{t-1} > 0 \\ A\exp\left[\dfrac{\phi-\psi}{\sigma}\epsilon_{t-1}\right] & for \quad \epsilon_{t-1} < 0 \end{cases}$$

$$\begin{aligned} A &\equiv \sigma^{2\beta}\exp\left[\omega - \psi\sqrt{2/\pi}\right] \\ \phi &< 0 \qquad \psi + \phi > 0 \end{aligned}$$

- The EGARCH allows good news and bad news to have different impact on volatility, while the standard GARCH does not.
- The EGARCH model allows big news to have a greater impact on volatility than GARCH model. EGARCH would have higher variances in both directions because the exponential curve eventually dominates the quadrature.

The Asymmetric GARCH(1,1) (Engle, 1990)

$$\sigma_t^2 = \omega + \alpha\left(\epsilon_{t-1} + \gamma\right)^2 + \beta\sigma_{t-1}^2$$

the NIC is

$$\sigma_t^2 = A + \alpha\left(\epsilon_{t-1} + \gamma\right)^2$$

$$A \equiv \omega + \beta\sigma^2$$

$$\omega > 0, 0 \leq \beta < 1, \sigma > 0, 0 \leq \alpha < 1.$$

is asymmetric and centered at $\epsilon_{t-1} = -\gamma$.
The Glosten-Jagannathan-Runkle model

$$\sigma_t^2 = \omega + \alpha\epsilon_t^2 + \beta\sigma_{t-1}^2 + \gamma S_{t-1}^- \epsilon_{t-1}^2$$

$$S_{t-1}^- = \begin{cases} 1 & if \quad \epsilon_{t-1} < 0 \\ 0 & otherwise \end{cases}$$

The NIC is

$$\sigma_t^2 = \begin{cases} A + \alpha\epsilon_{t-1}^2 & if \quad \epsilon_{t-1} > 0 \\ A + (\alpha + \gamma)\,\epsilon_{t-1}^2 & if \quad \epsilon_{t-1} < 0 \end{cases}$$

$$A \equiv \omega + \beta\sigma^2$$

$$\omega > 0, 0 \leq \beta < 1, \sigma > 0, 0 \leq \alpha < 1, \alpha + \beta < 1$$

is centered at $\epsilon_{t-1} = -\gamma$.

These differences between the news impact curves of the models have important implications for portfolio selection and asset pricing. Since predictable market volatility is related to market premium, the two models imply very different market risk premiums, and hence different risk premiums for individual stocks under conditional version of CAPM. Differences in predicted volatility after the arrival of some major news leads to a significant difference in the current option price and to different dynamic hedging strategies.

1.7 The GARCH-*in-mean* Model

The GARCH-in-mean (GARCH-M) proposed by Engle, Lilien and Robins (1987) consists of the system:

$$y_t = \gamma_0 + \gamma_1 x_t + \gamma_2 g\left(\sigma_t^2\right) + \epsilon_t$$

$$\sigma_t^2 = \beta_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-1}^2$$

$$\epsilon_t \mid \Phi_{t-1} \sim N(0, \sigma_t^2)$$

where y_t is a financial return. This model characterizes the evolution of the mean and the variance of a time series simultaneously. The process specifying the conditional variance is a GARCH(1,1) process. Engle, Lilien and Robbins ([13]) extend the Engle's ARCH model to allow the conditional variance to be a determinant of the conditional mean of the process, i.e., the expected risk premium. They consider an economy where risk averse economic agents choose among two kind of financial investment in order to maximize their expected utility. The first possibility is represented by a risky asset with normally distributed returns, i.e., the risky is measured by the asset return variance and the compensation by a rise in the expected returns. The second investment choice is represented by a riskless asset. The agents utility function maximization subject to the market clearing conditions lead to the traditional relation between the mean and the variance of the risky asset return. Engle, Lilien and Robbins investigate the previous relation when the risky asset variance changes over time and therefore the risky asset price will change as well. The above assumptions determine a relation between the mean and the variance of asset return that is still positive but not constant. The GARCH-M model therefore allows to analize the possibility of time-varying risk premium. When $y_t \equiv (r_t - r_f)$, where $(r_t - r_f)$ is the risk premium on holding the asset, then the GARCH-M represents a simple way to model the relation between risk premium and its conditional variance:

$$y_t = \gamma_0 + \gamma_1 x_t + \gamma_2 g\left(\sigma_t^2\right) + \epsilon_t$$

$$\sigma_t^2 = \beta_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-1}^2$$

$$\epsilon_t \mid \Phi_{t-1} \sim N(0, \sigma_t^2)$$

It turns out that:

$$y_t \mid \Phi_{t-1} = (r_t - r_f) \mid \Phi_{t-1} \sim N(\gamma_0 + \gamma_1 x_t + \gamma_2 g\left(\sigma_t^2\right), \sigma_t^2)$$

In applications, $g\left(\sigma_t^2\right) = \sqrt{\sigma_t^2}$, $g\left(\sigma_t^2\right) = \ln\left(\sigma_t^2\right)$ and $g\left(\sigma_t^2\right) = \sigma_t^2$ have been used.

Let $\phi = \left(\gamma_0, \gamma_1, \gamma_2, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p\right)$ be the parameters vector. The procedure utilized in estimating ϕ is the maximization of the conditional log likelihood function which, under the assumption of *i.i.d.* distribution of error process becomes:

$$L(\phi) = \sum_{t=1}^{T} L_t(\phi) = \sum_{t=1}^{T} (-\frac{1}{2} log(\sigma_t^2) - \frac{\epsilon_t^2}{2\sigma_t^2})$$

Moreover the consistency of the parameters estimation requires that both the first two conditional moments are correctly specified and simultaneously estimated. The GARCH-in-mean model can be used to estimate the conditional CAPM.

1.8 Long memory in stock returns

The Asymmetric Power ARCH (Ding, Engle and Granger, 1993)

$$r_t = \mu + \epsilon_t$$

$$\begin{aligned} \epsilon_t &= \sigma_t z_t \\ z_t &\sim N(0,1) \end{aligned}$$

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i \left(|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i}\right)^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta$$

where

$$\begin{aligned} \omega &> 0 \\ \delta &\geq 0 \\ \alpha_i &\geq 0 \qquad i = 1, \ldots, q \\ -1 &< \gamma_i < 1 \qquad i = 1, \ldots, q \\ \beta_j &\geq 0 \qquad j = 1, \ldots, p \end{aligned}$$

This model imposes a Box-Cox transformation of the conditional standard deviation process and the asymmetric absolute residuals. The asymmetric response of volatility to positive and negative "shocks" is the well known leverage effect.

If we assume the distribution of r_t is conditionally normal, then the condition for existence of $E\left[\sigma_t^\delta\right]$ and $E\left|\epsilon_t\right|^\delta$ is

$$\frac{1}{\sqrt{2\pi}} \sum_{i=1}^{q} \alpha_i \left\{ (1+\gamma_i)^\delta + (1-\gamma_i)^\delta \right\} 2^{\frac{\delta-1}{2}} \Gamma\left(\frac{\delta+1}{2}\right) + \sum_{j=1}^{p} \beta_j < 1.$$

If this condition is satisfied, then when $\delta \geq 2$ we have ϵ_t covariance stationary. But $\delta \geq 2$ is a sufficient condition for ϵ_t to be covariance stationary.

This generalized version of ARCH model includes seven other models as special cases.

1. ARCH(q) model, just let $\delta = 2$ and $\gamma_i = 0$, $i = 1, \ldots, q$, $\beta_j = 0$, $j = 1, \ldots, p$.

2. GARCH(p,q) model just let $\delta = 2$ and $\gamma_i = 0$, $i = 1, \ldots, q$.

3. Taylor/Schwert's GARCH in standard deviation model just let $\delta = 1$ and $\gamma_i = 0$, $i = 1, \ldots, q$.

4. GJR model just let $\delta = 2$.

   When $\delta = 2$ and $0 \leq \gamma_i < 1$

$$\begin{aligned}\sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i \left(|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i}\right)^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \\ &= \omega + \sum_{i=1}^{q} \alpha_i \left(|\epsilon_{t-i}|^2 + \gamma_i^2 \epsilon_{t-i}^2 - 2\gamma_i |\epsilon_{t-i}| \epsilon_{t-i}\right) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2\end{aligned}$$

$$\sigma_t^2 = \begin{cases} \omega + \sum_{i=1}^{q} \alpha_i^2 (1+\gamma_i)^2 \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 & \epsilon_{t-i} < 0 \\ \omega + \sum_{i=1}^{q} \alpha_i (1-\gamma_i)^2 \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 & \epsilon_{t-i} > 0 \end{cases}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (1-\gamma_i)^2 \epsilon_{t-i}^2 + \sum_{i=1}^{q} \alpha_i \left\{(1+\gamma_i)^2 - (1-\gamma_i)^2\right\} S_i^- \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (1-\gamma_i)^2 \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} 4\alpha_i \gamma_i S_i^- \epsilon_{t-i}^2$$

$$S_i^- = \begin{cases} 1 & if \quad \epsilon_{t-i} < 0 \\ 0 & otherwise \end{cases}$$

If we define

$$\begin{aligned}\alpha_i^* &= \alpha_i (1-\gamma_i)^2 \\ \gamma_i^* &= 4\alpha_i \gamma_i\end{aligned}$$

then we have

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i (1-\gamma_i)^2 \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{p} \gamma_i^* S_i^- \epsilon_{t-i}^2$$

which is the GJR model.

When $-1 \leq \gamma_i < 0$ we have

$$\begin{aligned}
\sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i \left(|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i}\right)^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \\
&= \omega + \sum_{i=1}^{q} \alpha_i \left(|\epsilon_{t-i}|^2 + \gamma_i^2 \epsilon_{t-i}^2 - 2\gamma_i |\epsilon_{t-i}| \epsilon_{t-i}\right) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \\
&= \omega + \sum_{i=1}^{q} \alpha_i (1-\gamma_i)^2 \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad \epsilon_{t-i} > 0 \\
&= \omega + \sum_{i=1}^{q} \alpha_i (1+\gamma_i)^2 \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad \epsilon_{t-i} < 0 \\
&= \omega + \sum_{i=1}^{q} \alpha_i (1+\gamma_i)^2 \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i \left\{(1-\gamma_i)^2 - (1+\gamma_i)^2\right\} S_i^+ \epsilon_{t-i}^2 \\
&= \omega + \sum_{i=1}^{q} \alpha_i (1+\gamma_i)^2 \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i \left\{1 + \gamma_i^2 - 2\gamma_i - 1 - \gamma_i^2 - 2\gamma_i\right\} S_i^+ \epsilon_{t-i}^2 \\
&= \omega + \sum_{i=1}^{q} \alpha_i (1+\gamma_i)^2 \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 - \sum_{i=1}^{q} 4\alpha_i \gamma_i S_i^+ \epsilon_{t-i}^2
\end{aligned}$$

$$S_i^+ = \begin{cases} 1 & if \quad \epsilon_{t-i} > 0 \\ 0 & otherwise \end{cases}$$

define

$$\begin{aligned}
\alpha_i^* &= \alpha_i (1+\gamma_i)^2 \\
\gamma_i^* &= -4\alpha_i \gamma_i
\end{aligned}$$

we have

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i^* \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \gamma_i^* S_i^+ \epsilon_{t-i}^2$$

which allows positive shocks to have a stronger effect on volatility.

5. Zakoian's TARCH model (Zakoian (1991)), let $\delta = 1$ and $\beta_j = 0$, $j = 1, \ldots, , p$. We have

$$\begin{aligned}
\sigma_t &= \omega + \sum_{i=1}^{q} \alpha_i \left(|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i}\right) \\
&= \omega + \sum_{i=1}^{q} \alpha_i (1-\gamma_i) \epsilon_{t-i}^+ - \sum_{i=1}^{q} \alpha_i (1+\gamma_i) \epsilon_{t-i}^-
\end{aligned}$$

where

$$\epsilon_{t-i}^{+} = \begin{cases} \epsilon_{t-i} & if \quad \epsilon_{t-i} > 0 \\ 0 & otherwise \end{cases}$$

and

$$\epsilon_{t-i}^{-} = \epsilon_{t-i} - \epsilon_{t-i}^{+}$$

Defining

$$\begin{aligned} \alpha_i^{+} &= \alpha_i \left(1 - \gamma_i\right) \\ \gamma_i^{-} &= \alpha_i \left(1 + \gamma_i\right) \end{aligned}$$

$$\sigma_t = \omega + \sum_{i=1}^{q} \alpha_i \left(1 - \gamma_i\right) \epsilon_{t-i}^{+} - \sum_{i=1}^{q} \alpha_i \left(1 + \gamma_i\right) \epsilon_{t-i}^{-}$$

If we let $\beta_j \neq 0,\ j = 1, \ldots, , q$ then we get a more general class of TARCH models.

