The XlQuant Team

Estimating and Forecasting GARCH Models using

XlQuant

Version 1 - (2020)

Contents

Front matter **i**

1 Univariate GARCH Models **1**
- 1.1 Conditional Mean Specification 1
- 1.2 Conditional Variance Specification 1
  - 1.2.1 GARCH Model 1
  - 1.2.2 EGARCH Model 2
  - 1.2.3 GJR Model 2
  - 1.2.4 APARCH Model 3
  - 1.2.5 IGARCH Model 3
  - 1.2.6 RiskMetrics^{TM} 4
  - 1.2.7 Fractionally Integrated Models 4
  - 1.2.8 Generalized Autoregressive Score (GAS) Models 4
  - 1.2.9 Forecasting the Conditional Variance of GARCH-type models 6
- 1.3 Estimation 7
- 1.4 Value-at-Risk (VaR) estimation using XlQuant 9

2 Multivariate GARCH Models **11**
- 2.1 Conditional mean specification 12
- 2.2 Multivariate GARCH specifications 12
  - 2.2.1 RiskMetrics 13
  - 2.2.2 BEKK models 13
  - 2.2.3 Conditional correlation models 15
- 2.3 Estimation 16
  - 2.3.1 Maximum Likelihood 17
  - 2.3.2 Two-step estimation 19
  - 2.3.3 Variance Targeting 20

Chapter 1

Univariate GARCH Models

1.1 Conditional Mean Specification

Let us consider a univariate time series y_t. If Ω_{t-1} is the information set at time $t-1$, we can define its functional form as:

$$y_t = \mu_t + \varepsilon_t, \tag{1.1}$$

where $\mu_t = E(y_t|\Omega_{t-1})$ is the conditional mean of y_t and ε_t is the disturbance term (or unpredictable part), with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_s) = 0, \forall\, t \neq s$.

Two of the most famous specifications of the conditional mean are the Autoregressive (AR) and Moving Average (MA) models. Combining these two processes, we obtain the ARMA(n, s) process

$$\Psi(L)(y_t - \mu) = \Theta(L)\varepsilon_t, \tag{1.2}$$

where L is the lag operator so that $L^k y_t = y_{t-k}$ and therefore $\Psi(L) = 1 - \sum_{i=1}^{n} \psi_i L^i$ and $\Theta(L) = 1 + \sum_{j=1}^{s} \theta_j L^j$.

1.2 Conditional Variance Specification

Estimating (1.1) by Ordinary Least Square (OLS) would require assuming the variance of the error term to be constant over time, i.e., $\varepsilon_t = \sigma z_t$, where z_t is an i.i.d. process with mean 0 and unit variance. This assumption in not realistic for most financial time series observed at the daily or weekly frequency. In the next sections, we discuss several specifications allowing to introduce dynamics in the conditional variance (denoted σ_t^2).

1.2.1 GARCH Model

The Generalized ARCH (GARCH) model of Bollerslev (1986) is based on an infinite ARCH specification and it allows to reduce the number of estimated parameters by imposing nonlinear restrictions on them. The GARCH (p, q) model can be expressed as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2. \tag{1.3}$$

Using the lag (or backshift) operator L, the GARCH (p, q) model becomes:

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2,$$

with $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_q L^q$ and $\beta(L) = \beta_1 L + \beta_2 L^2 + \ldots + \beta_p L^p$.

1.2.2 EGARCH Model

The Exponential GARCH (EGARCH) model, originally introduced by Nelson (1991), is re-expressed in Bollerslev and Mikkelsen (1996) as follows:

$$\log \sigma_t^2 = \omega + [1 - \beta(L)]^{-1} [1 + \alpha(L)] g(z_{t-1}). \tag{1.4}$$

The value of $g(z_t)$ depends on several elements. Nelson (1991) notes that, *"to accommodate the asymmetric relation between stock returns and volatility changes (...) the value of $g(z_t)$ must be a function of both the magnitude and the sign of z_t"*. That is why he suggests to express the function $g(.)$ as

$$g(z_t) \equiv \underbrace{\gamma_1 z_t}_{\text{sign effect}} + \underbrace{\gamma_2 [|z_t| - E|z_t|]}_{\text{magnitude effect}} \tag{1.5}$$

$E|z_t|$ depends on the assumption made on the unconditional density of z_t. $E(|z_t|) = \sqrt{2/\pi}$ for the normal distribution, $E(|z_t|) = \frac{4\xi^2}{\xi + 1/\xi} \frac{\Gamma\left(\frac{1+\upsilon}{2}\right)\sqrt{\upsilon - 2}}{\sqrt{\pi}\Gamma(\upsilon/2)}$ for an asymmetric Student-t (set $\xi = 1$ for the symmetric Student-t). Note that the use of a log transformation of the conditional variance ensures that σ_t^2 is always positive.

1.2.3 GJR Model

This popular model is proposed by Glosten, Jagannathan, and Runkle (1993). Its generalized version is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} (\alpha_i \varepsilon_{t-i}^2 + \gamma_i S_{t-i}^- \varepsilon_{t-i}^2) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{1.6}$$

where S_t^- is a dummy variable that takes the value 1 when ε_t is negative and 0 otherwise. In this model, it is assumed that the impact of ε_t^2 on the conditional variance σ_t^2 is different when ε_t is positive or negative.

A nice feature of the GJR model is that the null hypothesis of no leverage effect is easy to test. Indeed, $\gamma_1 = \ldots = \gamma_q = 0$ implies that the news impact curve is symmetric, i.e. past positive shocks have the same impact on today's volatility as past negative shocks.

1.2.4 APARCH Model

This model has been introduced by Ding, Granger, and Engle (1993). The APARCH (p, q) model can be expressed as:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i \left(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}\right)^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta, \tag{1.7}$$

where $\delta > 0$ and $-1 < \gamma_i < 1 \ (i = 1, ..., q)$.

The parameter δ plays the role of a Box-Cox transformation of σ_t while γ_i reflects the so-called leverage effect.

The APARCH includes seven other ARCH extensions as special cases:

- The ARCH of Engle (1982) when $\delta = 2,\ \gamma_i = 0\ (i = 1, \ldots, p)$ and $\beta_j = 0\ (j = 1, \ldots, p)$.
- The GARCH of Bollerslev (1986) when $\delta = 2$ and $\gamma_i = 0\ (i = 1, \ldots, p)$.
- Taylor (1986)/Schwert (1990)'s GARCH when $\delta = 1$, and $\gamma_i = 0\ (i = 1, \ldots, p)$.
- The GJR of Glosten, Jagannathan, and Runkle (1993) when $\delta = 2$.
- The TARCH of Zakoian (1994) when $\delta = 1$.
- The NARCH of Higgins and Bera (1992) when $\gamma_i = 0\ (i = 1, \ldots, p)$ and $\beta_j = 0\ (j = 1, \ldots, p)$.
- The Log-ARCH of Geweke (1986) and Pentula (1986), when $\delta \to 0$.

