GARCH Models of Volatility*

F. C. Palm

1. Introduction

Until some fifteen years ago, the focus of statistical analysis of time series centered on the conditional first moment. The increased role played by risk and uncertainty in models of economic decision making and the finding that common measures of risk and volatility exhibit strong variation over time lead to the development of new time series techniques for modeling time-variation in second moments.

In line with Box-Jenkins type models for conditional first moments, Engle (1982) put forward the Autoregressive Conditional Heteroskedastic (ARCH) class of models for conditional variances which proved to be extremely useful for analyzing economic time series. Since then an extensive literature has been developed for modeling higher order conditional moments. Many applications can be found in the field of financial time series. This vast literature on the theory and empirical evidence from ARCH modeling has been surveyed in Bollerslev et al. (1992), Nijman and Palm (1993), Bollerslev et al. (1994), Diebold and Lopez (1994), Pagan (1995) and Bera and Higgins (1995). A detailed treatment of ARCH models at a textbook level is also given by Gouriéroux (1992).

The purpose of this chapter is to provide a selective account of certain aspects of conditional volatility modeling in finance using ARCH and GARCH (generalized ARCH) models and to compare the ARCH approach to alternatives lines of research. The emphasis will be on recent developments for instance in multivariate modeling using factor-ARCH models. Finally, an evaluation of the state of the art will be given.

In Section 2, we introduce the univariate and multivariate GARCH models (including ARCH models), discuss their properties and the choice of the functional form and compare them with alternative volatility models. Section 3 will be devoted to problems of inference in these models. In Section 4, the statistical properties of GARCH models, their relationships with continuous time diffusion

* The author acknowledges many helpful comments by G. S. Maddala on an earlier version of the paper.
models and the forecasting volatility will be discussed. Finally in Section 5 we conclude and comment on potentially fruitful directions of future research.

2. GARCH models

2.1. Motivation

GARCH models have been developed to account for empirical regularities in financial data. As emphasized by Pagan (1995) and Bollerslev et al. (1994), many financial time series have a number of characteristics in common. First, asset prices are generally nonstationary, often have a unit root whereas returns are usually stationary. There is increasing evidence that some financial series are fractionally integrated. Second, return series usually show no or little auto-correlation. Serial independence between the squared values of the series however is often rejected pointing towards the existence of nonlinear relationships between subsequent observations. Volatility of the return series appears to be clustered. Heavy fluctuations occur for longer periods. Small values for returns tend to be followed by small values. These phenomena point towards time-varying conditional variances. Third, normality has to be rejected frequently in favor of some thick-tailed distribution. The presence of unconditional excess kurtosis in the series could be related to the time-variation in the conditional variance. Fourth, some series exhibit so-called leverage effects [see Black (1976)], that is changes in stock prices tend to be negatively correlated with changes in volatility. Some series have skewed unconditional empirical distributions pointing towards the inappropriateness of the normal distribution. Fifth, volatilities of different securities very often move together, indicating that there are linkages between markets and that some common factors may explain the temporal variation in conditional second moments. In the next subsection, we shall present several models which account for temporal dependence in conditional variances, for skewness and excess kurtosis.

2.2. Univariate GARCH models

Consider stochastic models of the form

\[ y_t = \epsilon_t \rho_t^{1/2}, \quad (2.1) \]

\[ h_t = \omega_0 + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \quad (2.2) \]

with \( \mathbb{E}\epsilon_t = 0, \) \( \text{Var}(\epsilon_t) = 1, \) \( \omega_0 > 0, \) \( \beta_i \geq 0, \) \( \alpha_i \geq 0, \) and \( \sum_{i=1}^{p} \beta_i + \sum_{i=1}^{q} \alpha_i < 1. \) This is the \((p, q)^{th}\) order GARCH model introduced by Bollerslev (1986). When \( \beta_i = 0, i = 1, 2, \ldots p, \) it specializes to the ARCH\((q)\) model put forward in a seminal paper by Engle (1982). The nonnegativity conditions imply a nonnegative variance, while the condition on the sum of the \( \alpha \)'s and \( \beta \)'s is required for wide sense
stationarity. These sufficient conditions for a nonnegative conditional variance can be substantially weakened as shown by Nelson and Cao (1992). The conditional variance of $y_t$ can become larger than the unconditional variance given by $\sigma^2 = \alpha_0 / (1 - \sum_{i=1}^{p} \beta_i - \sum_{i=1}^{q} \alpha_i)$ if past realizations of $y_t^2$ have been larger than $\sigma^2$.