Chapter 2
ESTIMATION PROCEDURES

The procedure most often used in estimating θ_0 in ARCH models involves the maximization of a likelihood function constructed under the auxiliary assumption of an i.i.d. distribution for the standardized innovation $z_t(\theta)$. Let $f(z_t(\theta);\eta)$ denote the density function for $z_t(\theta) \equiv \varepsilon_t(\theta)/\sigma_t(\theta)$, with mean zero and variance one, where η is the nuisance parameter, $\eta \in H \subseteq R^k$. Let $(y_T, y_{T-1}, \ldots, y_1)$ be a sample realization from an ARCH model as defined by equations (1.1) through (1.5), and $\psi' \equiv (\theta', \eta')$, the combined $(m+k)\times 1$ parameter vector to be estimated for the conditional mean, variance and density functions.

The log-likelihood function for the t-th observation is then given by

$$l_t(y_t;\psi) = \ln\{f[z_t(\theta);\eta]\} - \frac{1}{2}\ln\left[\sigma_t^2(\theta)\right] \qquad t = 1, 2, ... \tag{2.1}$$

The term $-\frac{1}{2}\ln\left[\sigma_t^2(\theta)\right]$ on the right hand side is the Jacobiam that arises in the transformation from the standardized innovations, $z_t(\theta)$, to the observables y_t ($f(y_t;\psi) = f(z_t(\theta);\eta)|J|$, where $J = \dfrac{\partial z_t}{\partial y_t} = \dfrac{1}{\sigma_t(\theta)}$).

The log-likelihood function for the full sample equals the sum of the conditional log likelihoods in eq.(2.1):

$$L_T(y_T, y_{T-1}, \ldots, y_1;\psi) = \sum_{t=1}^{T} l_t(y_t;\psi). \tag{2.2}$$

The maximum likelihood estimator for the true parameters $\psi_0' \equiv (\theta_0', \eta_0')$, say $\widehat{\psi}_T$ is found by the maximization of eq.(2.2). Assuming the conditional density and the $\mu_t(\theta)$ and $\sigma_t^2(\theta)$ functions to be differentiable for all $\psi \in \Theta \times H \equiv \Psi$, the maximum likelihood estimator is the solution to

$$S_T(y_T, y_{T-1}, \ldots, y_1;\psi) \equiv \sum_{t=1}^{T} s_t(y_t;\psi) = 0 \tag{2.3}$$

where $s_t \equiv \dfrac{\partial l_t(y_t,\psi)}{\partial \psi}$ is the score vector for the tth observation. In particular for the conditional mean and variance parameters

$$\frac{\partial l_t\left(y_t,\psi\right)}{\partial\theta}=f\left[z_t\left(\theta\right);\eta\right]^{-1}f'\left[z_t\left(\theta\right);\eta\right]\frac{\partial z_t\left(\theta\right)}{\partial\theta}-\frac{1}{2}\left[\sigma_t^2\left(\theta\right)\right]^{-1}\frac{\partial\sigma_t^2}{\partial\theta} \tag{2.4}$$

where $f'\left[z_t\left(\theta\right);\eta\right]\equiv\dfrac{\partial f\left(z_t\left(\theta\right);\eta\right)}{\partial z_t}$ and

$$\begin{aligned}\frac{\partial z_t\left(\theta\right)}{\partial\theta} &= \frac{\partial}{\partial\theta}\left(\frac{\varepsilon_t\left(\theta\right)}{\sqrt{\sigma_t^2}}\right)=\frac{-\frac{\partial\mu_t}{\partial\theta}\sqrt{\sigma_t^2}-\frac{1}{2}\left(\sigma_t^2\right)^{-1/2}\frac{\partial\sigma_t^2}{\partial\theta}\varepsilon_t\left(\theta\right)}{\sigma_t^2}\\ &= -\frac{\partial\mu_t}{\partial\theta}\left(\sigma_t^2\left(\theta\right)\right)^{-1/2}-\frac{1}{2}\left(\sigma_t^2\left(\theta\right)\right)^{-3/2}\frac{\partial\sigma_t^2}{\partial\theta}\varepsilon_t\left(\theta\right).\end{aligned}$$

where

$$\varepsilon_t\left(\theta\right)\equiv y_t-\mu_t\left(\theta\right).$$

In practice the solution to the set of $m+k$ non-linear equations in (2.3) is found by numerical optimization techniques.

In order to implement the maximum likelihood procedure an explicit assumption regarding the conditional density in eq.(2.1). The most commonly employed distribution in the literature is the normal:

$$f\left[z_t\left(\theta\right);\eta\right]=\left(2\pi\right)^{-1/2}\exp\left\{-\frac{z_t\left(\theta\right)^2}{2}\right\}$$

Since the normal distribution is uniquely determined by its first two moments, only the conditional mean and variance parameters enter the log-likelihood function in equation (2.2); i.e. $\psi=\theta$. The log-likelihood is:

$$l_t=-\frac{1}{2}\ln\left(2\pi\right)-\frac{1}{2}z_t\left(\theta\right)^2-\frac{1}{2}\ln\left(\sigma_t^2\right)$$

it follows that the score vector in eq.(2.4) takes the form:

$$\begin{aligned}s_t &= -z_t\frac{\partial z_t}{\partial\theta}-\frac{1}{2}\left(\sigma_t^2\left(\theta\right)\right)^{-1}\frac{\partial\left(\sigma_t^2\left(\theta\right)\right)}{\partial\theta}\\ &= -\frac{\varepsilon_t\left(\theta\right)}{\sqrt{\sigma_t^2}}\frac{\partial\left(\varepsilon_t\left(\theta\right)\left(\sigma_t^2\right)^{-1/2}\right)}{\partial\theta}-\frac{1}{2}\left(\sigma_t^2\left(\theta\right)\right)^{-1}\frac{\partial\left(\sigma_t^2\left(\theta\right)\right)}{\partial\theta}\\ &= -\frac{\varepsilon_t\left(\theta\right)}{\sqrt{\sigma_t^2}}\frac{\partial\left(\left(y_t-\mu_t\left(\theta\right)\right)\left(\sigma_t^2\right)^{-1/2}\right)}{\partial\theta}-\frac{1}{2}\left(\sigma_t^2\left(\theta\right)\right)^{-1}\frac{\partial\left(\sigma_t^2\left(\theta\right)\right)}{\partial\theta}\\ &= \frac{\varepsilon_t\left(\theta\right)}{\sqrt{\sigma_t^2}}\frac{\partial\mu_t\left(\theta\right)}{\partial\theta}\sigma_t^2\left(\theta\right)^{-1/2}+\frac{1}{2}\left(\sigma_t^2\left(\theta\right)\right)^{-3/2}\frac{\partial\sigma_t^2\left(\theta\right)}{\partial\theta}\frac{\varepsilon_t^2\left(\theta\right)}{\sqrt{\sigma_t^2\left(\theta\right)}}-\frac{1}{2}\left(\sigma_t^2\left(\theta\right)\right)^{-1}\frac{\partial\sigma_t^2\left(\theta\right)}{\partial\theta}\\ &= \frac{\partial\mu_t\left(\theta\right)}{\partial\theta}\frac{\varepsilon_t\left(\theta\right)}{\sigma_t^2\left(\theta\right)}+\frac{1}{2}\frac{\partial\sigma_t^2\left(\theta\right)}{\partial\theta}\frac{\varepsilon_t^2\left(\theta\right)}{\left(\sigma_t^2\left(\theta\right)\right)^2}-\frac{1}{2}\frac{\partial\sigma_t^2\left(\theta\right)}{\partial\theta}\left(\sigma_t^2\left(\theta\right)\right)^{-1}\end{aligned}$$

$$s_t = \frac{\partial \mu_t(\theta)}{\partial \theta}\frac{\varepsilon_t(\theta)}{\sigma_t^2(\theta)} + \frac{1}{2}\left(\sigma_t^2(\theta)\right)^{-1}\frac{\partial \sigma_t^2(\theta)}{\partial \theta}\left[\frac{\varepsilon_t^2(\theta)}{\sigma_t^2(\theta)} - 1\right] \tag{2.5}$$

Several other conditional distributions have been employed in the literature to capture the degree of tail fatness in speculative prices. We have seen above Student's t, GED, Generalized Student's t.

When $\theta = (\alpha', \beta')'$ where α are the conditional mean parameters and β are the conditional variance parameters, the score takes the form:

$$s_t = \begin{pmatrix} \frac{\partial l_t}{\partial \alpha} \\ \frac{\partial l_t}{\partial \beta} \end{pmatrix}$$

where

$$\frac{\partial l_t}{\partial \alpha} = \frac{\partial \mu_t(\alpha)}{\partial \alpha}\frac{\varepsilon_t(\alpha)}{\sigma_t^2(\beta)}$$

$$\frac{\partial l_t}{\partial \beta} = \frac{1}{2}\left(\sigma_t^2(\beta)\right)^{-1}\frac{\partial \sigma_t^2(\beta)}{\partial \beta}\left[\frac{\varepsilon_t^2(\alpha)}{\sigma_t^2(\beta)} - 1\right].$$

Weiss (1986) provided the first study of the asymptotic properties of the ARCH MLE. He showed that the MLE is consistent and asymptotically normal, requiring that the normalized data have finite fourth moments. This rules out IGARCH models. Bollerslev and Wooldridge derive the large sample distribution of the QMLE under high-level assumptions: asymptotic normality of the score vector and uniform weak convergence of the likelihood and its second derivatives. They do not verify conditions or show how they might be verified for GARCH models.

Lumsdaine (1996) imposed assumptions upon the rescaled varaible, ε_t/σ_t, rather than upon the observed data. As auxiliary assumptions, Lumsdaine assumed that the rescaled variable is independent and identically distributed (i.i.d.) and drawn from a symmetric unimodal density with 32nd moment finite. Lee and Hansen (1994) extended this literature to encompass a much broader class of GARCH processes. They focus on QMLE properties. They assume that conditional mean and variance equations have been specified correctly and that a likelihood is used as a vehicle to estimate the parameters. Lee and Hansen stress that there is no reason to assume that all of the conditional dependence is contained in the conditional mean and variance, so, ε_t/σ_t, the rescaled variable need not be independent over time. They allow for some time dependency, specifying that the rescaled variable is strictly stationary and ergodic. For the IGARCH case they are only able to prove the existence of a consistent root of the likelihood. For this result we need that the conditional $2 + \delta$ moment of the rescaled variable is uniformly bounded. Asymptotic normality is proved (including the IGARCH case) by adding the assumption that the conditional fourth moment of the rescaled variable is uniformly bounded.

2.1 Quasi-Maximum Likelihood Estimation

2.1.1 Kullback Information Criterion

In order to study the properties of a model given a set of observations, one can view a model as a good or bad approximation to the "true" but unknown distribution P_0 of the observations. In the first case, one assumes that the distribution P_0 generating the observations belongs to the family of distributions associated with the model, i.e., one assumes that $P_0 \in \mathcal{P}$. When the family is parametric so that $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, the distribution P_0 can be defined through a value θ_0 of the parameter and one has $P_0 = P_{\theta_0}$. This value is called the *true value of the parameter*. The distribution P_0 uniquely defines θ_0 if the mapping $\theta \mapsto P_\theta$ is bijective, i.e., the model is identified. When one believes a priori that the true distribution P_0 does not belong to $\mathcal{P}$, one says that there may be specification errors. Then it is interesting to find the element P_0^* in $\mathcal{P}$ that is closest to P_0 in order to assess the type of specification errors by comparing P_0 to P_0^*. To do this, one must have a measure of the proximity or dicrepancy between the probability distributions. This is provided by the *Kullback Information Criterion.*

Definition 7 *(Kullback Information Criterion). Given two distributions $P = (f\,(y) \cdot \mu)$ and $P^* = (f^*\,(y) \cdot \mu)$ the quantity*

$$I\,(P/P^*) = E_* \log\left(\frac{f^*\,(y)}{f\,(y)}\right) = \int_{\mathcal{Y}} \log\left(\frac{f^*\,(y)}{f\,(y)}\right) f^*\,(y)\,\mu\,(dy)$$

where μ is the common dominating measure, is called the Kullback Information Criterion.