1.2.5 IGARCH Model

In many high-frequency time-series applications, the conditional variance estimated using a GARCH(p, q) process exhibits a strong persistence, that is:

$$\sum_{j=1}^{p} \beta_j + \sum_{i=1}^{q} \alpha_i \approx 1.$$

The IGARCH(p, q) model of Engle and Bollerslev (1986) is nothing but a GARCH(p, q) that imposes the strong constrain $\sum_{j=1}^{p} \beta_j + \sum_{i=1}^{q} \alpha_i = 1$ so that $p + q - 1$ parameters are estimated, the last one being obtained using the above constrain.

1.2.6 RiskMetrics^{TM}

In October 1994, the risk management group at J.P. Morgan released a technical document describing its internal market risk management methodology (J.P.Morgan, 1996). This methodology, called RiskMetrics^{TM} soon became a standard in the market risk measurement due to its simplicity.

Basically, the RiskMetrics^{TM} model is an IGARCH(1,1) model where the ARCH and GARCH coefficients are fixed.

The model is given by:

$$\sigma_t^2 = \omega + (1-\lambda)\varepsilon_{t-1}^2 + \lambda\sigma_{t-1}^2, \tag{1.8}$$

where $\omega = 0$ and λ is generally set to 0.94 with daily data and to 0.97 with weekly data but this value can obviously be changed by the user.

1.2.7 Fractionally Integrated Models

Volatility tends to change quite slowly over time, and, as shown in Ding, Granger, and Engle (1993) among others, the effects of a shock can take a considerable time to decay. In their study of the daily S&P500 index, they find that the squared returns series has positive autocorrelations over more than 2,500 lags (or more than 10 years !). Therefore the distinction between stationary and unit root processes seems to be far too restrictive. Indeed, the propagation of shocks in a stationary process occurs at an exponential rate of decay (so that it only captures the short-memory), while for an unit root process the persistence of shocks is infinite.

To mimic the behavior of the correlogram of the observed volatility, Baillie, Bollerslev, and Mikkelsen (1996) (hereafter denoted BBM) introduce the Fractionally Integrated GARCH (FIGARCH) model.

The conditional variance of the FIGARCH (p,d,q) is given by:

$$\sigma_t^2 = \underbrace{\omega[1-\beta(L)]^{-1}}_{\omega^*} + \underbrace{\left\{1-[1-\beta(L)]^{-1}\phi(L)(1-L)^d\right\}}_{\lambda(L)}\varepsilon_t^2, \tag{1.9}$$

or $\sigma_t^2 = \omega^* + \sum_{i=1}^{\infty}\lambda_i L^i \varepsilon_t^2 = \omega^* + \lambda(L)\varepsilon_t^2$, with $0 \leq d \leq 1$. Setting $\phi_1 = 0$ gives the condition for the FIGARCH $(1,d,0)$. $\lambda(L)$ is an infinite summation which, in practice, has to be truncated. BBM propose to truncate $\lambda(L)$ at 1000 lags and replace the unobserved ε_t^2's by the empirical counterpart of $E(\varepsilon_t^2)$, i.e. $1/T\sum_{t=1}^{T}\hat{\varepsilon}_t^2$.

1.2.8 Generalized Autoregressive Score (GAS) Models

It is well known that financial series occasionally exhibit large changes, also known as jumps. Several authors have shown that these jumps affect future volatility less than what standard volatility models would predict. Many volatility models, such as GARCH, are based on the assumption that each return observation has the same relative impact on future volatility, regardless of the magnitude of the return. This assumption is at odds with an increasing body of evidence indicating that the largest return observations have a relatively smaller effect on future volatility than smaller shocks.

To overcome this problem, Harvey and Chakravarty (2008) and Creal, Koopman, and Lucas (2012) independently proposed a novel way to deal with large returns in a GARCH context. Their models rely on a potentially non-normal distribution for the innovations (z_t) and a GARCH-type equation for the conditional variance derived from the conditional score of the assumed distribution with respect to the second moment.

The general framework is as follows. Let ψ_t denote a time-varying parameter vector (e.g. the conditional variance σ_t^2 or its log) and x_t a possible vector of exogenous variables, at time t. Defining $Y_t = \{\varepsilon_1, \ldots, \varepsilon_t\}, \{\Psi_t = \psi_1, \ldots, \psi_t\}$, and $X_t = \{x_1, \ldots, x_t\}$. Y_t is a $(t \times 1)$ vector with the demeaned returns up to time t. It is assumed that ε_t is generated by the very general observation density: $f(\varepsilon_t|\psi_t, \Psi_{t-1}, Y_{t-1}, X_t; \theta), t = 1, \ldots, T$. If $\varepsilon_t = \sigma_t z_t$, $z_t \sim N(0,1)$ and $\psi_t = \sigma_t^2$, $f(\varepsilon_t|\psi_t, \Psi_{t-1}, Y_{t-1}, X_t; \theta) = f(\varepsilon_t|\sigma_t^2; \theta) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp(-\frac{\varepsilon_t^2}{2\sigma_t^2})$.

For a GAS(1,1) model, the updating equation for the time-varying parameters ψ_t is the well-known autoregressive updating function: $\psi_t = \omega + B_1\psi_{t-1} + A_1\kappa_{t-1}$. Harvey and Chakravarty (2008) and Creal, Koopman, and Lucas (2012) propose to update the time-varying parameters with $\kappa_t = S_t \triangledown_t$. $\triangledown_t$ is the score with respect to the parameter ψ_t, i.e. $\triangledown_t = \partial \log f(y_t|\psi_t, \Psi_{t-1}, Y_{t-1}, X_t; \theta)/\partial\psi_t$ and S_t is a time dependent scaling matrix.

Note that for a standard Normal-GARCH model, i.e. if $\varepsilon_t \sim N(0, \sigma_t^2)$, and $\psi_t = \sigma_t^2$ $\triangledown_t = \partial - 0.5\left(\log\sigma_t^2 + \epsilon_t^2\sigma_t^{-2}\right)/\partial\sigma_t^2 = 0.5(z_t^2 - 1)\sigma_t^2$. Therefore, a GARCH(1,1) model corresponds to a Normal-GAS(1,1) model (i.e. GAS(1,1) with $z_t \sim N(0,1)$) with $S_t = 2$, $A_1 = \alpha_1$ and $B_1 = \alpha_1 + \beta_1$.

Indeed, let us rewrite the GARCH(1,1) model as follows:

$$\sigma_t^2 = \omega + \alpha_1 \underbrace{z_{t-1}^2\sigma_{t-1}^2}_{\varepsilon_{t-1}^2} + \beta_1\sigma_{t-1}^2 \tag{1.10}$$

or equivalently

$$\sigma_t^2 = \omega + \alpha_1 \underbrace{(z_{t-1}^2 - 1)}_{u_{t-1}}\sigma_{t-1}^2 + \underbrace{(\alpha_1 + \beta_1)}_{B_1}\sigma_{t-1}^2. \tag{1.11}$$

In the above equation $u_t = z_{t-1}^2 - 1$ is proportional to the score of the conditional distribution of ε_t with respect to σ_{t-1}^2 and therefore is a natural choice of updating scheme in a 'Newton-Raphson' sense.