As shown by Anderson (1992), 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 $t$. Adding the assumption of normality, the model can be written as

$$y_t \mid \Phi_{t-1} \sim N(0, h_t)$$

(2.3)

with $h_t$ being given by (2.2) and $\Phi_{t-1}$ being the set of information available at time $t-1$. Anderson (1994) distinguishes between deterministic, conditionally heteroskedastic, conditionally stochastic and contemporaneously stochastic volatility processes. Loosely speaking, the volatility process is deterministic if the information set ($\sigma$-field) $\Phi$ is identical to the $\sigma$-field of all random vectors in the system up to and including time $t = 0$, the process is conditionally heteroskedastic if $\Phi$ contains information available and observable at time $t-1$, the process is conditionally stochastic if $\Phi$ contains all random vectors up to period $t-1$ whereas the volatility process is contemporaneously stochastic if the information set $\Phi$ contains the random vectors up to period $t$. Notice the order imposed on the information structure of the various volatility representations.

When $\sum_{i=1}^{p} \beta_i + \sum_{i=1}^{q} \alpha_i = 1$, the integrated GARCH (IGARCH) model arises [see Engle and Bollerslev (1986)]. From the GARCH($p, q$) model in (2.2), we obtain that $[1 - \alpha(L) - \beta(L)]y_t^2 = \alpha_0 + [1 - \beta(L)]y_{t-1}$, where $y_t = y_t^2 - h_t$ are the innovations in the conditional variance process and $\alpha(L) = \sum_{i=1}^{q} \alpha_i L^i$ and $\beta(L) = \sum_{i=1}^{p} \beta_i L^i$. The fractionally integrated GARCH model [FIGARCH($p$, $d$, $q$)] proposed by Baillie, Bollerslev and Mikkelsen (1993) arises when the polynomial in the lag operator $L, 1 - \alpha(L) - \beta(L)$, can be factorized as $\phi(L)(1 - L)^d$ where the roots of $\phi(z) = 0$ lie outside the unit circle and $0 \leq d \leq 1$. The FIGARCH model nests the GARCH($p$, $q$) model for $d = 0$, and the IGARCH($p$, $q$) model for $d = 1$. Allowing $d$ to take a value in the interval between zero and one gives additional flexibility that may be important when modeling long-run dependence in the conditional variance.

In the empirical analysis of financial data, GARCH(1,1) or GARCH(1,2) models have often been found to appropriately account for conditional heteroskedasticity. This finding is similar to that low order ARMA models usually describe the dynamics of the conditional mean of many economic time series quite well.

It is important to notice that for the above models positive and negative past values have a symmetric effect on the conditional variance. Many financial series however are strongly asymmetric. Negative equity returns are followed by larger increases in volatility than equally large positive returns. Black (1976) interpreted this phenomenon as the leverage effect according to which large declines in equity values would not be matched by a decrease in the value of debt and would raise the debt to equity ratio. Models such as the exponential GARCH (EGARCH)

Nelson's EGARCH model reads as follows

\[
\ln h_t = \alpha_0 + \sum_{i=1}^{p} \beta_i \ln h_{t-i} + \sum_{i=1}^{q} \alpha_i (e_{t-i} + \phi | e_{t-i} | - \phi E | e_{t-i} |),
\]

(2.4)

where the parameters \(\alpha_0, \alpha_i, \beta_i\) are not restricted to be nonnegative. 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 \(\psi < 0\). Similarly, when fractional integration is allowed for in an exponential GARCH model, the FIEGARCH model is obtained.