2.1.2 Quasi-Maximum Likelihood Estimation Theory

To study the relationship between an endogenous variable y and some exogenous variables x, one considers a conditional model specifying the form of the conditional distribution of $y_1, \ldots, y_T$ given $x_1, \ldots, x_T$. It's assumed that the model is parametrized by $\theta \in \Theta$ which is an open subset of $\mathcal{R}^p$ the densities can be written as

$$L\,(y_1, \ldots, y_T\,|x_1, \ldots, x_T\,;\theta) = \prod_{i=1}^{T} f\,(y_i\,|x_i\,;\theta) \qquad \theta \in \Theta$$

thus the model implies the mutual independence of the variables $Y_1, \ldots, Y_T$ conditionally on $X_1, \ldots, X_T$ and the equality of the conditional densities $f\,(y_i\,|x_i\,;\theta)$ across observations. We consider the case where the model is misspecified. The true distribution of the observations is given by the density

$$L_0\,(y_1, \ldots, y_T\,|x_1, \ldots, x_T) = \prod_{i=1}^{T} f_0\,(y_i\,|x_i)$$

where $f_0 (y_i |x_i)$ does not belong to the specified density family, $f_0 (y_i |x_i) \notin \{f (y |x ; \theta), \theta \in \Theta\}$.

It's possible to evaluate the discrepancy between the true density f_0 and the model $\{f (y |x ; \theta), \theta \in \Theta\}$ by the Kullback Information Criterion. This leads naturally to the concept of quasi true value θ_0^* of the parameter θ that corresponds to the distribution in the model that is closest to f_0. This quasi true value is a solution to

$$\max_{\theta \in \Theta} E_X E_0 \log f (Y |X ; \theta)$$

where E_0 denotes the conditional expectation of Y given X under f_0. We assume θ_0^* is unique.

Definition 8 *A quasi (or pseudo) maximum likelihood (QML) estimator $\widehat{\theta}_T$ of θ is a solution $\widehat{\theta}_T$ to*

$$\max_{\theta \in \Theta} \sum_{i=1}^{T} \log f (y_i |x_i ; \theta).$$

Thus $\widehat{\theta}_T$ is a maximum likelihood estimator based on a misspecified model. Under regularity conditions, the QML estimator converges almost surely to the pseudo true value θ_0^*.

Proposition 9 *Under regularity conditions, the QML estimator is asymptotically normal distributed with*

$$\sqrt{T} \left(\widehat{\theta}_T - \theta_0^* \right) \xrightarrow{d} N \left(0, A^{-1} B A^{-1} \right)$$

The matrices A and B are, respectively, equal to:

$$A = -\frac{1}{T} E_0 \left[\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'} \right]$$

$$B = \frac{1}{T} E_0 \left[\frac{\partial \log L(\theta)}{\partial \theta} \frac{\partial \log L(\theta)}{\partial \theta'} \right]$$

The matrices A and B are not, in general, equal when specification errors are present. Thus comparing estimates of the matrices A and B can be useful for detecting specification errors.

In the case of univariate GARCH models and when the parameter vector θ is decomposable such as $\theta = (\alpha', \beta')'$ where α are the conditional mean parameters and β are the conditional variance parameters we can show the $A = B$ only under special circumstances.

The second derivatives matrix of the tth log-likelihood function is equal to

$$\begin{aligned}
\frac{\partial^2 l_t}{\partial\theta\partial\theta'} &= \frac{1}{2}\left(\sigma_t^2(\theta)\right)^{-2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}-\frac{1}{2}\left(\sigma_t^2(\theta)\right)^{-1}\frac{\partial^2\sigma_t^2(\theta)}{\partial\theta\partial\theta'}\\
&-\frac{\varepsilon_t^2(\theta)}{\left(\sigma_t^2(\theta)\right)^3}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\\
&+\frac{1}{2}\frac{\varepsilon_t^2(\theta)}{\left(\sigma_t^2(\theta)\right)^2}\frac{\partial^2\sigma_t^2(\theta)}{\partial\theta\partial\theta'}-\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\frac{\varepsilon_t(\theta)}{\left(\sigma_t^2(\theta)\right)^2}\\
&+\frac{\varepsilon_t(\theta)}{\sigma_t^2(\theta)}\frac{\partial^2\mu_t(\theta)}{\partial\theta\partial\theta'}-\left(\sigma_t^2(\theta)\right)^{-1}\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\partial\mu_t(\theta)}{\partial\theta'}\\
&-\frac{\varepsilon_t(\theta)}{\left(\sigma_t^2(\theta)\right)^2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\mu_t(\theta)}{\partial\theta'}.
\end{aligned}$$

given that $E_{t-1}\left[\frac{\varepsilon_t(\theta)}{\left(\sigma_t^2(\theta)\right)^{1/2}}\right]=0$ and $E_{t-1}\left[\frac{\varepsilon_t^2(\theta)}{\left(\sigma_t^2(\theta)\right)}\right]=1$, we have that

$$\begin{aligned}
A_t &= E_0\left[-\frac{\partial^2 l_t}{\partial\theta\partial\theta'}\right]=E_0\left[-\frac{1}{2}\left(\sigma_t^2(\theta)\right)^{-2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}+\frac{1}{2}\left(\sigma_t^2(\theta)\right)^{-1}\frac{\partial^2\sigma_t^2(\theta)}{\partial\partial\theta\partial\theta'}\right]+\\
&E_0\left[\frac{1}{\left(\sigma_t^2(\theta)\right)^2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}E_{t-1}\left[\frac{\varepsilon_t^2(\theta)}{\left(\sigma_t^2(\theta)\right)}\right]\right]+\\
&E_0\left[-\frac{1}{2}\frac{1}{\sigma_t^2(\theta)}\frac{\partial^2\sigma_t^2(\theta)}{\partial\partial\theta\partial\theta'}E_{t-1}\left[\frac{\varepsilon_t^2(\theta)}{\left(\sigma_t^2(\theta)\right)}\right]\right]+\\
&E_0\left[\frac{1}{\left(\sigma_t^2(\theta)\right)^{3/2}}\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}E_{t-1}\left[\frac{\varepsilon_t(\theta)}{\left(\sigma_t^2(\theta)\right)^{1/2}}\right]\right]+\\
&E_0\left[-\frac{1}{\left(\sigma_t^2(\theta)\right)^{1/2}}\frac{\partial^2\mu_t(\theta)}{\partial\theta\partial\theta'}E_{t-1}\left[\frac{\varepsilon_t(\theta)}{\left(\sigma_t^2(\theta)\right)^{1/2}}\right]+\left(\sigma_t^2(\theta)\right)^{-1}\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\partial\mu_t(\theta)}{\partial\theta'}\right]+\\
&E_0\left[\frac{1}{\left(\sigma_t^2(\theta)\right)^2}E_{t-1}\left[\frac{\varepsilon_t(\theta)}{\left(\sigma_t^2(\theta)\right)^{1/2}}\right]\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\mu_t(\theta)}{\partial\theta'}\right].
\end{aligned}$$

$$A_t=E_0\left[\frac{1}{2}\left(\sigma_t^2(\theta)\right)^{-2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}+\left(\sigma_t^2(\theta)\right)^{-1}\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\partial\mu_t(\theta)}{\partial\theta'}\right]$$

The information matrix is

$$
\begin{aligned}
B_t &= E_0\left[\frac{\partial l_t}{\partial\theta}\frac{\partial l_t}{\partial\theta'}\right] = \\
& E_0\left[\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\varepsilon_t(\theta)}{\sigma_t^2(\theta)}+\frac{1}{2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\varepsilon_t^2(\theta)}{\left(\sigma_t^2(\theta)\right)^2}-\frac{1}{2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\left(\sigma_t^2(\theta)\right)^{-1}\right]\times \\
& \left[\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\varepsilon_t(\theta)}{\sigma_t^2(\theta)}+\frac{1}{2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\varepsilon_t^2(\theta)}{\left(\sigma_t^2(\theta)\right)^2}-\frac{1}{2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\left(\sigma_t^2(\theta)\right)^{-1}\right]' \\
&= E_0\left[\frac{1}{4}\frac{1}{\left(\sigma_t^2(\theta)\right)^2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}+\frac{1}{4}\frac{1}{\left(\sigma_t^2(\theta)\right)^2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}E_{t-1}\left[\frac{\varepsilon_t^4(\theta)}{\left(\sigma_t^2(\theta)\right)^2}\right]\right]+ \\
& E_0\left[\left(\sigma_t^2(\theta)\right)^{-1}\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\partial\mu_t(\theta)}{\partial\theta'}E_{t-1}\left[\frac{\varepsilon_t^2(\theta)}{\left(\sigma_t^2(\theta)\right)}\right]\right]+E_0\left[-\frac{1}{2}\frac{1}{\left(\sigma_t^2(\theta)\right)^2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\right]+ \\
& E_0\left[\frac{1}{2}\frac{1}{\left(\sigma_t^2(\theta)\right)^{3/2}}\left(\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\mu_t(\theta)}{\partial\theta'}+\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\right)E_{t-1}\left[\frac{\varepsilon_t^3(\theta)}{\left(\sigma_t^2(\theta)\right)^{3/2}}\right]\right].
\end{aligned}
$$

$$
\begin{aligned}
B_t &= E_0\left[\frac{1}{4}\frac{1}{\left(\sigma_t^2(\theta)\right)^2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\left(K_t(\theta)-1\right)+\left(\sigma_t^2(\theta)\right)^{-1}\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\partial\mu_t(\theta)}{\partial\theta'}\right]+ \\
& E_0\left[\frac{1}{2}\frac{1}{\left(\sigma_t^2(\theta)\right)^3}\left(\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\mu_t(\theta)}{\partial\theta'}+\frac{\partial\mu_t(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\right)M_{3t}(\theta)\right]
\end{aligned}
$$

$M_{3t}(\theta) = E_{t-1}\left[\varepsilon_t^3(\theta)\right]$ and $K_t(\theta) = \dfrac{1}{\left(\sigma_t^2(\theta)\right)^2}E_{t-1}\left[\varepsilon_t^4(\theta)\right]$.

Whenever it is possible to decompose the parameter vector in $\theta = (\alpha', \beta')'$, the hessian matrix for the tth is:

$$
A_t = \begin{bmatrix} E\left[\left(\sigma_t^2(\beta)\right)^{-1}\frac{\partial\mu_t(\alpha)}{\partial\alpha}\frac{\partial\mu_t(\alpha)}{\partial\alpha'}\right] & 0 \\ 0 & E\left[\frac{1}{2}\left(\sigma_t^2(\theta)\right)^{-2}\frac{\partial\sigma_t^2(\beta)}{\partial\beta}\frac{\partial\sigma_t^2(\beta)}{\partial\beta'}\right] \end{bmatrix}
$$

$$
B_t = \begin{bmatrix} E\left[\left(\sigma_t^2(\theta)\right)^{-1}\frac{\partial\mu_t(\alpha)}{\partial\alpha}\frac{\partial\mu_t(\alpha)}{\partial\alpha'}\right] & E\left[\frac{1}{2\left(\sigma_t^2(\alpha)\right)^3}\frac{\partial\mu_t(\alpha)}{\partial\alpha}\frac{\partial\sigma_t^2(\beta)}{\partial\beta'}M_{3t}(\theta)\right] \\ E\left[\frac{1}{2\left(\sigma_t^2(\alpha)\right)^3}\frac{\partial\sigma_t^2(\beta)}{\partial\beta}\frac{\partial\mu_t(\alpha)}{\partial\alpha'}M_{3t}(\theta)\right] & E\left[\frac{1}{4\left(\sigma_t^2(\theta)\right)^2}\frac{\partial\sigma_t^2(\beta)}{\partial\beta}\frac{\partial\sigma_t^2(\beta)}{\partial\beta'}\left(K_t(\theta)-1\right)\right] \end{bmatrix}
$$

The asymptotic variance-covariance matrices of the QML estimators $\widehat{\alpha}_T$ and $\widehat{\beta}_T$ are

$$
Var^{asy}\left[\sqrt{T}\left(\widehat{\alpha}_T-\alpha\right)\right] = \left[\frac{1}{T}\sum E\left[\left(\sigma_t^2(\theta)\right)^{-1}\frac{\partial\mu_t(\alpha)}{\partial\alpha}\frac{\partial\mu_t(\alpha)}{\partial\alpha'}\right]\right]^{-1}
$$

$$\begin{aligned} Var^{asy}\left[\sqrt{T}\left(\widehat{\beta}_T-\beta\right)\right] &= \left[\frac{1}{T}\sum E\left[\frac{1}{2}\left(\sigma_t^2\left(\theta\right)\right)^{-2}\frac{\partial\sigma_t^2\left(\beta\right)}{\partial\beta}\frac{\partial\sigma_t^2\left(\beta\right)}{\partial\beta'}\right]\right]^{-1}\times \\ &\left[\frac{1}{T}\sum E\left[\frac{1}{4\left(\sigma_t^2\left(\theta\right)\right)^2}\frac{\partial\sigma_t^2\left(\beta\right)}{\partial\beta}\frac{\partial\sigma_t^2\left(\beta\right)}{\partial\beta'}\left(K_t\left(\theta\right)-1\right)\right]\right]\times \\ &\left[\frac{1}{T}\sum E\left[\frac{1}{2}\left(\sigma_t^2\left(\theta\right)\right)^{-2}\frac{\partial\sigma_t^2\left(\beta\right)}{\partial\beta}\frac{\partial\sigma_t^2\left(\beta\right)}{\partial\beta'}\right]\right]^{-1} \end{aligned}$$