The only difference between Harvey and Chakravarty (2008) and Creal, Koopman, and Lucas (2012) is on S_t. Creal, Koopman, and Lucas (2012) discuss different choices for S_t and recommend using $S_t = 1$ or $S_t = (E_{t-1} \bigtriangledown_t \bigtriangledown_t')^{-1}$ while Harvey and Chakravarty (2008) set $S_t = 2$. XlQuant follows Harvey and Chakravarty (2008) and sets $S_t = 2$.

The specification of the GAS(1,1) model of Harvey and Chakravarty (2008) combined with a normal, and Student-t is given below:

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}\sigma_{t-1}^2 + \psi_1\sigma_{t-1}^2, \tag{1.12}$$

where

$$u_t = z_t^2 - 1 \text{ if } z_t \sim N(0,1); \tag{1.13}$$

$$u_t = \frac{(\nu+1)z_t^2}{\nu - 2 + z_t^2} - 1 \text{ if } z_t \sim t(0,1,\nu). \tag{1.14}$$

$$\tag{1.15}$$

Harvey and Chakravarty (2008) call the above GAS model with a T distribution 'Beta-t-GARCH' because, for this distribution, $(u_t+1)/(\nu+1)$ has a Beta distribution. Note that the specification of the GAS models in Harvey and Chakravarty (2008) and Harvey and Succarat (2012) slightly differ from those implemented in XlQuant because the score is derived for standardized distributions (i.e. $E(z_t) = 0$ and $V(z_t) = 1$) and not just centered distributions (i.e. $E(z_t) = 0$ but $V(z_t) \neq 1$).

1.2.9 Forecasting the Conditional Variance of GARCH-type models

In the simple GARCH(p,q) case, the optimal h-step-ahead forecast of the conditional variance, i.e. $\hat{\sigma}_{t+h|t}^2$ is given by:

$$\sigma_{t+h|t}^2 = \hat{\omega} + \sum_{i=1}^{q} \hat{\alpha}_i \varepsilon_{t+h-i|t}^2 + \sum_{j=1}^{p} \hat{\beta}_j \sigma_{t+h-j|t}^2, \tag{1.16}$$

where $\varepsilon_{t+i|t}^2 = \sigma_{t+i|t}^2$ for $i > 0$ while $\varepsilon_{t+i|t}^2 = \varepsilon_{t+i}^2$ and $\sigma_{t+i|t}^2 = \sigma_{t+i}^2$ for $i \leq 0$. Equation (1.16) is usually computed recursively, even if a closed form solution of $\sigma_{t+h|t}^2$ can be obtained by recursive substitution in Equation (1.16).

Similarly, one can easily obtain the h-step-ahead forecast of the conditional variance of an ARCH, IGARCH and FIGARCH model. By contrast, for thresholds models, the computation of out-of-sample forecasts is more complicated. Indeed, for EGARCH, GJR and APARCH models, the assumption made on the innovation process may have an effect on the forecast (especially for $h > 1$).

For instance, for the GJR (p,q) model, we have

$$\hat{\sigma}^2_{t+h|t} = \hat{\omega} + \sum_{i=1}^{q} (\hat{\alpha}_i \varepsilon^2_{t-i+h|t} + \hat{\gamma}_i S^-_{t-i+h|t}\varepsilon^2_{t-i+h|t}) + \sum_{j=1}^{p} \hat{\beta}_j \sigma^2_{t-j+h|t}. \quad (1.17)$$

When $\gamma_i = 0$ for all i, we obtain the forecast of the GARCH model. Otherwise, $S^-_{t-i+h|t}$ has to be computed. Note first that $S^-_{t+i|t} = S^-_{t+i}$ for $i \leq 0$. However, when $i > 0$, $S^-_{t+i|t}$ depends on the choice of the distribution of z_t. When the distribution of z_t is symmetric around 0 (for the Gaussian, Student), the probability that ε_{t+i} is negative is $S^-_{t+i|t} = 0.5$. If z_t is (standardized) skewed-Student distributed with asymmetry parameter ξ and degree of freedom υ, $S^-_{t+i|t} = \frac{1}{1+\xi^2}$ since ξ^2 is the ratio of probability masses above and below the mode.

For the APARCH (p,q) model,

$$\begin{aligned}
\hat{\sigma}^{\delta}_{t+h|t} &= E\left(\sigma^{\delta}_{t+h}|\Omega_t\right) \\
&= E\left(\hat{\omega} + \sum_{i=1}^{q} \hat{\alpha}_i \left(|\varepsilon_{t+h-i}| - \hat{\gamma}_i \varepsilon_{t+h-i}\right)^{\hat{\delta}} + \sum_{j=1}^{p} \hat{\beta}_j \sigma^{\hat{\delta}}_{t+h-j} \mid \Omega_t\right) \\
&= \hat{\omega} + \sum_{i=1}^{q} \hat{\alpha}_i E\left[(\varepsilon_{t+h-i} - \hat{\gamma}_i \varepsilon_{t+h-i})^{\hat{\delta}}|\Omega_t\right] + \sum_{j=1}^{p} \hat{\beta}_j \sigma^{\hat{\delta}}_{t+h-j|t}, \quad (1.18)
\end{aligned}$$

where $E\left[(\varepsilon_{t+k} - \hat{\gamma}_i \varepsilon_{t+k})^{\hat{\delta}}|\Omega_t\right] = \kappa_i \sigma^{\hat{\delta}}_{t+k|t}$, for $k > 1$ and $\kappa_i = E\left(|z| - \gamma_i z\right)^{\hat{\delta}}$.

For the EGARCH (p,q) model,

$$\begin{aligned}
\log \hat{\sigma}^2_{t+h|t} &= E\left(\log \sigma^2_{t+h}|\Omega_t\right) \\
&= E\left\{\hat{\omega} + \left[1 - \hat{\beta}(L)\right]^{-1} [1 + \hat{\alpha}(L)]\hat{g}(z_{t+h-1}) \mid \Omega_t\right\} \\
&= \left[1 - \hat{\beta}(L)\right]\hat{\omega} + \hat{\beta}(L)\log \hat{\sigma}^2_{t+h|t} + [1 + \hat{\alpha}(L)]\hat{g}(z_{t+h-1|t}) \quad (1.19)
\end{aligned}$$

where $\hat{g}(z_{t+k|t}) = \hat{g}(z_{t+k})$ for $k \leq 0$ and 0 for $k > 0$.

For the GAS(p,q) model,

$$\sigma^2_{t+h|t} = \hat{\omega} + \sum_{i=1}^{q} \hat{\alpha}_i u_{t+h-i|t}\sigma^2_{t+h-i|t} + \sum_{j=1}^{p} \hat{\phi}_j \sigma^2_{t+h-j|t}, \quad (1.20)$$

where $u_{t+i|t}\sigma^2_{t+i|t} = 0$ for $i > 1$.