The QGARCH model is written by Sentana (1991) as

\[
h_t = \sigma_t^2 + \psi x_{t-q} + x_{t-q}^T A x_{t-q} + \sum_{i=1}^{p} \beta_i h_{t-i},
\]

(2.5)

where \(x_{t-q} = (y_{t-1}, y_{t-2}, \ldots, y_{t-q})^T\). The linear term allows for asymmetry. The off-diagonal elements of \(A\) account for interaction effects of lagged values of \(x_t\) on the conditional variance. The various quadratic variance functions proposed in the literature are nested in (2.5). The augmented GARCH (GAARCH) model of Bera and Lee (1990) assumes \(\psi = 0\). Engle's (1982) ARCH model restricts \(\psi = 0, \beta_i = 0\) and \(A\) to be diagonal. The asymmetric GARCH model of Engle (1990) and Engle and Ng (1993) assumes \(A\) to be diagonal. The linear standard deviation model studied by Robinson (1991) restricts \(\beta_i = 0, \sigma_t^2 = \rho^2, \psi = 2\rho\phi\) and \(A = \phi \rho\), a matrix of rank 1. The conditional variance then becomes \(h_t = (\rho + \rho^T x_{t-q})^2\).

The TGARCH model put forward by Zakoian (1994) is given by

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

(2.6)

where \(y_{t}^+ = \max\{y_t, 0\}\) and \(y_{t}^- = \min\{y_t, 0\}\). It accounts for asymmetries by allowing the coefficients \(\alpha_i^+\) and \(\alpha_i^-\) to differ.

As shown by Hentschel (1994) many members of the family of GARCH models (taking \(p = q = 1\)) can be embedded in a Box-Cox transformation of the absolute GARCH (AGARCH) model

\[
(\sigma_t^\lambda - 1)/\lambda = \alpha_0 + \alpha_1 \sigma_{t-1}^\lambda f^\lambda(e_{t-1}) + \beta(\sigma_{t-1}^\lambda - 1)/\lambda,
\]

(2.7)

where \(\sigma_t = h_t^{1/2}, f^\lambda (e_t) = \max\{e_t - b, 0\} \cdot e_t^\lambda + \min\{e_t - b, 0\} \cdot e_t^\lambda\) is the news impact curve introduced by Pagan and Schwert (1990). For \(\lambda > 1\), the Box-Cox transformation is convex, for \(\lambda < 1\), it is concave. For \(\lambda = \nu = 1\) and \(|c| < 1\) expression (2.7) specializes to become the AGARCH model. The model for the conditional standard deviation suggested by Taylor (1986) and Schwert (1989) arises when \(\lambda = \nu = 1\) and \(b = c = 0\). The exponential GARCH model (2.4) for \(p = q = 1\) arises from (2.7)
when $\lambda = 0, \nu = 1$ and $b = 0$. The TGARCH model for the standard deviation is obtained from (2.7) when $\lambda = \nu = 1, b = 0$ and $| c | \leq 1$. The GARCH model (2.2) arises if $\lambda = \nu = 2$ and $b = c = 0$. Engle and Ng's (1993) nonlinear asymmetric GARCH corresponds to the values of $\lambda = \nu = 2$ and $c = 0$ whereas the GARCH model proposed by Glosten-Jagannathan-Runkle (1993) is obtained when $\lambda = \nu = 2, b = 0$. The nonlinear ARCH model of Higgins and Bera (1992) leaves $\lambda$ free and $\nu$ equal to $\lambda$ with $b = c = 0$. The asymmetric power ARCH (APARCH) of Ding, Granger and Engle (1993) leaves $\lambda$ free and $\nu$ equal to $\lambda, b = 0$ and $| c | \leq 1$. Sentana's (1991) QGARCH is not nested in the specification (2.7). As shown by Hentschel (1994), nesting existing GARCH models in a general specification like (2.7) highlights the relations between these models and offers opportunities for testing sequences of nested hypotheses regarding the functional form for conditional second order moments. Crouhy and Rockinger (1994) put forward the general so-called hysteresis GARCH (HGARCH) model, in which, in addition to a threshold GARCH part, they include a short term, up to a few days, and a long term, up to a few weeks, impact of returns on volatility.

Engle, Lilien and Robins (1987) introduce the ARCH in mean (ARCH-M) model in which the conditional mean is a function of the conditional variance of the process

$$ y_t = g(z_{t-1}, h_t) + h_t^{1/2} e_t, $$

(2.8)

where $z_{t-1}$ is a vector of predetermined variables, $g$ is some function of $z_{t-1}$ and $h_t$ is generated by an ARCH($q$) process. Of course, when $h_t$ follows a GARCH process, expression (2.8) will be a GARCH in mean equation. The most simple ARCH-M model has $g(z_{t-1}, h_t) = \delta h_t$. GARCH in mean models arise in a natural way in theories of finance where for instance $g(z_{t-1}, h_t)$ could denote expected return on some asset with $h_t$ being a measure of risk. The mean equation (2.8) would then reflect the trade-off between risk and expected return. Pagan and Ullah (1988) refer to these models as models with risk terms.