When the true conditional distribution is normal $M_{3t}\left(\theta\right)=0$ and $K_t\left(\theta\right)=3$. In this case, the expressions for A_t and B_t coincide. The asymptotic variance-covariance matrices of the QML estimator $\widehat{\beta}_T$ reduces to:

$$Var^{asy}\left[\sqrt{T}\left(\widehat{\beta}_T-\beta\right)\right]=\left[\frac{1}{T}\sum E\left[\frac{1}{2}\left(\sigma_t^2\left(\theta\right)\right)^{-2}\frac{\partial\sigma_t^2\left(\beta\right)}{\partial\beta}\frac{\partial\sigma_t^2\left(\beta\right)}{\partial\beta'}\right]\right]^{-1}.$$

2.2 Testing in GARCH models

2.2.1 The GARCH(1,1) case

Suppose the true model is:

$$y_t=\mu+\varepsilon_{0t}$$

$$\sigma_{0t}^2=\omega_0+\alpha_0\varepsilon_{0t-1}^2+\beta_0\sigma_{0t-1}^2$$

$$\varepsilon_{0t}\left|\Psi_{t-1}\right.\sim N\left(0,\sigma_{0t}^2\right)$$

$$\theta_0=\left(\mu_0,\omega_0,\alpha_0,\beta_0\right)'$$

The estimated model is:

$$y_t=\mu+\varepsilon_t$$

$$\sigma_t^2=\omega+\alpha\varepsilon_{t-1}^2+\beta\sigma_{t-1}^2$$

$$\theta=\left(\mu,\omega,\alpha,\beta\right)'$$

$$L_T=\sum l_t\left(\theta\right)$$

$$A_0=-\frac{1}{T}E\left[\frac{\sum\partial^2 l_t\left(\theta_0\right)}{\partial\theta\partial\theta'}\right]$$

$$A_T(\theta) = -\frac{1}{T}\sum \frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}$$

$$A_T(\theta) = \frac{1}{T}\left[\frac{1}{2}\sum_{t=1}^{T}\left(\sigma_t^2\right)^{-2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'} + \sum_{t=1}^{T}\left(\sigma_t^{-2}\right)\frac{\partial\varepsilon_t}{\partial\theta}\frac{\partial\varepsilon_t}{\partial\theta'}\right]$$

$$B_0 = \frac{1}{T}E\left[\frac{\partial L(\theta_0)}{\partial\theta}\frac{\partial L(\theta_0)}{\partial\theta'}\right]$$

$$B_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t}{\partial\theta}\frac{\partial l_t}{\partial\theta'}$$

$A_T\left(\widehat{\theta}\right)$ and $B_T\left(\widehat{\theta}\right)$ are consistent estimators of A and B, $\widehat{\theta}$ is the maximum likelihood estimator of θ_0. When ε_0 is conditionally normal $A = B$. Moreover

$$D\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N(0, I_k)$$

D can be $A^{1/2}$ in the conditionally case or $B^{-1/2}A$ in the general case. Robust form of the t statistics

$$B_T^{-1/2}A_T\sqrt{T}\left(\widehat{\theta} - \theta_0\right)$$

non robust form is:

$$A_T^{1/2}\sqrt{T}\left(\widehat{\theta} - \theta_0\right)$$

The t statistics has to be compared to a standard normal distribution. Define the null hypothesis:

$$H_0 : \alpha_0 + \beta_0 = 1$$

$$H_0 : g(\theta_0) = \alpha_0 + \beta_0 - 1 = 0.$$

Define $\widehat{\theta}_U$ the ML of the unrestricted model and $\widehat{\theta}_R$ the estimator of the restricted model. The statistics are:

$$\xi_{LR} = -2\left[L\left(\widehat{\theta}_R\right) - L\left(\widehat{\theta}_U\right)\right]$$

$$\xi_{LM}^{NR} = \frac{1}{T}\left(\frac{\partial L\left(\widehat{\theta}_R\right)}{\partial\theta}\right)' A_T^{-1}\left(\widehat{\theta}_R\right)\left(\frac{\partial L\left(\widehat{\theta}_R\right)}{\partial\theta}\right)$$

$$\begin{aligned}\xi_{LM}^{R} &= \frac{1}{T}\left[\left(\frac{\partial L\left(\widehat{\theta}_R\right)}{\partial\theta}\right)' A_T^{-1}\left(\widehat{\theta}_R\right)\left(\frac{\partial g\left(\widehat{\theta}_R\right)}{\partial\theta}\right)\right]\times\\ &\left[\left(\frac{\partial g\left(\widehat{\theta}_R\right)}{\partial\theta'}\right) A_T^{-1}\left(\widehat{\theta}_R\right) B_T\left(\widehat{\theta}_R\right) A_T^{-1}\left(\widehat{\theta}_R\right)\right]^{-1}\times\\ &\left[\left(\frac{\partial g\left(\widehat{\theta}_R\right)}{\partial\theta'}\right) A_T^{-1}\left(\widehat{\theta}_R\right)\left(\frac{\partial L\left(\widehat{\theta}_R\right)}{\partial\theta}\right)\right]\end{aligned}$$

$$\xi_W^R = Tg\left(\widehat{\theta}_U\right)'\left[\left(\frac{\partial g\left(\widehat{\theta}_U\right)}{\partial\theta'}\right) A_T^{-1}\left(\widehat{\theta}_U\right) B_T\left(\widehat{\theta}_U\right) A_T^{-1}\left(\widehat{\theta}_U\right)\left(\frac{\partial g\left(\widehat{\theta}_R\right)}{\partial\theta}\right)\right]^{-1} g\left(\widehat{\theta}_U\right)$$

$$\xi_W^{NR} = Tg\left(\widehat{\theta}_U\right)'\left[\left(\frac{\partial g\left(\widehat{\theta}_U\right)}{\partial\theta'}\right) A_T^{-1}\left(\widehat{\theta}_U\right)\left(\frac{\partial g\left(\widehat{\theta}_R\right)}{\partial\theta}\right)\right]^{-1} g\left(\widehat{\theta}_U\right)$$

The Wald statistics (ξ_W^R and ξ_W^{NR}) are the squares of the (robust and non robust, respectively) t statistics for $g\left(\theta\right) = 0$.

2.3 Testing for ARCH disturbances

We want to test for the presence of ARCH effect. This can be done with a LM test. The test is based upon the score under the null and information matrix under the null. The null hypothesis is

$$\alpha_1 = \alpha_2 = \ldots = \alpha_q = 0$$

Consider the ARCH model with $\sigma_t^2 = \sigma^2\left(z_t\alpha\right)$, where $\sigma^2\left(\cdot\right)$ is a differentiable function. $z_t = \left(1, \widehat{\epsilon}_{t-1}^2, \ldots, \widehat{\epsilon}_{t-q}^2\right)$, $\alpha = \left(\alpha_0, \alpha_1, \ldots, \alpha_q\right)'$ where $\widehat{\epsilon}_t$ are the OLS residuals. Under the null, σ_t^2 is a constant $\sigma_t^2 = \sigma_0^2$. The derivative of σ_t^2 with respect to α is

$$\frac{\partial\sigma_t^2}{\partial\alpha} = \sigma^{2\prime} z_t'$$

where $\sigma^{2\prime}$ is the scalar derivative of $\sigma^2\left(z_t\alpha\right)$. Recalling that the log-likelihood function is

$$L_T = \sum_{t=1}^{T} l_t\left(\alpha\right) = \sum_{t=1}^{T}\left[-\frac{1}{2}\log\left(\sigma_t^2\right) - \frac{1}{2}\frac{\widehat{\epsilon}_t^2}{\sigma_t^2}\right]$$

the derivative of l_t with respect to α is:

$$\frac{\partial l_t}{\partial \alpha} = \frac{\sigma^{2\prime} z_t'}{2\sigma_t^2}\left[\frac{\widehat{\epsilon}_t^2}{\sigma_t^2} - 1\right]$$

the score under the null is

$$\frac{\partial L_T}{\partial \alpha}|_0 = \frac{\sigma^{2\prime}}{2\sigma_0^2}\sum_t z_t'\left(\frac{\widehat{\epsilon}_t^2}{\sigma_0^2} - 1\right) = \frac{\sigma^{2\prime}}{2\sigma_0^2} Z' f^0$$

where $f^0 = \left[\left(\frac{\widehat{\epsilon}_1^2}{\sigma_0^2} - 1\right), \ldots, \left(\frac{\widehat{\epsilon}_T^2}{\sigma_0^2} - 1\right)\right]'$ and $Z' = (z_1', \ldots, z_T')$ is a $((q+1)\times T)$ matrix. The second derivatives matrix is

$$\begin{aligned}
\frac{\partial^2 l_t}{\partial \alpha \partial \alpha'} &= -\frac{\sigma^{2\prime} z_t' \sigma^{2\prime} z_t}{2\sigma_t^4}\left[\frac{\widehat{\epsilon}_t^2}{\sigma_t^2} - 1\right] + \frac{\sigma^{2\prime} z_t'}{2\sigma_t^2}\left[\frac{-\sigma^{2\prime} z_t \widehat{\epsilon}_t^2}{\sigma_t^4}\right] \\
&= -\frac{1}{2}\left(\frac{\sigma^{2\prime}}{\sigma_t^2}\right)^2 \frac{\widehat{\epsilon}_t^2}{\sigma_t^2} z_t' z_t + \frac{1}{2}\left(\frac{\sigma^{2\prime}}{\sigma_t^2}\right)^2 z_t' z_t - \frac{1}{2}\left(\frac{\sigma^{2\prime}}{\sigma_t^2}\right)^2 \frac{\widehat{\epsilon}_t^2}{\sigma_t^2} z_t' z_t \\
&= -\left(\frac{\sigma^{2\prime}}{\sigma_t^2}\right)^2 \frac{\widehat{\epsilon}_t^2}{\sigma_t^2} z_t' z_t + \frac{1}{2}\left(\frac{\sigma^{2\prime}}{\sigma_t^2}\right)^2 z_t' z_t
\end{aligned}$$

This yields the information matrix under the null:

$$\begin{aligned}
A_{\alpha\alpha,0} &= -\frac{1}{T} E\left[\frac{\partial^2 L_T}{\partial \alpha \partial \alpha'}\right]|_0 = -\frac{1}{T} E\left[E\left[\sum \frac{\partial^2 l_t}{\partial \alpha \partial \alpha'}|\Phi_{t-1}\right]\right]|_0 \\
&= -\frac{1}{T} E\left[\sum E\left[\frac{\partial^2 l_t}{\partial \alpha \partial \alpha'}|\Phi_{t-1}\right]\right]|_0 \\
&= -\frac{1}{T} E\left[\sum E\left[-\left(\frac{\sigma^{2\prime}}{\sigma_t^2}\right)^2 \frac{\widehat{\epsilon}_t^2}{\sigma_t^2} z_t' z_t + \frac{1}{2}\left(\frac{\sigma^{2\prime}}{\sigma_t^2}\right)^2 z_t' z_t \,|\Phi_{t-1}\right]\right]|_0 \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{-\frac{1}{2}\left(\frac{\sigma^{2\prime}}{\sigma_0^2}\right)^2 E\left[z_t' z_t\right] + \left(\frac{\sigma^{2\prime}}{\sigma_0^2}\right)^2 E\left[z_t' z_t\right]\right\} = \\
&= \frac{1}{2}\left(\frac{\sigma^{2\prime}}{\sigma_0^2}\right)^2 \frac{1}{T}\sum_{t=1}^{T} E\left[z_t' z_t\right].
\end{aligned}$$

The LM statistic is given by

$$\xi_{LM} = \frac{1}{T}\left(\frac{\partial L_T}{\partial \alpha}|_0\right)' A_{\alpha\alpha,0}^{-1}\left(\frac{\partial L_T}{\partial \alpha}|_0\right)$$

$$\begin{aligned}
\xi_{LM} &= f^{0\prime} Z \frac{\sigma^{2\prime}}{2\sigma_0^2}\left[\frac{1}{2}\left(\frac{\sigma^{2\prime}}{\sigma_0^2}\right)^2 \sum_{t=1}^{T} E\left[z_t' z_t\right]\right]^{-1} \frac{\sigma^{2\prime}}{2\sigma_0^2} Z' f^0 \\
&= f^{0\prime} Z\left(\sum_{t=1}^{T} E\left[z_t' z_t\right]\right)^{-1} Z' f^0 / 2
\end{aligned}$$

it can be consistently estimated by

$$\xi_{LM} = f^{0\prime} Z \left(Z'Z\right)^{-1} Z' f^0 / 2.$$

When we assume normality $plim\left(f^{0\prime} f^0/T\right) = 2$. Thus an asymptotically equivalent statistic would be

$$\xi^* = T f^{0\prime} Z \left(Z'Z\right)^{-1} Z' f^0 / \left(f^{0\prime} f^0\right) = TR^2$$

where R^2 is the squared multiple correlation between f^0 and Z. Since adding a constant and multiplying by a scalar will not change the R^2 of a regression, this is also the R^2 of the regression of $\widehat{\epsilon}_t^2$ on an intercept and q lagged values of $\widehat{\epsilon}_t^2$. The statistic will be asymptotically distributed as chi square with q degrees of freedom when the null hypothesis is true.