1.3 Estimation

Estimation of GARCH-type models is commonly done by maximum likelihood so that one has to make an additional assumption about the innovation process z_t, i.e. choosing a density function $D(0,1)$ with a mean 0 and a unit variance.

Three distributions are available: the usual Gaussian (normal) distribution, the Student-t distribution, and the skewed-Student distribution (which is available for all models but the GAS model).

The logic of ML is to interpret the density as a function of the parameter set, conditional on a set of sample outcomes. This function is called the *likelihood function*.

If we express the mean equation as in Equation (1.1) and $\varepsilon_t = z_t\sigma_t$, the log-likelihood function of the standard normal distribution is given by:

$$L_{norm} = -\frac{1}{2}\sum_{t=1}^{T}\left[\log\left(2\pi\right) + \log\left(\sigma_t^2\right) + z_t^2\right], \tag{1.21}$$

where T is the number of observations.

For a Student-t distribution, the log-likelihood is:

$$\begin{aligned} L_{Stud} &= T\left\{\log\Gamma\left(\frac{\upsilon+1}{2}\right) - \log\Gamma\left(\frac{\upsilon}{2}\right) - \frac{1}{2}\log\left[\pi(\upsilon-2)\right]\right\} \\ &- \frac{1}{2}\sum_{t=1}^{T}\left[\log(\sigma_t^2) + (1+\upsilon)\log\left(1+\frac{z_t^2}{\upsilon-2}\right)\right], \end{aligned} \tag{1.22}$$

where υ is the degrees of freedom, $2 < \upsilon \leq \infty$ and $\Gamma(.)$ is the gamma function.

The main drawback of the previous densities is that even if they may account for fat tails, they are symmetric. Skewness and kurtosis are important in financial applications in many respects (in asset pricing models, portfolio selection, option pricing theory or Value-at-Risk among others).

The log-likelihood of a standardized (zero mean and unit variance) skewed-Student is:

$$\begin{aligned} L_{SkSt} &= T\left\{\log\Gamma\left(\frac{\upsilon+1}{2}\right) - \log\Gamma\left(\frac{\upsilon}{2}\right) - 0.5\log\left[\pi\left(\upsilon-2\right)\right] + \log\left(\frac{2}{\xi+\frac{1}{\xi}}\right) + \log\left(s\right)\right\} \\ &- 0.5\sum_{t=1}^{T}\left\{\log\sigma_t^2 + (1+\upsilon)\log\left[1+\frac{\left(sz_t+m\right)^2}{\upsilon-2}\xi^{-2I_t}\right]\right\}, \end{aligned} \tag{1.23}$$

where

$$I_t = \begin{cases} 1 \text{ if } z_t \geq -\frac{m}{s} \\ -1 \text{ if } z_t < -\frac{m}{s} \end{cases},$$

ξ is the asymmetry parameter, υ is the degree of freedom of the distribution,

$$m = \frac{\Gamma\left(\frac{\upsilon-1}{2}\right)\sqrt{\upsilon-2}}{\sqrt{\pi}\Gamma\left(\frac{\upsilon}{2}\right)}\left(\xi - \frac{1}{\xi}\right),$$

and

$$s = \sqrt{\left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2}.$$

Note that XlQuant does not estimate ξ but $\log(\xi)$ to facilitate inference about the null hypothesis of symmetry (since the skewed-Student equals the symmetric Student distribution when $\xi = 1$ or $\log(\xi) = 0$). The estimated value of $\log(\xi)$ is reported in the output under the label “Asymmetry”. See Bauwens and Laurent (2005) for more details.

1.4 Value-at-Risk (VaR) estimation using XlQuant

In recent years, the tremendous growth of trading activity and the widely publicized trading loss of well-known financial institutions (see Jorion, 2000, for a brief history of these events) has led financial regulators and supervisory authorities to favor quantitative techniques which appraise the possible loss that these institutions can incur. Value-at-Risk has become one of the most sought-after techniques as it provides a simple answer to the following question: with a given probability (say α), what is my predicted financial loss over a given time horizon? The answer is the VaR at level α, which gives an amount in the currency of the traded assets (in dollar terms for example) and is thus easily understandable.

It turns out that the VaR has a simple statistical definition: the VaR at level α for a sample of returns is defined as the corresponding empirical quantile at $\alpha\%$. Because of the definition of the quantile, we have that, with probability $1 - \alpha$, the returns will be larger than the VaR. In other words, with probability $1 - \alpha$, the losses will be smaller than the dollar amount given by the VaR.[1] From an empirical point of view, the computation of the VaR for a collection of returns thus requires the computation of the empirical quantile at level α of the distribution of the returns of the portfolio.

In this section, we show how to use XlQuant to predict the VaR and test the adequacy of the selected model in forecasting the VaR of the investigated series.

The long side of the daily VaR is defined as the VaR level for traders having long positions in the relevant equity index: this is the “usual” VaR where traders incur losses when negative returns are observed. Correspondingly, the short side of the daily VaR

[1] Contrary to some wide-spread beliefs, the VaR does not specify the maximum amount that can be lost.

is the VaR level for traders having short positions, i.e. traders who incur losses when stock prices increase.[2]

In-sample one-step-ahead VaR computed in $t-1$ for long trading positions is given by $\mu_t + z_\alpha \sigma_t$, for short trading positions it is equal to $\mu_t + z_{1-\alpha}\sigma_t$, with z_α being the left quantile at $\alpha\%$ for the assumed distribution and $z_{1-\alpha}$ is the right quantile at $\alpha\%$.[3]

Similarly, h-step-ahead out-of-sample forecasts of the VaR for long trading positions are given by $\mu_{t+h|t} + z_\alpha \sigma_{t+h|t}$, for short trading positions it is equal to $\mu_{t+h|t} + z_{1-\alpha}\sigma_{t+h|t}$.

[2]An asset is short-sold by a trader when it is first borrowed and subsequently sold on the market. By doing this, the trader hopes that the price will fall, so that he can then buy the asset at a lower price and give it back to the lender.

[3]Note that when computing the VaR, μ_t and σ_t are evaluated by replacing the unknown parameters by their maximum likelihood estimates (MLE).

Chapter 2

Multivariate GARCH Models

It is now widely accepted that financial volatilities move together over time across assets and markets.

Recognizing this feature through a multivariate modeling framework leads to more relevant empirical models than working with separate univariate models. From a financial point of view, it opens the door to better decision tools in various areas, such as asset pricing, portfolio selection, option pricing, hedging, and risk management. Indeed, unlike at the beginning of the 1990s, several institutions have now developed the necessary skills to use the econometric theory in a financial perspective.

MGARCH models were initially developed in the late eighties and the first half of the nineties, and after a period of tranquility in the second half of the nineties, this area seems to be experimenting again a quick expansion phase. See Bauwens, Laurent, and Rombouts (2006) for a survey on MGARCH models.