2.3. Alternative models for conditional volatility

Measures of volatility which are not based on ARCH type specifications have also been put forward in the literature. For instance, French et al. (1987) construct monthly stock return variance estimates by taking the average of the squared daily returns and fit ARMA models to these monthly variance estimates. A procedure which uses high frequency data to estimate the conditional variance of low frequency observations does not make efficient use of all the data. Also, the conventional standard errors from the second stage estimation may not be appropriate. Nevertheless, the computational simplicity of this procedure and a related one put forward by Schwert (1989), in which the conditional standard deviation is measured by the absolute value of the residuals from a first step estimate of the conditional mean, makes them appealing alternatives to more complicated ARCH type models for preliminary data analysis.
A related estimator for the volatility may be obtained from the inter-period highs and lows. As shown by Parkinson (1980), a high-low estimator for the variance in a random walk with constant variance and continuous time parameter is more efficient than the conventional sample variance based on the same number of end-of-interval observations. Along these lines, the relationship between volatility and the bid-ask spread for prices could be used to construct variance estimates for returns [see e.g. Bollerslev and Domowitz (1993)]. Similarly, the recent efforts into developing option pricing formulae in the presence of stochastic volatility [see e.g. Melino and Turnbull (1990)] have established a positive relationship between the value of an option and the variance of the underlying security, that could be used to assess the volatility of the security price. Finally, information on the price returns distribution across assets at given points in time could also be used to quantify market volatility.

When deciding on the form of the specification for the conditional variance one has to define the conditioning set of information and to select a functional form for the mapping between the conditioning set and the conditional variance. Usually, the conditioning set is restricted to include past values of the series itself. A simple two-step estimator of the conditional residual variance can be obtained from a regression of the square residuals against their own lagged values [see Davidian and Carroll (1987)]. Pagan and Schwert (1990) show that the OLS estimator is consistent although not efficient. This two-step estimator's role is that of a benchmark which can be computed in a straightforward way. Jump or mixture models possibly combined with a GARCH specification for the conditional variance have been used to describe time-variation in volatility measures, fat-tails and skewness of financial series. In the Poisson jump model it is assumed that upon the arrival of abnormal information a jump occurs in the returns. The number of jumps occurring at time \( t \), \( n_t \), is generated by a Poisson distribution with parameter \( \lambda \). Conditionally on the number of jumps \( n_t \), returns are normally distributed with mean \( n_t \theta \) and variance \( \sigma^2 = \sigma^2_n + n_t \sigma^2_\theta \). The parameter \( \theta \) denotes the expected jump size. The conditional mean and variance of the returns depend on the number of jumps at period \( t \). Additional time dependency could be introduced by assuming that \( \sigma^2_\theta \) is generated by a GARCH-type process.

In the finance literature, stochastic jumps have been usually modeled by means of a Poisson process [see e.g. Ball and Torous (1985), Jorion (1988), Hsieh (1989), Nieuwland et al. (1991) and Ball and Roma (1993)]. Vlaar and Palm (1993) compare the Poisson jump process with the Bernoulli jump model for weekly exchange rate data from the European Monetary System (EMS). The performance of both models is very similar in most instances. Using the Bernoulli process has the advantage that one avoids making a truncation error when cutting off the infinite sum in a Poisson process.

The mixing parameter \( \lambda \) could be allowed to vary over time. For instance Vlaar and Palm (1994) assume that the mixing parameter \( \lambda \) of a Bernoulli jump model for risk premia on European currencies depends on the inflation differential with respect to Germany.
Another way of allowing for time dependence is to assume that the probabilities of being in state 1 during period $t$ differ, depending on whether the economy was in say state 1 or state 2 in period $t-1$. Such a model has been put forward by Hamilton (1989) and applied to exchange rates [Engel and Hamilton (1990)], interest rates [Hamilton (1988)] and stock returns [Pagan and Schwert (1990)].