The test procedure is to run the OLS regression and save the residuals. Regress the squared residuals on a constant and q lags and test TR^2 as a χ_q^2. This will be an asymptotically locally most powerful test.

Lee and King (1993) derive a locally most powerful (LMMP) - based score test for the presence of ARCH and GARCH disturbances. Wald and likelihood ratio (LR) criteria could be used to test the hypothesis of conditional homoskedasticity e.g. against a GARCH(1,1) alternative.

The statistic associated with $H_0 : \alpha_1 = \beta_1 = 0$ against $H_1 : \alpha_1 \geq 0$ or $\beta_1 \geq 0$ with at least one strict inequality do not have a χ^2 distribution with two degrees of freedom can be shown to be conservative.

2.4 Test for Asymmetric Effects

Implicit in any volatility model is a particular news impact curve. The standard GARCH model has news impact curve which is symmetric and centered at $\varepsilon_{t-1} = 0$. That is, positive and negative return shocks of the same magnitude produce the same amount of volatility. Also, larger return shocks forecast more volatility at a rate proportional to the square of the size of the return shock. If a negative return shock causes more volatility than a positive return shock of the same size, the GARCH model underpredicts the amount of volatility following bad news and overpredicts the amount of volatility following good news. Furthermore, if large return shocks cause more volatility than a quadratic function allows,, then the standard GARCH model underpredicts volatility after a large return shock and overpredicts volatility after a small return shock.

Engle and Ng [16] put forward three diagnostic tests for volatility models: the *Sign Bias Test*, the *Negative Size Bias Test*, and the *Positive Size Bias Test*. These tests examine whether we can predict the squared normalized residual by some variables observed in the past which are not included in the volatility model being used. If these variables can predict the squared normalized residual, then the variance model is misspecified. The sign bias test examines the impact of positive and negative return shocks on volatility not predicted by the model under consideration. The negative size bias test focuses on the different effects that large and small negative return shocks have on volatility which are not predicted by the volatilty model. The positive size bias test focuses on the different impacts that large and small positive return shocks may have on volatility, which are not explained by the volatility model.

To derive the optimal form of these tests, we assume that the volatility model under the null hypothesis is a special case of a more general model of the following form:

$$\log\left(\sigma_t^2\right) = \log\left(\sigma_{0t}^2\left(\delta_0' z_{0t}\right)\right) + \delta_a' z_{at} \tag{2.6}$$

where $\sigma_{0t}^2\left(\delta_0' z_{0t}\right)$ is the volatility model hypothesized under the null, δ_0 is a $(k \times 1)$ vector of parameters under the null, z_{0t} is a $(k \times 1)$ vector of explanatory variables under the null, δ_a is a $(m \times 1)$ vector of additional parameters, z_{at} is a $(m \times 1)$ vector of missing explanatory variables.

This form encompasses both the GARCH and EGARCH models. For the GARCH(1,1) model

$$\sigma_{0t}^2\left(\delta_0' z_{0t}\right) = \delta_0' z_{0t}$$

$$z_{0t} \equiv \left[1, \sigma_{t-1}^2, \varepsilon_{t-1}^2\right]'$$

$$\delta_0 \equiv [\omega, \beta, \alpha]'$$

$$\delta_a = [\beta^*, \phi^*, \psi^*]'$$

$$z_{at} = \left[\log\left(\sigma_{t-1}^2\right), \frac{\varepsilon_{t-1}}{\sigma_{t-1}}, \left(\frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - \sqrt{2/\pi}\right)\right]'$$

The encompassing model is

$$\log\left(\sigma_t^2\right) = \log\left[\omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2\right] + \beta^* \log\left(\sigma_{t-1}^2\right) + \phi^* \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \psi^* \left(\frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - \sqrt{2/\pi}\right)$$

when $\alpha = \beta = 0$ is an EGARCH(1,1) while with $\beta^* = \phi^* = \psi^* = 0$ is a GARCH(1,1) model.

The null hypothesis is $\delta_a = 0$. Let υ_t be the normalized residual corresponding to observation t under the volatility model hypothesized. That is, $\upsilon_t \equiv \dfrac{\varepsilon_t}{\sigma_t}$. The LM test statistic for $H_0 : \delta_a = 0$ in (2.6) is a test of $\delta_a = 0$ in the auxiliary regression

$$\upsilon_t^2 = z_{0t}^{*\prime}\delta_0 + z_{at}^{*\prime}\delta_a + u_t \tag{2.7}$$

where $z_{0t}^{*} \equiv \sigma_{0t}^{-2}\left(\frac{\partial \sigma_t^2}{\partial \delta_0}\right)$, $z_{at}^{*} \equiv \sigma_{0t}^{-2}\left(\frac{\partial \sigma_t^2}{\partial \delta_a}\right)$. Both $\frac{\partial \sigma_t^2}{\partial \delta_0}$ and $\frac{\partial \sigma_t^2}{\partial \delta_a}$ are evaluated at $\delta_a = 0$ and δ_0 (the maximum likelihood estimator of δ_0 under H_0). If the parameters restrictions are met, the right-hand side variables in (2.7) should have no explanatory variables power at all. Thus, the test is often computed as

$$\xi_{LM} = TR^2$$

where R^2 is the squared multiple correlation of (2.7), and T is the number of observations in the sample*. The LM statistic is aasymptotically distributed as chi-square with m degrees of freedom when the null hypothesis is true, where m is the number of parameter restrictions. Under the encompassing model (2.6), $\left(\frac{\partial \sigma_t^2}{\partial \delta_a}\right)$ evaluated under the null is equal to† $\sigma_{0t}^2 z_{at}$, hence $z_{at}^{*} = z_{at}$. The regression actually involves regressing υ_t^2 on a constant z_{0t}^{*} and z_{at}. The variables in z_{at} are S_{t-1}, $S_{t-1}^{-}\varepsilon_{t-1}$ and $S_{t-1}^{+}\varepsilon_{t-1}$. The optimal form for conducting the *sign bias test* is:

$$\upsilon_t^2 = a + b_1 S_{t-1}^{-} + \gamma' z_{0t}^{*} + e_t$$

where

$$S_{t-1}^{-} = \begin{cases} 1 & \varepsilon_{t-1} < 0 \\ 0 & otherwise \end{cases}$$

the regression for the *negative size bias test* is:

$$\upsilon_t^2 = a + b_2 S_{t-1}^{-}\varepsilon_{t-1} + \gamma' z_{0t}^{*} + e_t$$

the *positive size bias test statistic*:

$$\upsilon_t^2 = a + b_3 S_{t-1}^{+}\varepsilon_{t-1} + \gamma' z_{0t}^{*} + e_t$$

$$S_{t-1}^{+} = \begin{cases} 1 & \varepsilon_{t-1} > 0 \\ 0 & otherwise \end{cases}$$

*However, for highly nonlinear models, the numerical optimization algorithm generally does not guarantee exact orthogonality of υ_t^2 to z_{0t}^{*}. Engle and Ng ([16]) propose to regress y_t^2 on z_{0t} alone, and use the residuals from this regression (which are now guaranteed to be orthogonal to z_{0t}) in place of υ_t^2.

†In fact,

$$\sigma_t^2 = \sigma_{0t}^2\left(\delta_0' z_{0t}\right)\exp\left(\delta_a' z_{at}\right)$$

$$\frac{\partial \sigma_t^2}{\partial \delta_a} = \sigma_{0t}^2 z_{at}\exp\left(\delta_a' z_{at}\right)$$

under the null, $\delta_a = 0$, $\frac{\partial \sigma_t^2}{\partial \delta_a} = \sigma_{0t}^2 z_{at}$.

The t-ratios for b_1, b_2 and b_3 are the sign bias, the negative size bias, and the positive size bias test statistics, respectively. The joint test is the LM test for adding the three variables in the variance equation (2.6) under the maintained specification:

$$\upsilon_t^2 = a + b_1 S_{t-1}^- + b_2 S_{t-1}^- \varepsilon_{t-1} + b_3 S_{t-1}^+ \varepsilon_{t-1} + \gamma' z_{0t}^* + e_t$$

The test statistics is TR^2. If the volatility model is correct then $b_1 = b_2 = b_3 = 0$, $\gamma = 0$ and e_t is i.i.d. If z_{0t}^* is not included the test will be conservative; the size will be less than or equal to the nominal size, and the power may be reduced.

Chapter 3
MULTIVARIATE GARCH MODELS

3.1 Introduction

The extension from a univariate GARCH model to an N-variate model requires allowing the conditional variance-covariance matrix of the N-dimensional zero mean random variables ε_t depend on the elements of the information set.

Let $\{z_t\}$ be a sequence of $(N \times 1)$ i.i.d. random vector with the following characteristics:

$$E[z_t] = 0$$

$$E[z_t z_t'] = I_N$$

$$z_t \sim G(0, I_N)$$

with G continuous density function. Let $\{\varepsilon_t\}$ be a sequence of $(N \times 1)$ random vectors generated as:

$$\varepsilon_t = H_t^{1/2} z_t$$

where

$$E_{t-1}(\varepsilon_t) = 0$$

$$E_{t-1}(\varepsilon_t \varepsilon_t') = H_t$$

where H_t is a matrix $(N \times N)$ positive definite and measurable with respect to the information set Ψ_{t-1}, that is the σ-field generated by the past observations: $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$. The parametrization of H_t as a multivariate GARCH, which means as a function of the information set Ψ_{t-1}, allows each element of H_t to depend on q lagged of the squares and cross-products of ε_t, as well as p lagged values of the elements of H_t, and a $(J \times 1)$ vector of dummies. So the elements of the covariance matrix follow a vector of ARMA process in squares and cross-products of the disturbances.

3.2 Vech representation

Let *vech* denote the vector-half operator, which stacks the lower triangular elements of an $N \times N$ matrix as an $[N(N+1)/2] \times 1$ vector. Since the conditional covariance matrix H_t is symmetric, $vech(H_t)$ contains all the unique elements in H_t. Following Bollerslev et al. [6], a natural multivariate extension of the univariate GARCH(p,q) model is

$$\begin{aligned} vech(H_t) &= W + \sum_{i=1}^{q} A_i^* vech\left(\varepsilon_{t-i}\varepsilon_{t-i}'\right) + \sum_{j=1}^{p} B_j^* vech(H_{t-j}) \\ &= W + A^*(L)\, vech(\varepsilon_t \varepsilon_t') + B^*(L)\, vech(H_t) \end{aligned} \tag{3.1}$$

W is a $[N(N+1)/2] \times 1$ vector, the A_i^* and B_j^* are $[(N(N+1)/2) \times (N(N+1)/2)]$ matrices. This general formulation is termed *vec representation* by Engle and Kroner [15]. The number of parameters is $\left[N(N+1)/2 + (p+q)[N(N+1)/2]^2\right]$. Even for low dimensions of N and small values of p and q the number of parameters is very large; for $N = 5$ and $p = q = 1$ the unrestricted version of (3.1) contains 465 parameters.

For any parametrization to be sensible, we require that H_t be positive definite for all values of ε_t in the sample space in the *vech* representation this restriction can be difficult to check, let alone impose during estimation.

3.2.1 Diagonal vech model

A natural restriction that was first used in the ARCH context by Engle, Granger and Kraft [14] and in the GARCH context by Bollerslev et al [6] is the diagonal representation, in which each element of the covariance matrix depends only on past values of itself and past values of $\varepsilon_{jt}\varepsilon_{kt}$. In the diagonal model the A_i^* and B_j^* matrices are all taken to be diagonal. For $N = 2$ and $p = q = 1$, the diagonal model is written as:

$$\begin{aligned} \begin{bmatrix} h_{11,t} \\ h_{21,t} \\ h_{22,t} \end{bmatrix} &= \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} + \begin{bmatrix} a_{11}^* & 0 & 0 \\ 0 & a_{22}^* & 0 \\ 0 & 0 & a_{33}^* \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix} \\ &\quad + \begin{bmatrix} b_{11}^* & 0 & 0 \\ 0 & b_{22}^* & 0 \\ 0 & 0 & b_{33}^* \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{bmatrix} \end{aligned}$$

Thus the $(i,j)\,th$ element in H_t depends on the corresponding $(i,j)\,th$ element in $\varepsilon_{t-1}\varepsilon_{t-1}'$ and H_{t-1}. This restriction reduces the number of parameters to $[N(N+1)/2](1+p+q)$. This model does not allow for causality in variance, co-persistence in variance and asymmetries.