Consider a vector stochastic process $\{y_t\}$ of dimension $N \times 1$. As usual, we condition on the sigma field, denoted by Ω_{t-1}, generated by the past information (here the y_t's) until time $t-1$. We denote by θ a finite vector of parameters and we write:

$$y_t = \mu_t(\theta) + \epsilon_t, \tag{2.1}$$

where $\mu_t(\theta)$ is the conditional mean vector and,

$$\epsilon_t = H_t^{1/2}(\theta) z_t, \tag{2.2}$$

where $H_t^{1/2}(\theta)$ is a $N \times N$ positive definite matrix. Furthermore, we assume the $N \times 1$ random vector z_t to have the following first two moments:

$$\begin{aligned} \mathbb{E}(z_t) &= 0 \\ \text{Var}(z_t) &= I_N, \end{aligned} \tag{2.3}$$

where I_N is the identity matrix of order N. We still have to explain what $H_t^{1/2}$ is (for convenience we leave out θ in the notation). To make this clear we calculate the conditional variance matrix of y_t:

$$\begin{aligned}\text{Var}(y_t|\Omega_{t-1}) = \text{Var}_{t-1}(y_t) &= \text{Var}_{t-1}(\epsilon_t) \\ &= H_t^{1/2}\text{Var}_{t-1}(z_t)(H_t^{1/2})' \\ &= H_t. \end{aligned} \tag{2.4}$$

Hence $H_t^{1/2}$ is any $N \times N$ positive definite matrix such that H_t is the conditional variance matrix of y_t, *e.g.* $H_t^{1/2}$ may be obtained by the Cholesky factorization of H_t. Both H_t and μ_t depend on the unknown parameter vector θ, which can in most cases be split into two disjoint parts, one for μ_t and one for H_t. A case where this is not true is that of GARCH-in-mean models, where μ_t is functionally dependent on H_t.

In the next subsections we review different specifications of H_t implemented in XlQuant.

2.1 Conditional mean specification

Recall that the conditional mean equation is specified as follows:

$$y_t = \mu_t(\theta) + \epsilon_t, \tag{2.5}$$

where $\mu_t(\theta) = \{\mu_{1t}, \ldots, \mu_{Nt}\}$ is the conditional mean vector of y_t.

ARMA specifications are available for all the MGARCH models described in the next sections. XlQuant provides diagonal ARMA models in the sense that an ARMA specification is fitted on each univariate series, i.e.

$$\Psi_i\,(L)\,(y_{it} - \mu_i) = \Theta_i\,(L)\,\epsilon_{it}, \tag{2.6}$$

where L is the lag operator[1], $\Psi_i\,(L) = 1 - \sum\limits_{j=1}^{n} \psi_{ij} L^j$ and $\Theta_i\,(L) = 1 + \sum\limits_{j=1}^{s} \theta_{ij} L^j$.

2.2 Multivariate GARCH specifications

The models in this category are multivariate extensions of the univariate GARCH model. When we consider VARMA models for the conditional mean of several time series the number of parameters increases rapidly. The same happens for multivariate GARCH models as straightforward extensions of the univariate GARCH model. Furthermore, since H_t is a variance matrix, positive definiteness has to be ensured.

[1] Recall that $L^k y_t = y_{t-k}$.

2.2.1 RiskMetrics

J.P.Morgan (1996) uses the exponentially weighted moving average model (EWMA) to forecast variance and covariances. Practitioners who study volatility processes often observe that their model is very close to the unit root case. To take this into account, Riskmetrics defines the variances and covariances as IGARCH type models (Engle and Bollerslev, 1986):

Definition 1. *The RiskMetrics model is defined as:*

$$H_t = (1-\lambda)\epsilon_{t-1}\epsilon'_{t-1} + \lambda H_{t-1}, \tag{2.7}$$

or alternatively

$$H_t = \frac{(1-\lambda)}{(1-\lambda)^{t-1}} \sum_{i=1}^{t-1} \lambda^{i-1} \epsilon_{t-1}\epsilon'_{t-1}. \tag{2.8}$$

The decay factor λ $(0 < \lambda < 1)$ proposed by Riskmetrics is equal to 0.94 for daily data and 0.97 for monthly data. The decay factor is not estimated but suggested by Riskmetrics. In this respect, this model is easy to work with in practice. However, imposing the same dynamics on every component in a multivariate GARCH model, no matter which data are used, is difficult to justify.

2.2.2 BEKK models

Engle and Kroner (1995) propose a parametrization for H_t that easily imposes its positivity, i.e. the BEKK model (the acronym comes from synthesized work on multivariate models by Baba, Engle, Kraft and Kroner).

Definition 2. *The BEKK(p, q) model is defined as:*

$$H_t = C'C + \sum_{i=1}^{q} A'_i \epsilon_{t-i}\epsilon'_{t-i} A_i + \sum_{j=1}^{p} G'_i H_{t-j} G_j, \tag{2.9}$$

where C, the A's and the G's matrices are of dimension $N \times N$ but C is upper triangular.

The original BEKK model is a bit more general since it involves a summation over K terms. We restrict K to be equal to 1. The BEKK model is actually a special case of the VEC model of Bollerslev, Engle, and Wooldridge (1988).

The number of ARCH and GARCH parameters in the BEKK(1,1) model is $N(5N+1)/2$. To reduce this number, and consequently to reduce the generality, one can impose a diagonal BEKK model, *i.e.* A_i and G_j in (2.9) are diagonal matrices.

Another way to reduce the number of parameters is to use a scalar BEKK model, *i.e.* A_i and G_j are equal to a scalar times the identity matrix.

XlQuant provides two of these models, i.e. the Diag-BEKK and Scalar-BEKK model.

The Diag-BEKK and Scalar-BEKK are covariance-stationary if $\sum_{i=1}^{q} a_{nn,i}^2 + \sum_{j=1}^{p} g_{nn,j}^2 < 1$, $\forall n = 1, \ldots, N$, $\sum_{i=1}^{q} {a_i}^2 + \sum_{j=1}^{p} {g_j}^2 < 1$, respectively. These conditions are imposed during the estimation.

When it exists, the unconditional variance matrix $\Sigma \equiv \mathbb{E}(H_t)$ of the BEKK model, is given by

$$vec(\Sigma) = \left[I_{N^2} - \sum_{i=1}^{q} (A_i \otimes A_i)' - \sum_{j=1}^{p} (G_j \otimes G_j)' \right]^{-1} vec(C'C), \qquad (2.10)$$

where vec denotes the operator that stacks the columns of a matrix as a vector.

Similar expressions can be obtained for the Diag-BEKK and Scalar-BELL models.

Variance Targeting

What renders most MGARCH models difficult for estimation is their high number of parameters. A simple trick to ensure a reasonable value of the model-implied unconditional covariance matrix, which also helps to reduce the number of parameters in the maximization of the likelihood function, is referred to as variance targeting by Engle and Mezrich (1996). The conditional variance matrix of the BEKK model (and all its particular cases), may be expressed in terms of the unconditional variance matrix and other parameters. Doing so one can reparametrize the model using the unconditional variance matrix and replace it by a consistent estimator (before maximizing the likelihood).