In Hamilton's basic model, an unobserved state variable $z_t$ can take the values 0 or 1. The transition probabilities from state $j$ in period $t-1$ to state $i$ in period $t$, $P_{ij}$ are constant and given by $P_{11} = p$, $P_{10} = 1 - p$, $P_{00} = q$ and $P_{01} = 1 - q$. As shown by Pagan (1995), $z_t$ evolves as an AR(1) process. Observed returns $y_t$ in Hamilton's model are assumed to be generated by

$$y_t = \beta_0 + \beta_1 z_t + (\sigma^2 + \phi \sigma^2)^{1/2} \epsilon_t,$$

with $\epsilon_t \sim NID(0, \sigma^2)$. The expected values of $y_t$ in the two states are $\beta_0$ and $\beta_0 + \beta_1$ respectively. The variances are $\sigma^2$ and $\sigma^2 + \phi$. The model therefore generates states with high volatility and states with low volatility. Expected returns can also vary across these types of states. The variance of returns conditional on the state in the period $t-1$ can be expressed as

$$\text{Var}(y_t | z_{t-1}) = [\sigma^2 + (1 - q) \phi](1 - z_{t-1}) + [\rho \phi + \sigma^2] z_{t-1}.$$  \hspace{1cm} (2.10)

Quite obviously the conditional variance (2.10) exhibits time dependence.

Hamilton and Susmel (1994) generalize the Markov switching regime model by allowing the disturbances to be ARCH. Their model is called switching regime ARCH model (SWARCH). As in equation (2.9), the conditional mean of the SWARCH model depends linearly on the state variable $z_t$.

The disturbance term of $y_t$ is assumed to follow an autoregressive process of order $p$ with an error $\epsilon_t = \sqrt{g_t} \tilde{u}_t$ where $\tilde{u}_t$ follows an ARCH($q$) process with leverage effects as in the model of Glosten et al. (1993) and $g_t$ is constant factor which differs across regimes. The innovation $\tilde{u}_t$ is assumed to have a conditional student $t$-distribution with mean zero. Transitions between regimes are governed by an unobserved Markov chain. The authors use weekly returns on the value-weighted portfolio of stocks traded on the New York Stock Exchange for the period July 3, 1962 to December 29, 1987. Various ARCH models are compared to SWARCH models allowing for up to four regimes. The SWARCH specification with leverage terms, a conditional student $t$-distribution with a low number of degrees of freedom and allowing for four regimes is found to perform best. Along similar lines using a two-state SWARCH model, Cai (1994) examines the issue of volatility persistence in monthly returns of three-month treasury bills in the period 1964,8 to 1991,11. The persistence in ARCH processes found in previous studies can be accounted for by discrete shifts in the intercept in the conditional variance of the process. Two periods during which a regime shift occurred are the period of the oil crisis 1974,2 – 1974,8 and the period 1979, 9 – 1982,8 associated with a policy change of the Federal Reserve Bank.

Estimates of the conditional variance which do not depend on specific assumptions about the functional form can be obtained using nonparametric

The kernel estimator of a conditional moment of \( y \), denoted by \( g(y) \) with a finite number of conditioning variables \( x \), reads as

\[
\hat{E}(g(y) \mid x) = \frac{1}{\sum_{j=1}^{r} g(y_j) K(x_j - x)} \sum_{j=1}^{r} K(x_j - x),
\]

where \( K \) is a kernel function which smoothes the data. Various types of kernels might be employed. A popular one is the normal kernel which has also been used by Pagan and Schwert (1990)

\[
K(x_j - x) = (2\pi)^{-1/2} \exp\left[-\frac{1}{2}(x_j - x)^T H(x_j - x)\right].
\]

\( H \) is a diagonal matrix with \( k^{th} \) diagonal element set equal to the bandwidth \( \hat{\sigma}_k \tau^{-(k+1)/4} \), with \( \hat{\sigma}_k \) being the standard deviation of \( x_k \), \( k = 1, \ldots q \), with \( q \) being the dimension of the conditioning set.