3.3 BEKK representation

Engle and Kroner ([15]) propose a parametrization that impose positive definiteness restrictions. Consider the following model

$$H_t = CC' + \sum_{k=1}^{K}\sum_{i=1}^{q} A_{ik}\varepsilon_{t-i}\varepsilon_{t-i}' A_{ik}' + \sum_{k=1}^{K}\sum_{i=1}^{p} B_{ik}H_{t-i}B_{ik}' \tag{3.2}$$

where C, A_{ik} and B_{ik}. The intercept matrix is decomposed into CC', where C is a lower triangular matrix. Without any further assumption CC' is positive semidefinite. This representation is general that it includes all positive definite diagonal representations and nearly all positive definite *vech* representations. For exposition simplicity we will assume that $K = 1$:

$$H_t = CC' + \sum_{i=1}^{q} A_i\varepsilon_{t-i}\varepsilon_{t-i}' A_i' + \sum_{i=1}^{p} B_i H_{t-i} B_i' \tag{3.3}$$

To illustrate the BEKK model, consider the simple GARCH(1,1) model:

$$H_t = CC' + A_1\varepsilon_{t-1}\varepsilon_{t-1}' A_1' + B_1 H_{t-1} B_1' \tag{3.4}$$

Proposition 10 *(Engle and Kroner [15]) Suppose that the diagonal elements in C are restricted to be positive and that a_{11} and b_{11} are also restricted to be positive. Then if $K = 1$ there exists no other C, A_1, B_1 in the model (3.4) that will give an equivalent representation.*

The purpose of the restrictions is to eliminate all other observationally equivalent structures. For example, as relates to the term $A_1\varepsilon_{t-1}\varepsilon_{t-1}'A_1'$ the only other observationally equivalent structure is obtained by replacing A_1 by $-A_1$. The restriction that a_{11} (b_{11}) be positive could be replaced with the condition that a_{ij} (b_{ij}) be positive for a given i and j, as this condition is also sufficient to eliminate $-A_1$ from the set of admissible structures.

In the bivariate case the BEKK becomes

$$\begin{aligned} H_t &= CC' + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1}^2 & \varepsilon_{1t-1}\varepsilon_{2t-1} \\ \varepsilon_{2t-1}\varepsilon_{1t-1} & \varepsilon_{2t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \\ &\quad + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} h_{11t-1} & h_{12t-1} \\ h_{21t-1} & h_{22t-1} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}' \end{aligned}$$

For what concerns the positive definiteness of H_t, we have the following result.

Proposition 11 *(Engle and Kroner [15]) (Sufficient condition) In a GARCH(p,q) model, if H_0, $H_{-1}, \ldots, H_{-p+1}$ are all positive definite, then the BEKK parametrization (with $K = 1$) yields a positive definite H_t for all possible values of ε_t if C is a full rank matrix or if any B_i $i = 1, \ldots, p$ is a full rank matrix (the intersection of the null spaces of C' and B_i' $i = 1, \ldots, p$ is null).*

Proof. For simplicity consider the GARCH(1,1) model. The BEKK parameterization is

$$H_t = CC' + +A_1\varepsilon_{t-1}\varepsilon_{t-1}'A_1' + B_1H_{t-1}B_1'$$

The proof proceeds by induction. First H_t is p.d. for $t = 1$: The term $A_1\varepsilon_0\varepsilon_0'A_1'$ is positive semidefinite because $\varepsilon_0\varepsilon_0'$ is positive semidefinite. Also if the null spaces of the matrices of C and B_1 intersect only at the origin, that is at least one of two is full rank then

$$CC' + B_1H_0B_1'$$

is positive definite. This is true if C or B_1 has full rank. To show that the null space condition is sufficient $CC' + B_1H_0B_1'$ is p.d. if and only if

$$x'\left(CC' + B_1H_0B_1'\right)x > 0 \qquad \forall x \neq 0$$

or

$$\left(C'x\right)'\left(C'x\right) + \left(H_0^{1/2}B_1'x\right)'\left(H_0^{1/2}B_1'x\right) > 0 \qquad \forall x \neq 0 \tag{3.5}$$

where $H_0 = H_0^{1/2\prime}H_0^{1/2}$ and $H_0^{1/2}$ is full rank. Defining $N(P)$ to be the null space of the matrix P, (3.5) is true if and only if

$$N\left(C'\right) \cap N\left(H_0^{1/2}B_1'\right) = \emptyset.$$

$N\left(H_0^{1/2}B_1'\right) = N\left(B_1'\right)$ because $H_0^{1/2}$ is full rank. This implies that $CC' + B_1H_0B_1'$ is positive definite if and only if $N\left(C'\right) \cap N\left(H_0^{1/2}B_1'\right) = \emptyset$. Now suppose that H_t is positive definite for $t = \tau$. Then

$$H_{\tau+1} = CC' + A_1\varepsilon_\tau\varepsilon_\tau'A_1' + B_1H_\tau B_1'$$

is positive definite if and only if, given that $A_1\varepsilon_\tau\varepsilon_\tau'A_1'$ is positive semidefinite, the null space condition holds, because H_τ is positive definite by the induction assumption. ■

We now examine the relationship between the BEKK and *vech* parametrizations. The mathematical relationship between the parameters of the two models can be found simply vectorizing the equation (3.3):

$$vec(H_t) = vec(CC') + \sum_{i=1}^{q} vec(A_i\varepsilon_{t-i}\varepsilon_{t-i}'A_i') + \sum_{i=1}^{p} vec(B_iH_{t-i}B_i')$$

where $vec\,()$ is an operator such that given a matrix A $(n \times n)$, $vec(A)$ is a $(n^2 \times 1)$ vector. The $vec\,()$ satisfies

$$vec\,(ABC) = (C' \otimes A)\,vec\,(B)$$

then

$$\begin{aligned} vec(H_t) &= vec(CC') + \sum_{i=1}^{q} (A_i \otimes A_i)\,vec(\varepsilon_{t-i}\varepsilon'_{t-i}) \\ &+ \sum_{i=1}^{p} (B_i \otimes B_i)\,vec(H_{t-i}) \end{aligned}$$

For A $(n \times n)$ symmetric, then $vech\,(A)$ contains precisely the $n\,(n+1)\,/2$ distinct elements of A and the elements of $vec\,(A)$ are those of $vech\,(A)$ with some repetitions. Hence there exists a unique $n^2 \times n\,(n+1)\,/2$ which transforms, for symmetric A, $vech\,(A)$ into $vec\,(A)$. This matrix is called the *duplication matrix* and is denoted D_n:

$$vec\,(A) = D_n vech\,(A)$$

where D_n is the duplication matrix.

$$\begin{aligned} D_N vech\,(H_t) &= D_N vech(CC') + \sum_{i=1}^{q} (A_i \otimes A_i)\,D_N vech(\varepsilon_{t-i}\varepsilon'_{t-i}) \\ &+ \sum_{i=1}^{p} (B_i \otimes B_i)\,D_N vech(H_{t-i}) \end{aligned}$$

If D_N is a full column rank matrix we can define the generalized inverse of D_N as:

$$D_N^+ = (D'_N D_N)^{-1}\,D'_N$$

that is a $(N\,(N+1)\,/2) \times (N^2)$ matrix, where

$$D_N^+ D_N = I_N$$

This implies that premultiplying by D_N^+

$$\begin{aligned} vech\,(H_t) &= vech(CC') + D_N^+ \left(\sum_{i=1}^{q} (A_i \otimes A_i) \right) D_N vech(\varepsilon_{t-i}\varepsilon'_{t-i}) \\ &+ D_N^+ \left(\sum_{i=1}^{p} (B_i \otimes B_i) \right) D_N vech(H_{t-i}) \end{aligned}$$

One implication of this result is that the *vech* model implied by any given BEKK model is unique, while the converse is not true. The transformation from a *vech* model to a BEKK model (when it exists) is not unique, because for a given A_1^* the choice of A_1 is not unique. This can be seen recognizing that $(A_i \otimes A_i) = (-A_i \otimes -A_i)$ so while $A_i^* = D_N^+ (A_i \otimes A_i) D_N$ is unique, the choice of A_i is not unique. It can also be shown that all positive definite diagonal *vech* models can be written in the BEKK framework.

Given A_i diagonal matrix, then $D_N^+ (A_i \otimes A_i) D_N$ is also diagonal, with diagonal elements given by $a_{ii}a_{jj}$ $(1 \leq j \leq i \leq N)$ (See Magnus [22]).

3.3.1 Covariance Stationarity

Given the *vech* model

$$vech\,(H_t) = W + A^*\,(L)\,vech\,(\varepsilon_t\varepsilon_t') + B^*\,(L)\,vech\,(H_t)$$

the necessary and sufficient condition for covariance stationary of $\{\varepsilon_t\}$ is that all the eigenvalues of $A^*(1) + B^*(1)$ are less than one in modulus. But defining $A^*(1) = D_N^+ \left(\sum_{i=1}^{q} (A_i \otimes A_i)\right) D_N$ and $B^*(1) = D_N^+ \left(\sum_{i=1}^{q} (B_i \otimes B_i)\right) D_N$. This implies also that in the BEKK model, $\{\varepsilon_t\}$ is covariance stationary if and only if all the eigenvalues of $D_N^+ \left(\sum_{i=1}^{q} (A_i \otimes A_i)\right) D_N + D_N^+ \left(\sum_{i=1}^{q} (B_i \otimes B_i)\right) D_N$ are less than one in modulus. Let $\lambda_1, \ldots, \lambda_N$ the eigenvalues of A_i, the eigenvalues of $D_N^+ \left(\sum_{i=1}^{q} (A_i \otimes A_i)\right) D_N$ are $\lambda_i\lambda_j$ $(1 \leq j \leq i \leq N)$. (Magnus, [22])

For a GARCH(p,q) in *vech* form, the unconditional covariance matrix, when it exists, is given by*

$$E\,(vech(\varepsilon_t\varepsilon_t')) = vech\,(E\,(\varepsilon_t\varepsilon_t')) = \left[I_{N^*} - A^*\,(1) - B^*\,(1)\right]^{-1} W$$

and in the BEKK model†

$$vech\,(E\,(\varepsilon_t\varepsilon_t')) = \left[I_{N^*} - D_N^+\,(A_1 \otimes A_1)\,D_N - D_N^+\,(B_1 \otimes B_1)\,D_N\right]^{-1} vech\,(CC')$$

---

*Given that

$$vech\,(\varepsilon_t\varepsilon_t') = vech\,(H_t) + vech\,(V_t)$$

with $E\,(vech\,(V_t)) = vech(E\,(V_t)) = 0$

$$vech\,(\varepsilon_t\varepsilon_t') = W + A_1^* vech\left(\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right)\right) + B_1^*\left[vech\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) - vech\,(V_{t-1})\right]$$

$$vech\,(E\,(\varepsilon_t\varepsilon_t')) = W + A_1^* vech\left(E\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right)\right) + B_1^* vech\left(E\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right)\right)$$

†Given that

$$\varepsilon_t\varepsilon_t' = H_t + V_t$$

$N^* = N(N+1)/2$. The diagonal *vech* model is stationary if and only if the sum $a^*_{ii} + b^*_{ii} < 1$ for all i. In the diagonal BEKK model the covariance stationary condition is that $a^2_{ii} + b^2_{ii} < 1$. Only in the case of diagonal models the stationarity properties are determined solely by the diagonal elements of the A_i and B_i matrices.