Applying variance targeting to the BEKK models implies replacing CC' by unvec $\left[I_{N^2} - \sum_{i=1}^{q} (A_i \otimes A_i)' - \sum_{j=1}^{p} (G_j \otimes G_j)' \right] \bar{\Sigma}$, where $\bar{\Sigma}$ is the unconditional variance-covariance matrix of ϵ and unvec is the reverse of the vec operator.[2]

The difficulty when estimating a BEKK model is the high number of unknown parameters, even after imposing several restrictions. It is thus not surprising that these models are rarely used when the number of series is larger than 3 or 4.

[2] When explanatory variables appear in the BEKK equaton and the variance targeting option is selected, these variables are simply centered as explaned in Section **??**.

2.2.3 Conditional correlation models

This section collects models that may be viewed as nonlinear combinations of univariate GARCH models. This allows for models where one can specify separately, on the one hand, the individual conditional variances, and on the other hand, the conditional correlation matrix. For models of this category, theoretical results on stationarity, ergodicity and moments may not be so straightforward to obtain as for models presented in the preceding sections. Nevertheless, they are less greedy in parameters than the models of the first category, and therefore they are more easily estimable.

The conditional variance matrix for this class of models is specified in a hierarchical way. First, one chooses a GARCH-type model for each conditional variance. Second, based on the conditional variances one models the conditional correlation matrix (imposing its positive definiteness $\forall t$).

Bollerslev (1990) proposes a class of MGARCH models in which the conditional correlations are constant and thus the conditional covariances are proportional to the product of the corresponding conditional standard deviations. This restriction highly reduces the number of unknown parameters and thus simplifies estimation.

Definition 3. *The CCC model is defined as:*

$$H_t = D_t R D_t = \left(\rho_{ij}\sqrt{h_{iit}h_{jjt}}\right), \tag{2.11}$$

where

$$D_t = \text{diag}\,(h_{11t}^{1/2} \dots h_{NNt}^{1/2}), \tag{2.12}$$

h_{iit} *can be defined as any univariate GARCH model, and*

$$R = (\rho_{ij}) \tag{2.13}$$

is a symmetric positive definite matrix with $\rho_{ii} = 1, \forall\, i$.

R is the matrix containing the constant conditional correlations ρ_{ij}. The original CCC model has a GARCH(1,1) specification for each conditional variance in D_t:

$$h_{iit} = \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i h_{ii,t-1} \quad i = 1, \dots, N. \tag{2.14}$$

This CCC model contains $N(N+5)/2$ parameters. H_t is positive definite if and only if all the N conditional variances are positive and R is positive definite. The unconditional variances are easily obtained, as in the univariate case, but the unconditional covariances are difficult to calculate because of the nonlinearity in (2.11).

The assumption that the conditional correlations are constant may seem unrealistic in many empirical applications. Engle (2002) proposes a generalization of the CCC model by making the conditional correlation matrix time dependent. The model is then called a dynamic conditional correlation (DCC) model. An additional difficulty is that the time dependent conditional correlation matrix has to be positive definite $\forall t$. The DCC models guarantee this under simple conditions on the parameters.

Definition 4. *The DCC model of Engle (2002) is defined as*

$$H_t = D_t R_t D_t, \tag{2.15}$$

where D_t is defined in (2.12), h_{iit} can be defined as any univariate GARCH model, and

$$R_t = diag\,(q_{11,t}^{-1/2} \ldots q_{NN,t}^{-1/2})\; Q_t\; diag\,(q_{11,t}^{-1/2} \ldots q_{NN,t}^{-1/2}), \tag{2.16}$$

where the $N \times N$ symmetric positive definite matrix $Q_t = (q_{ij,t})$ is given by:

$$Q_t = (1 - \alpha - \beta)\overline{Q} + \alpha u_{t-1}u'_{t-1} + \beta Q_{t-1}, \tag{2.17}$$

with $u_t = (u_{1t}, \ldots, u_{Nt})'$, where $u_{it} = \epsilon_{it}/\sqrt{h_{iit}}$. $\overline{Q}$ is the $N \times N$ unconditional variance matrix of u_t, and α and β are nonnegative scalar parameters satisfying $\alpha+\beta < 1$.

The elements of $\overline{Q}$ can be estimated or alternatively set to their empirical counterpart to render the estimation even simpler.

Interestingly, CCC and DCC models can be estimated consistently in two steps (see Section 2.3.2) which makes this approach feasible when N is high. Of course, when N is large, the restriction of common dynamics gets tighter, but for large N the problem of maintaining tractability also gets harder.

2.3 Estimation

In the previous section we have defined existing specifications of conditional variance matrices that enter the definition either of a data generating process (DGP) or of a model to be estimated. In Section 2.3.1, we discuss maximum likelihood (ML) estimation of these models, and in Section 2.3.2 we explain a two-step approach for estimating conditional correlation models. Finally, we review briefly the variance targeting issue in Section 2.3.3.

2.3.1 Maximum Likelihood

Suppose the vector stochastic process $\{y_t\}$ (for $t = 1, \ldots, T$) is a realization of a DGP whose conditional mean, conditional variance matrix and conditional distribution are respectively $\mu_t(\theta_0)$, $H_t(\theta_0)$ and $p(y_t|\zeta_0, \Omega_{t-1})$, where $\zeta_0 = (\theta_0\ \eta_0)$ is a r-dimensional parameter vector and η_0 is the vector that contains the parameters of the distribution of the innovations z_t (there may be no such parameter). Importantly, to justify the choice of the estimation procedure, we assume that the model to be estimated encompasses the true formulations of $\mu_t(\theta_0)$ and $H_t(\theta_0)$.

The procedure most often used in estimating θ_0 involves the maximization of a likelihood function constructed under the auxiliary assumption of an $i.i.d.$ distribution for the standardized innovations z_t. The $i.i.d.$ assumption may be replaced by the weaker assumption that z_t is a martingale difference sequence with respect to Ω_{t-1}, but this type of assumption does not translate into the likelihood function. The likelihood function for the $i.i.d.$ case can then be viewed as a quasi-likelihood function.

Consequently, one has to make an additional assumption on the innovation process by choosing a density function, denoted $g(z_t(\theta)|\eta)$ where η is a vector of nuisance parameters. The problem to solve is thus to maximize the sample log-likelihood function $L_T(\theta, \eta)$ for the T observations (conditional on some starting values for μ_0 and H_0), with respect to the vector of parameters $\zeta = (\theta, \eta)$, where

$$L_T(\zeta) = \sum_{t=1}^{T} \log f(y_t|\zeta, \Omega_{t-1}), \tag{2.18}$$

with

$$f(y_t|\zeta, \Omega_{t-1}) = |H_t|^{-1/2} g\left(H_t^{-1/2}(y_t - \mu_t)|\eta\right), \tag{2.19}$$

and the dependence with respect to θ occurs through μ_t and H_t. The term $|H_t|^{-1/2}$ is the Jacobian that arises in the transformation from the innovations to the observables.

The most commonly employed distribution in the literature is the multivariate normal, uniquely determined by its first two moments (so that $\zeta = \theta$ since η is empty). In this case, the sample log-likelihood is:

$$L_T(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left[N\log(2\pi) + \log \mid H_t \mid + (y_t - \mu_t)' H_t^{-1}(y_t - \mu_t)\right]. \tag{2.20}$$

It is well-known that the normality of the innovations is rejected in most applications dealing with daily or weekly data. In particular, the kurtosis of most financial asset returns is larger than three, which means that they have too many extreme values to be normally distributed. Moreover, their unconditional distribution has often fatter tails than what is implied by a conditional normal distribution: the increase of the kurtosis coefficient brought by the dynamics of the conditional variance is not usually sufficient to match adequately the unconditional kurtosis of the data.

However, as shown by Bollerslev and Wooldridge (1992), a consistent estimator of θ_0 may be obtained by maximizing (2.20) with respect to θ *even* if the DGP is not conditionally Gaussian. This estimator, called (Gaussian) quasi-maximum likelihood (QML) or pseudo-maximum likelihood (PML) estimator, is consistent provided the conditional mean and the conditional variance are specified correctly. Jeantheau (1998) proves the strong consistency of the Gaussian QML estimator of multivariate GARCH models. He also provides sufficient identification conditions for the CCC model. See Gourieroux (1997) for a detailed description of the QML method in a MGARCH context and its asymptotic properties. For these reasons and as far as the purpose of the analysis is to estimate consistently the first two conditional moments, estimating MGARCH models by QML is justified.

Nevertheless, in certain situations it is desirable to search for a better distribution for the innovation process. For instance, when one is interested in obtaining density forecasts, (see Diebold, Gunther, and Tay, 1998, in the univariate case and Diebold, Hahn, and Tay, 1999, in the multivariate case) it is natural to relax the normality assumption, keeping in mind the risk of inconsistency of the estimator (see Newey and Steigerwald, 1997).

A natural alternative to the multivariate Gaussian density is the Student density, see Harvey, Ruiz, and Shephard (1992) and Fiorentini, Sentana, and Calzolari (2003). The latter has an extra scalar parameter, the degrees of freedom parameter, denoted ν hereafter. When this parameter tends to infinity, the Student density tends to the normal density. When it tends to zero, the tails of the density become thicker and thicker. The parameter value indicates the order of existence of the moments, *e.g.* if $\nu = 2$, the second order moments do not exist, but the first order moments exist. For this reason, it is convenient (although not necessary) to assume that $\nu > 2$, so that H_t is always interpretable as a conditional covariance matrix. Under this assumption, the Student density can be defined as:

$$g(z_t|\theta,\nu) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left[\pi(\nu-2)\right]^{\frac{N}{2}}}\left[1+\frac{z_t'z_t}{\nu-2}\right]^{-\frac{N+\nu}{2}}, \tag{2.21}$$

where $\Gamma(.)$ is the Gamma function. Note that in this case $\eta = \nu$. The density function of y_t is easily obtained by applying (2.19).

2.3.2 Two-step estimation

A useful feature of the CCC and DCC models presented in Section 2.2.3 is that they can be estimated consistently using a two-step approach. Engle and Sheppard (2001) show that in the case of a DCC model, the log-likelihood can be written as the sum of a mean and volatility part (depending on a set of unknown parameters θ_1^*) and a correlation part (depending on θ_2^*).

Indeed, recalling that the conditional variance matrix of a DCC model can be expressed as $H_t = D_t R_t D_t$, an inefficient but consistent estimator of the parameter θ_1^* can be found by replacing R_t by the identity matrix in (2.20). In this case the quasi-loglikelihood function corresponds to the sum of loglikelihood functions of N univariate models:

$$QL1_T(\theta_1^*) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[\log(2\pi) + \log(h_{iit}) + \frac{(y_{it} - \mu_{it})^2}{h_{iit}}\right]. \qquad (2.22)$$

Given θ_1^* and under appropriate regularity conditions, a consistent, but inefficient, estimator of θ_2^* can be obtained by maximizing:

$$QL2_T(\theta_2^*|\theta_1^*) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log|R_t| + u_t' R_t^{-1} u_t\right), \qquad (2.23)$$

where $u_t = D_t^{-1}(y_t - \mu_t)$.

The sum of the likelihood functions in (2.22) and (2.23) plus half of the total sum of squared standardized residuals ($\sum_t u_t' u_t/2$, which is almost equal to $NT/2$), is equal to the log-likelihood in (2.20). It is thus possible to compare the log-likelihood of the two-step approach with that of the one-step approach and of other models.

Engle and Sheppard (2001) explain that the estimators $\hat{\theta}_1^*$ and $\hat{\theta}_2^*$, obtained by maximizing (2.22) and (2.23) separately, are not fully efficient (even if z_t is normally distributed) since they are limited information estimators.

XlQuant implements the two-step approach described above (for the CCC and DCC models) but also allows the estimation of these model in one-step. Note that when choosing the one-step approach, the model is first estimated with the two-step approach to get accurate starting values.

Importantly, XlQuant allows the selection of non-standard ARCH models for the conditional variances (like APARCH, GJR, etc.) and an ARMA specification for the conditional mean.

2.3.3 Variance Targeting

We have seen that what renders most MGARCH models difficult for estimation is their high number of parameters. A simple trick to ensure a reasonable value of the model-implied unconditional covariance matrix, which also helps to reduce the number of parameters in the maximization of the likelihood function, is referred to as variance targeting by Engle and Mezrich (1996). For example, in the VEC model (and all its particular cases), the conditional variance matrix may be expressed in terms of the unconditional variance matrix (see Section 2.2.2) and other parameters. Doing so one can reparametrize the model using the unconditional variance matrix and replace it by a consistent estimator (before maximizing the likelihood). When doing this, one should correct the covariance matrix of the estimator of the other parameters for the uncertainty in the preliminary estimator. In DCC models, this can also be done with the constant matrix of the correlation part, *e.g.* $\overline{Q}$ in (2.17). In this case, the two-step estimation procedure explained in Section 2.3.2 becomes a three-step procedure.

Bibliography

AIELLI, G. (2009): "Dynamic conditional correlations: on properties and estimation," Department of Statistics, University of Florence, Mimeo.

ALEXANDER, C. (2001.): *Market Models*. Wiley, New York.

ALEXANDER, C., and A. CHIBUMBA (1997): "Multivariate Orthogonal Factor GARCH," University of Sussex, Mimeo.

BAILLIE, R. T., T. BOLLERSLEV, and H. O. MIKKELSEN (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

BAUWENS, L., and S. LAURENT (2005): "A New Class of Multivariate Skew Densities, with Application to GARCH Models," *Journal of Business and Economic Statistics*, 23, 346–354.