3.4 Constant Correlations Model

In the constant correlations model put forward by Bollerslev ([3]) the time-varying conditional covariances are parametrized to be proportional to the product of the corresponding conditional standard deviations. This assumption greatly simplifies the computational burden in estimation, and conditions for H_t to be positive definite a.s for all t are easy to impose. The model assumptions are:

$$E_{t-1}[\varepsilon_t] = 0$$

$$E_{t-1}[\varepsilon_t \varepsilon_t'] = H_t$$

$$\{H_t\}_{ii} = \sigma^2_{it}$$

$$\{H_t\}_{ij} = \sigma_{ijt} = \rho_{ij}\sigma_{it}\sigma_{jt} \qquad i \neq j$$

Let D_t denote the $(N \times N)$ diagonal matrix with the conditional variances along the diagonal, $\{D_t\}_{ii} = \sigma^2_{it}$. Let Γ_t denote the matrix of constant correlations with $ij-th$ element given by

$$\{\Gamma_t\}_{ij} = \{H_t\}_{ij}\left[\{H_t\}_{ii}\{H_t\}_{jj}\right]^{-1/2} \qquad i, j = 1, \ldots, N$$

the model assumes $\Gamma_t = \Gamma$

$$H_t = D_t^{1/2}\Gamma D_t^{1/2}$$

---

with $E(V_t) = 0$

$$vec(\varepsilon_t\varepsilon_t') = vec(CC') + (A_1 \otimes A_1)\, vec\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) + (B_1 \otimes B_1)\left[vec\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) - vec(V_{t-1})\right]$$

$$E\left[vec(\varepsilon_t\varepsilon_t')\right] = vec(CC') + \left[(A_1 \otimes A_1) + (B_1 \otimes B_1)\right] E\left[vec\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right)\right]$$

$$E\left[vec(\varepsilon_t\varepsilon_t')\right] = \left[I_{N^2} - (A_1 \otimes A_1) - (B_1 \otimes B_1)\right]^{-1} vec(CC')$$

or in *vech* representation as

$$D_N vech(E(\varepsilon_t\varepsilon_t')) = D_N vech(CC') + (A_1 \otimes A_1)\, D_N vech\left(E\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right)\right) + (B_1 \otimes B_1)\, D_N vech\left(E\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right)\right)$$

$$H_t = \begin{bmatrix} \sigma_{1t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{Nt} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & & \rho_{1N} \\ \rho_{21} & 1 & & \vdots \\ \vdots & \vdots & \ddots & \rho_{N-1N} \\ \rho_{N1} & & \rho_{NN-1} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{Nt} \end{bmatrix}.$$

When $N = 2$

$$\begin{aligned} H_t &= \begin{bmatrix} \sigma_{1t} & 0 \\ 0 & \sigma_{2t} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1t} & 0 \\ 0 & \sigma_{2t} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{1t}^2 & \rho_{12}\sigma_{1t}\sigma_{2t} \\ \rho_{12}\sigma_{1t}\sigma_{2t} & \sigma_{2t}^2 \end{bmatrix}. \end{aligned}$$

If the conditional variances along the diagonal in the D_t matrices are all positive, and the conditional correlation matrix Γ is positive definite, the sequence of conditional covariance matrices $\{H_t\}$ is guaranteed to be positive definite a.s. for all t. Furthermore the inverse of H_t is given by

$$H_t^{-1} = D_t^{-1/2}\Gamma^{-1}D_t^{-1/2}.$$

When calculating the log-likelihood function only one matrix inversion is required for each evaluation.

3.5 Factor ARCH model

The Factor GARCH model, introduced by Engle et al. ([17]), can be thought of as an alternative simple parametrization of the BEKK model. Suppose that the $(N \times 1)$ y_t has a factor structure with K factors given by the $K \times 1$ vector f_t and a time invariant factor loadings given by the $N \times K$ matrix B:

$$y_t = Bf_t + \varepsilon_t \tag{3.6}$$

Assume that the idiosyncratic shocks ε_t have conditional covariance matrix Ψ which is constant in time and positive semidefinite, and that the common factors are characterized by

$$E_{t-1}(f_t) = 0$$

$$E_{t-1}(f_t f_t') = \Lambda_t$$

$\Lambda_t = diag(\lambda_1, \ldots, \lambda_K)$ and positive definite. The conditioning set is $\{y_{t-1}, f_{t-1}, \ldots, y_1, f_1\}$. Also suppose that $E(f_t\varepsilon_t') = 0$. The conditional covariance matrix of y_t equals

$$E_{t-1}(y_t y_t') = H_t = \Psi + B\Lambda_t B' = \Psi + \sum_{k=1}^{K} \beta_k \beta_k' \lambda_{kt}$$

where β_k denotes the kth column in B. Thus, there are K statistics which determine the full covariance matrix. Forecasts of the variances and covariances or of any portfolio of assets, will be based only on the forecasts of these K statistics.

There exists *factor-representing portfolios* with portfolio weights that are orthogonal to all but one set of factor loadings:

$$r_{kt} = \phi_k' y_t$$

$$\phi_k' \beta_j = \begin{cases} 1 & k = j \\ 0 & otherwise \end{cases}$$

the vector of factor-representing portfolios is

$$r_t = \Phi' y_t$$

where the columns of matrix Φ are the ϕ_k vectors. The conditional variance of r_{kt} is given by

$$\begin{aligned} Var_{t-1}(r_{kt}) &= \phi_k' E_{t-1}(y_t y_t') \phi_k = \phi_k' H_t \phi_k \\ &= \phi_k' (\Psi + B \Lambda_t B') \phi_k \\ &= \psi_k + \lambda_{kt} \end{aligned}$$

where $\psi_k = \phi_k' \Psi \phi_k$. The portfolio has the exact time variation as the factors, which is why they are called factor-representing portfolios. In order to estimate this model, the dependence of the λ_{kt}'s upon the past information set must also be parametrized:

$$\theta_{kt} \equiv \phi_k' H_t \phi_k = Var_{t-1}(r_{kt}) = \psi_k + \lambda_{kt}$$

So we get that

$$\sum_{k=1}^{K} \beta_k \beta_k' \theta_{kt} = \sum_{k=1}^{K} \beta_k \beta_k' \psi_k + \sum_{k=1}^{K} \beta_k \beta_k' \lambda_{kt}$$

$$\sum_{k=1}^{K} \beta_k \beta_k' \lambda_{kt} = \sum_{k=1}^{K} \beta_k \beta_k' \theta_{kt} - \sum_{k=1}^{K} \beta_k \beta_k' \psi_k$$

$$\begin{aligned} H_t &= \Psi + \sum_{k=1}^{K} \beta_k \beta_k' \lambda_{kt} = \Psi + \sum_{k=1}^{K} \beta_k \beta_k' \theta_{kt} - \sum_{k=1}^{K} \beta_k \beta_k' \psi_k \\ &= \Psi^* + \sum_{k=1}^{K} \beta_k \beta_k' \theta_{kt} \end{aligned}$$

where $\Psi^* = \left(\Psi - \sum_{k=1}^{K} \beta_k \beta_k' \psi_k\right)$. The simplest assumption is that there is a set of factor-representing portfolios with univariate GARCH(1,1) representations. The conditional variance θ_{kt} follows a GARCH(1,1) process

$$\theta_{kt} = \omega_k + \alpha_k \left(\phi_k' \varepsilon_{t-1}\right)^2 + \gamma_k E_{t-2}\left(r_{kt-1}^2\right)$$

$$\theta_{kt} = \omega_k + \alpha_k \phi_k' \left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) \phi_k + \gamma_k E_{t-2}\left[\left(\phi_k' y_t\right)\left(\phi_k' y_t\right)\right]$$

$$\theta_{kt} = \omega_k + \alpha_k \phi_k' \left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) \phi_k + \gamma_k \left[\phi_k' E_{t-2}\left(y_t y_t'\right) \phi_k\right]$$

$$\theta_{kt} = \omega_k + \alpha_k \phi_k' \left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) \phi_k + \gamma_k \left[\phi_k' H_{t-1} \phi_k\right]$$

(A lezione si era assunto $\omega_k = \psi_k$) The conditional variance-covariance matrix of y_t can be written as

$$\begin{aligned} H_t &= \Psi^* + \sum_{k=1}^{K} \beta_k \beta_k' \theta_{kt} \\ &= \Psi^* + \sum_{k=1}^{K} \beta_k \beta_k' \left\{\omega_k + \alpha_k \left[\phi_k' \left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) \phi_k\right] + \gamma_k \left[\phi_k' H_{t-1} \phi_k\right]\right\} \\ &= \left(\Psi^* + \sum_{k=1}^{K} \beta_k \beta_k' \omega_k\right) + \sum_{k=1}^{K} \beta_k \beta_k' \left\{\alpha_k \left[\phi_k' \left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) \phi_k\right] + \gamma_k \left[\phi_k' H_{t-1} \phi_k\right]\right\} \end{aligned}$$

$$H_t = \Gamma + \sum_{k=1}^{K} \beta_k \beta_k' \theta_{kt}$$

where $\Gamma = \Psi^* + \sum_{k=1}^{K} \beta_k \beta_k' \omega_k$, therefore

$$H_t = \Gamma + \sum_{k=1}^{K} \alpha_k \left[\beta_k \phi_k' \left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) \phi_k \beta_k'\right] + \sum_{k=1}^{K} \gamma_k \left[\beta_k \phi_k' H_{t-1} \phi_k \beta_k'\right]$$

so that the factor GARCH model is a special case of the BEKK parametrization. Estimation of the factor GARCH model is carried out by maximum likelihhod estimation. It is often convenient to assume that the factor-representing portfolios are known a priori.

3.6 Asymmetric Multivariate GARCH-in-mean model

A general multivariate model can be written as:

$$y_t = \mu + \Pi\left(L\right) y_{t-1} + \Phi x_{t-1} + \Lambda vech\left(H_t\right) + \epsilon_t \tag{3.7}$$

where y_t is a $(N \times 1)$ vector of weakly stationary variables (that is, asset returns), $\Pi\left(L\right) = \Pi_1 + \Pi_2 L + \cdots + \Pi_k L^{k-1}$, x_{t-1} contains predetermined variables. ϵ_t is the vector of innovation with respect to the information set formed exclusively of past realizations of y_t. Λ is a $\left(N \times N\left(N+1\right)/2\right)$:

$$H_t = E_{t-1}\left(\epsilon_t \epsilon_t'\right)$$

$$H_t = CC' + \sum_{i=1}^{q} A_i \left(\varepsilon_{t-i} + \gamma\right)\left(\varepsilon_{t-i} + \gamma\right)' A_i' + \sum_{j=1}^{p} B_j H_{t-j} B_j' \tag{3.8}$$

We can consider a multivariate generalization of the size effect and sign effect:

$$H_t = CC' + A_1 \varepsilon_{t-1}\varepsilon_{t-1}' A_1' + B_1 H_{t-1} B_1' + D v_{t-1} v_{t-1}' D' + G_1^* \varepsilon_{t-1}\varepsilon_{t-1}' G^{*\prime}$$

where $v_t = \left|z_t\right| - E\left|z_t\right|$, with $z_{it} = \varepsilon_{it}/\sqrt{h_{ii,t}}$ and

$$G^* = \begin{bmatrix} I\left(\varepsilon_{1t-1} < 0\right) g_{11} & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & & 0 & I\left(\varepsilon_{Nt-1} < 0\right) g_{NN} \end{bmatrix}$$

When $N = 2$

$$\begin{aligned} v_{t-1} v_{t-1}' &= \begin{bmatrix} \left|\varepsilon_{1t-1}/\sqrt{h_{11,t-1}}\right| - E\left|\varepsilon_{1t-1}/\sqrt{h_{11,t-1}}\right| \\ \left|\varepsilon_{2t-1}/\sqrt{h_{22,t-1}}\right| - E\left|\varepsilon_{2t-1}/\sqrt{h_{22,t-1}}\right| \end{bmatrix} \begin{bmatrix} \left|\varepsilon_{1t-1}/\sqrt{h_{11,t-1}}\right| - E\left|\varepsilon_{1t-1}/\sqrt{h_{11,t-1}}\right| \\ \left|\varepsilon_{2t-1}/\sqrt{h_{22,t-1}}\right| - E\left|\varepsilon_{2t-1}/\sqrt{h_{22,t-1}}\right| \end{bmatrix}' \\ &= \begin{bmatrix} \left(\left|z_{1t}\right| - E\left|z_{1t}\right|\right)^2 & \left(\left|z_{1t}\right| - E\left|z_{1t}\right|\right)\left(\left|z_{2t}\right| - E\left|z_{2t}\right|\right) \\ \left(\left|z_{2t}\right| - E\left|z_{2t}\right|\right)\left(\left|z_{1t}\right| - E\left|z_{1t}\right|\right) & \left(\left|z_{2t}\right| - E\left|z_{2t}\right|\right)^2 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} G_1^* \varepsilon_{t-1}\varepsilon_{t-1}' G^{*\prime} &= \begin{bmatrix} g_{11}^{*2}\varepsilon_{1t-1}^2 & g_{11}^* g_{22}^* \varepsilon_{1t-1}\varepsilon_{2t-1} \\ g_{11}^* g_{22}^* \varepsilon_{1t-1}\varepsilon_{2t-1} & g_{22}^{*2}\varepsilon_{2t-1}^2 \end{bmatrix} \\ &= \begin{bmatrix} I\left(\varepsilon_{1t-1} < 0\right) g_{11}^2 \varepsilon_{1t-1}^2 & \delta_{12} g_{11} g_{22} \varepsilon_{1t-1}\varepsilon_{2t-1} \\ \delta_{12} g_{11} g_{22} \varepsilon_{1t-1}\varepsilon_{2t-1} & I\left(\varepsilon_{2t-1} < 0\right) g_{22}^2 \varepsilon_{2t-1}^2 \end{bmatrix} \end{aligned}$$

$$\delta_{12} = I\left(\varepsilon_{1t-1} < 0\right) I\left(\varepsilon_{2t-1} < 0\right)$$

3.7 Estimation procedure

Given the model (3.7)-(3.8), the log-likelihood function for $\{\varepsilon_T, \dots, \varepsilon_1\}$ obtained under the assumption of conditional multivariate normality is:

$$L_T(\varepsilon_T, \dots, \varepsilon_1; \theta) = -\frac{1}{2}\left[TN\ln(2\pi) + \sum_{t=1}^{T}\left(\ln|H_t| + \varepsilon_t' H_t^{-1}\varepsilon_t\right)\right]$$

The function corresponds directly to the conditional likelihood function for the univariate ARCH model, used in maximum likelihood or quasi-maximum likelihood estimation. Because maximum likelihood under normality is so widely used, it is important to investigate its properties in a general setting. In general, the assumption of conditional normality can be quite resctrictive. The symmetry imposed under normality is difficult to justify, and the tails of even conditional distributions often seem fatter than that of normal distribution.