BAUWENS, L., S. LAURENT, and J. ROMBOUTS (2006): "A review of Multivariate GARCH Models with Applications to Financial Data," *Journal of Applied Econometrics*, 21.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

——— (1990): "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH model," *Review of Economics and Statistics*, 72, 498–505.

BOLLERSLEV, T., R. F. ENGLE, and J. M. WOOLDRIDGE (1988): "A Capital Asset Pricing Mode1 with Time Varying Covariances," *Journal of Political Economy*, 96, 116–131.

BOLLERSLEV, T., and H. O. MIKKELSEN (1996): "Modeling and Pricing Long-Memory in Stock Market Volatility," *Journal of Econometrics*, 73, 151–184.

BOLLERSLEV, T., and J. M. WOOLDRIDGE (1992): "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances," *Econometric Reviews*, 11, 143–172.

BOSWIJK, H., and R. VAN DER WEIDE (2006): "Wake me up before you GO-GARCH," UvA-Econometrics Discussion Paper 2006/03.

BRUNETTI, C., and C. GILBERT (2000): "Bivariate FIGARCH and Fractional Cointegration," *Journal of Empirical Finance*, 7, 509–530.

COMTE, F., and O. LIEBERMAN (2003): "Asymptotic Theory for Multivariate GARCH Processes," *Journal of Multivariate Analysis*, 84, 61–84.

CREAL, D., S. KOOPMAN, and A. LUCAS (2012): "A General Framework for Observation Driven Time-Varying Parameter Models," Forthcoming in Journal of Applied Econometrics.

DIEBOLD, F. X., T. A. GUNTHER, and A. S. TAY (1998): "Evaluating Density Forecasts, with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.

DIEBOLD, F. X., J. HAHN, and A. S. TAY (1999): "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange," *Review of Economics and Statistics*, 81, 863–883.

DING, Z., C. W. J. GRANGER, and R. F. ENGLE (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

DUFOUR, J.-M., L. KHALAF, and M.-C. BEAULIEU (2003): "Finite-sample Diagnostics for Multivariate Regression with Applications to Linear Asset Pricing Models," Working Paper No. 06-2003, CIREQ.

ENGLE, R. (2002): "Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

ENGLE, R., and F. K. KRONER (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.

ENGLE, R., and J. MEZRICH (1996): "GARCH for Groups," *RISK*, 9, 36–40.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

ENGLE, R. F., and T. BOLLERSLEV (1986): "Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

ENGLE, R. F., and B. T. KELLY (2008): "Dynamic Equicorrelation," Mimeo, Stern School of Business.

ENGLE, R. F., and K. SHEPPARD (2001): "Theorical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH," Mimeo, UCSD.

FIORENTINI, G., E. SENTANA, and G. CALZOLARI (2003): "Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroskedastic Dynamic Regression Models with Student t Innovations," *Journal of Business and Economic Statistics*, 21, 532–546.

GEWEKE, J. (1986): "Modeling the Persistence of Conditional Variances: A Comment," *Econometric Reviews*, 5, 57–61.

GLOSTEN, L. R., R. JAGANNATHAN, and D. E. RUNKLE (1993): "On the Relation Between Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

GOURIEROUX, C. (1997): *ARCH Models and Financial Applications*. Springer Verlag, Berlin.

HARVEY, A., and T. CHAKRAVARTY (2008): "Beta-t-(E)GARCH," Working Paper series, University of Cambridge.

HARVEY, A., E. RUIZ, and N. SHEPHARD (1992): "Unobservable Component Time Series Models with ARCH Disturbances," *Journal of Econometrics*, 52, 129–158.

HARVEY, A., and G. SUCCARAT (2012): "EGARCH models with fat tails, skewness and leverage," CWPE 1236.

HIGGINS, M. L., and A. K. BERA (1992): "A Class of Nonlinear ARCH Models," *International Economic Review*, 33, 137–158.

HOSKING, J. (1980): "The Multivariate Portmanteau Statistic," *Journal of American Statistical Association*, 75, 602–608.

JEANTHEAU, T. (1998): "Strong Consistency of Estimators for Multivariate ARCH models," *Econometric Theory*, 14, 70–86.

JORION, P. (2000): *Value-at-Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, New York.

J.P.MORGAN (1996): *Riskmetrics Technical Document, 4th ed.* J.P.Morgan, New York.

KARIYA, T. (1988): "MTV Model and its Application to the Prediction of Stock Prices," in *Proceedings of the Second International Tampere Conference in Statistics*, ed. by T. Pullila, and S. Puntanen, Finland. University of Tampere.

KRONER, F. K., and V. K. NG (1998): "Modelling Asymmetric Comovements of Asset Returns," *The Review of Financial Studies*, 11, 817–844.

LEE, S. W., and B. E. HANSEN (1994): "Asymptotic Properties of the Maximum Likelihood Estimator and Test of the Stability of Parameters of the GARCH and IGARCH Models," *Econometric Theory*, 10, 29–52.

LI, W., and A. MCLEOD (1981): "Distribution of the residual autocorrelation in multivariate ARMA time series models," *Journal of the Royal Statistical Society B*, 43, 231–239.

LING, S., and M. MCALEER (2003): "Asymptotic Theory for a Vector ARMA-GARCH Model," *Econometric Theory*, 19, 280–310.

LUMSDAINE, R. L. (1996): "Asymptotic Properties of the Quasi Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models," *Econometrica*, 64, 575–596.

NELSON, D. B. (1991): "Conditional Heteroskedasticity in Asset Returns: a New Approach," *Econometrica*, 59, 349–370.

NEWEY, W. K., and D. S. STEIGERWALD (1997): "Asymptotic Bias for Quasi Maximum Likelihood Estimators in Conditional Heteroskedasticity Models," *Econometrica*, 3, 587–599.

PENTULA, S. G. (1986): "Modeling the Persistence of Conditional Variances: A Comment," *Econometric Reviews*, 5, 71–74.

SCHWERT, W. (1990): "Stock volatility and the crash of '87," *Review of financial Studies*, 3, 77–102.

TAYLOR, S. J. (1986): *Modelling Financial Time Series*. J. Wiley & Sons, New York.

TSE, Y. (2000): "A Test for Constant Correlations in a Multivariate GARCH Model," *Journal of Econometrics*, 98, 107–127.

TSE, Y., and A. TSUI (1999): "A Note on Diagnosing Multivariate Conditional Heteroscedasticity Models," *Journal of Time Series Anaysis*, 20, 679–691.

——— (2002): "A Multivariate GARCH Model with Time-Varying Correlations," *Journal of Business and Economic Statistics*, 20, 351–362.

VAN DER WEIDE, R. (2002): "GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model," *Journal of Applied Econometrics*, 17, 549–564.

ZAKOIAN, J.-M. (1994): "Threshold Heteroskedasticity Models," *Journal of Economic Dynamics and Control*, 15, 931–955.