Let $\{(y_t, x_t) : t = 1, 2, \dots\}$ be a sequence of observables random vectors with y_t $(N \times 1)$ and x_t $(L \times 1)$. The vector y_t contains the "endogenous" variables and x_t contains contemporaneous "exogenous" variables. Let $w_t = (x_t, y_{t-1}, x_{t-1}, \dots, y_1, x_1)$.

The conditional mean and variance functions are jointly parametrized by a finite dimensional vector θ:

$$\{\mu_t(w_t, \theta), \theta \in \Theta\}$$

$$\{H_t(w_t, \theta), \theta \in \Theta\}$$

where Θ is a subset of R^P and μ_t and H_t are known functions of w_t and θ.

In the analysis, the validity of most of the inference procedures is proven under the null hypothesis that the first two conditional moments are correctly specified, for some $\theta_0 \in \Theta$,

$$E(y_t | w_t) = \mu_t(w_t, \theta_0) \tag{3.9}$$

$$Var(y_t | w_t) = H_t(w_t, \theta_0) \quad t = 1, 2, \dots \tag{3.10}$$

The procedure most often used to estimate θ_0 is maximization of a likelihood function that is constructed under the assumption that $y_t | w_t \sim N(\mu_t, H_t)$. The approach taken here is the same, but the subsequent analysis does not assume that y_t has a conditional normal distribution.

For observation t the quasi-conditional log-likelihood is

$$l_t(\theta; y_t, w_t) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|H_t(w_t, \theta)| - \frac{1}{2}(y_t - \mu_t(w_t, \theta))' H_t^{-1}(w_t, \theta)(y_t - \mu_t(w_t, \theta))$$

Letting $\varepsilon_t(y_t, w_t, \theta_0) \equiv y_t - \mu_t(w_t, \theta)$ denote the $N \times 1$ residual function, and in amore concise notation

$$l_t(\theta) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|H_t(\theta)| - \frac{1}{2}\varepsilon_t'(\theta) H_t^{-1}(\theta)\varepsilon_t(\theta) \tag{3.11}$$

$$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta)$$

If $\mu_t(w_t, \theta)$ and $H_t(w_t, \theta)$ are differentiable on Θ for all relevant w_t, and if $H_t(w_t, \theta)$ is nonsingular with probability one for all $\theta \in \Theta$, then the differentiation of (3.11) yields the $(1 \times P)$ score function $s_t(\theta)$:

$$\begin{aligned} s_t(\theta)' &= \nabla_\theta l_t(\theta)' - \nabla_\theta \mu_t(\theta)' H_t^{-1}(\theta)\varepsilon_t(\theta) + \\ &\quad \frac{1}{2}\nabla_\theta H_t(\theta)' \left[H_t^{-1}(\theta) \otimes H_t^{-1}(\theta)\right] vec\left[\varepsilon_t(\theta)\varepsilon_t(\theta)' - H_t(\theta)\right] \end{aligned}$$

where $\nabla_\theta \mu_t(\theta)$ is the $(N \times P)$ derivative of $\mu_t(w_t, \theta)$ and $\nabla_\theta H_t(\theta)$ is the $(N^2 \times P)$ derivative of $H_t(\theta)$. If the first conditional two moments are correctly specified, that is if the (3.9) holds then the true error vector is defined as $\varepsilon_t^0 \equiv \varepsilon_t(\theta_0) = y_t - \mu_t(w_t, \theta_0)$ and $E(\varepsilon_t^0 | w_t) = 0$. If in addition, (3.10) holds then $E(\varepsilon_t^0 \varepsilon_t^{0\prime} | w_t) = H_t(w_t, \theta_0)$. It follows that under correct specification of the first two conditional moments of y_t given w_t:

$$E[s_t(\theta_0) | w_t] = 0$$

The score evaluated at the true parameter is a vector of martingale difference with respect to the $\sigma - fields$ $\{\sigma(y_t, w_t) : t = 1, 2, \ldots\}$. This result can be used to establish weak consistency of the quasi-maximum likelihood estimator (QMLE).

For robust inference we also need an expression for the hessian $h_t(\theta)$ of $l_t(\theta)$. Define the $(P \times P)$ positive semidefinite matrix $a_t(\theta_0) = -E[\nabla_\theta s_t(\theta_0) | w_t] = E[-h_t(\theta_0) | w_t]$. A straightforward calculation shows that, under (3.9) and (3.10),

$$\begin{aligned} a_t(\theta_0) &= \nabla_\theta \mu_t(\theta_0)' H_t^{-1}(\theta_0) \nabla_\theta \mu_t(\theta_0) \\ &\quad + \frac{1}{2}\nabla_\theta H_t(\theta)' \left[H_t^{-1}(\theta) \otimes H_t^{-1}(\theta)\right] \nabla_\theta H_t(\theta) \end{aligned}$$

When the normality assumption holds the matrix $a_t(\theta_0)$ is the conditional information matrix. However, if y_t does not have a conditional normal distribution then $Var[s_t(\theta_0) | w_t]$ is generally not equal to $a_t(\theta_0)$ and the information matrix equality is violated.

The QMLE has the following properties:

$$\left[A_T^{0-1} B_T^0 A_T^{0-1}\right]^{-1/2} \sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{d} N(0, I_P)$$

where

$$A_T^0 \equiv -\frac{1}{T}\sum_{t=1}^{T} E\left[h_t\left(\theta_0\right)\right] = \frac{1}{T}\sum_{t=1}^{T} E\left[a_t\left(\theta_0\right)\right]$$

and

$$B_T^0 \equiv Var\left[T^{-1/2} S_T\left(\theta_0\right)\right] = \frac{1}{T}\sum_{t=1}^{T} E\left[s_t\left(\theta_0\right)' s_t\left(\theta_0\right)\right]$$

in addition

$$\widehat{A}_T - A_T^0 \xrightarrow{p} 0$$

$$\widehat{B}_T - B_T^0 \xrightarrow{p} 0$$

The matrix $\widehat{A}_T^{-1}\widehat{B}_T\widehat{A}_T^{-1}$ is a consistent estimator od the robust asymptotic covariance matrix of $\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right)$. In practice, one treats $\widehat{\theta}_T$ as if it is normally distributed with mean θ_0 and variance $\widehat{A}_T^{-1}\widehat{B}_T\widehat{A}_T^{-1}/T$. Under normality, the variance estimator can be replaced by $\widehat{A}_T^{-1}/T$ (Hessian form) or $\widehat{B}_T^{-1}/T$ (outer product of the gradient form).

We can derive a robust form for Wald statistics for testing hypotheses about θ_0. Assume that the null hypothesis can be stated as

$$H_0 : r\left(\theta_0\right) = 0$$

where $r : \Theta \rightarrow \Re^Q$ is continuously differentiable on $int\left(\Theta\right)$ and $Q < P$. Let $R\left(\theta\right) = \nabla_\theta r\left(\theta\right)$ be the $\left(Q \times P\right)$ gradient of r on $int\left(\Theta\right)$. If $\theta_0 \in int\left(\Theta\right)$ and $rank\left(R\left(\theta_0\right)\right) = Q$ then the Wald statistic

$$\xi_W = Tr\left(\widehat{\theta}_T\right)' \left[R\left(\widehat{\theta}_T\right)\widehat{A}_T^{-1}\widehat{B}_T\widehat{A}_T^{-1}R\left(\widehat{\theta}_T\right)'\right]^{-1} r\left(\widehat{\theta}_T\right) \underset{H_0}{\xrightarrow{d}} \chi_Q^2.$$

Chapter 4
REFERENCES

1. Bera A.K., M.L. Higgins (1993) "ARCH Models: properties, Estimation and Testing", *Journal of Economic Surveys*, vol.7,n.4, 305-62.
2. Bollerslev T., (1986) "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*,31,307-327.
3. Bollerlsev T. (1990) "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach" *Review of Economics and Statistics*, 72, 498-505.
4. Bollerslev T:,R.F Engle.(1993) "Common Persistence in Conditional Variances", *Econometrica*,61,166-187.
5. Bollerslev T., R.F. Engle, D.B. Nelson (1994) "ARCH Models" in *Handbook of Econometrics*, vol. IV, edited by R.F.Engle and D.L. McFadden. Elsevier Science.
6. Bollerslev T., R.F. Engle, J. Wooldridge (1988) "A Capital Asset Pricing Model with Time Varying Covariances", *Journal of Political Economy*, 96, 116-131.
7. Bollerslev T., J. Wooldridge (1992) "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances", *Econometric Reviews*, 11(2), 143-172.
8. Ding Z., C.W.J.Granger and R.Engle (1993) "A long memory property of stock market returns and a new model", *Journal of Empirical Finance*, 1, 83-106.
9. Engle R.F. (1982) "Autoregressive Conditional Heteroskedasticity with estimates of the Variance of U.K. Inflation", *Econometrica,* 50, 987-1008.
10. Engle R.F. (1990) "Discussion: Stock Market Volatility and the Crash of 87" *Review of Financial Studies*,3,103-106.
11. Engle R.F., T.Bollerslev (1986) "Modeling the persistence of conditional variances" *Econometric Review*,5,1-50.
12. Engle R. G.J.Lee (1999) "A Permanent and Transitory Component Model of Stock Return Volatility", in ed. R. Engle and H. White *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, Oxford University Press, 475-497.
13. Engle R.F., Lilien D.M. and Robins R.P. (1987) "Estimating time varying risk premia in the term structure: the ARCH-M model", *Econometrica*, 55, 391-407.
14. Engle R.F., C.W.J. Granger, D.F. Kraft (1984) "Combining competing forecasts of

inflation using a bivariate ARCH model" *Journal of Economic Dynamics and Control*, 8, 151-165.

15. Engle R.F. K.F. Kroner (1995)"Multivariate Simultaneous Generalized ARCH" *Econometric Theory*, 11, 122-150.
16. Engle R.F., V.K.Ng (1993) "Measuring and Testing the Impact of News on Volatility", *Journal of Finance*,48,1749-1778.
17. Engle R.F., V.K.Ng, M.Rothschild (1990) "Asset pricing with a factor-ARCH covariance structure", *Journal of Econometrics*, 213-237.
18. Engle R.F., G. Gonzalez-Rivera (1991) "Semiparametric ARCH Models", *Journal of Business and Economic Statistics*, 19, 3-29.
19. Glosten L., R.Jagannathan, D.Runkle (1993) "On the Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks" *Journal of Finance*,48,1779-1801.
20. Gourieroux C. (1992) *Modeles ARCH et applications financières*, Economica: Paris.
21. Hamilton J. (1994) *Time Series Analysis*, Princenton University Press.
22. Magnus J.R. (1988) *Linear Structures*, Charles Griffin & Company Lim., London.
23. Nelson D.B. (1990) "Stationarity and Persistence in the GARCH(1,1) Model", *Econometric Theory, 6, 318-334.*
24. Nelson D.(1991) "Conditional Heteroskedasticity in asset returns: A new approach", *Econometrica*, 59,347-370.
25. Nelson D., C.Q.Cao (1992) "Inequality Constraints in the Univariate GARCH Model" *Journal of Business and Economic Statistics,*10,229-235.
26. Pagan A.(1996)"The Econometrics of financial markets" *Journal of Empirical Finance*,3,15-102.
27. Robinson P. (1991)
28. Sentana E. (1995) "Quadratic ARCH Models" *Review of Economic Studies,* 62(4), 639-61.
29. Taylor S. (1986) *Modeling Financial Time Series*, Wiley and Sons: New York, NY.