An alternative nonparametric estimator involves a global approximation of the conditional variance using a series expansion. Among the many existing series expansions, the Flexible Fourier Form (FFF) proposed by Gallant (1981) has been used extensively in finance. The conditional variance is represented as the sum of a low-order polynomial and trigonometric terms constructed from past \( \hat{e}_t \)'s (the residuals from a regression for \( y_t \)). Then, the specification for \( \hat{\sigma}_t^2 \) becomes

\[
\hat{\sigma}_t^2 = \sigma^2 + \sum_{j=1}^{L} \left\{ \alpha_j \hat{e}_{t-j} + \beta_j \hat{e}_{t-j}^2 + \sum_{k=1}^{2} \phi_{jk} \cos(k \hat{e}_{t-j}) + \phi_{jk} \sin(k \hat{e}_{t-j}) \right\}.
\]

In theory, the number of trigonometric terms should tend to infinity, but in practice in terms of significance, it is often not worthwhile to go beyond an order of two. A drawback of (2.13) is the possibility that estimates of \( \hat{\sigma}_t^2 \) can be negative. The estimator in (2.13) has been applied to stock returns by Pagan and Schwert (1990) for \( L = 1 \). The estimate of \( \hat{\sigma}_t^2 \) is roughly constant and similar for the kernel, GARCH(1,2) and FFF estimation methods across most of the range of \( \hat{e}_{t-1} \). Only for large positive and negative values of \( \hat{e}_{t-1} \) the estimators exhibit a different behavior. For negative values of \( \hat{e}_{t-1} \), the volatility estimates increase dramatically. Also, the trigonometric terms in (2.13) appear to be highly significant when tested jointly using a \( \chi^2 \)-test.

The nonparametric estimates of conditional volatility using kernels or Fourier series differ from the parametric estimates for the GARCH, EGARCH and Hamilton model in periods when stock prices fall. In particular, large negative unexpected returns lead to a large increase in volatility. Parametric estimates appear to slowly adjust to large shocks and the effects of these shocks exhibit persistence. The parametric methods use the persistent aspects while the nonparametric methods use the highly nonlinear response to large negative shocks. While the nonparametric estimators of conditional volatility have a much higher
explanatory power than the parametric GARCH, EGARCH and Hamilton models, in particular in explaining asymmetries, they are inefficient compared with parametric methods. This suggests that improvements could be obtained by merging the two approaches to capture a richer set of specifications than are currently employed.

Other nonparametric approaches have been put forward in the literature. Gouriéroux and Monfort (1992) propose to approximate the unknown relation between \( y_t \) and \( e_t \) by a step function of the form

\[
y_t = \sum_{j=1}^{J} \alpha_j I_{A_j}(y_{t-1}) + \sum_{j=1}^{J} \beta_j I_{A_j}(y_{t-1})e_t, \tag{2.14}
\]

where \( A_j, j = 1, 2, \ldots, J \) is a partition of the set of values of \( y_{t-1} \), \( I_{A_j}(y_{t-1}) \) is an indicator variable taking the value 1 when \( y_{t-1} \) is in \( A_j \) and zero otherwise and \( e_t \) is white noise. This model is called Qualitative Threshold Autoregressive Conditionally Heteroskedastic (QTARCH) model.

If regime \( j \) applies to the variable \( y_{t-1} \), the conditional mean and variance of \( y_t \) are given by \( \alpha_j \) and \( \beta_j \) respectively. The process of \( y_t \) is determined by qualitative state variables \( z_t = (I_{A_1}(y_t), \ldots, I_{A_J}(y_t)) \) which are generated by a Markov chain. For instance, the partition \( A_1, \ldots, A_J \) may correspond to the different stages of expansion and contraction of the financial market. By refining the partition \( A_1, \ldots, A_J \) sufficiently, one can use (2.14) to approximate more complex specifications for the conditional mean and variance of \( y_t \). Alternatively the conditional variance specification could be refined by adding a GARCH term. The pseudo-maximum likelihood estimators of \( \alpha_j \) and \( \beta_j \) are the sample mean and variance computed for regime \( j \). The QTARCH model approximates the conditional mean and variance by step functions whereas the TARCH model of Zakoian (1994) relies on a piecewise linear approximation of the conditional variance function. The nonparametric kernel estimators smooth the conditional moments and the FFF estimators approximate the conditional moments using functions which are smoother than piecewise linear or step functions. Along similar lines, Engle and Ng (1993) use linear splines to estimate the shape of the response to news. Their procedure is called partially nonparametric (PNP) as the long memory component is modeled as parametric and the relationship between news and volatility is treated nonparametrically.

Among semiparametric methods extensively used in analyzing dependencies in financial data, we should mention the semiparametric (SNP) models based on a series expansion with a Gaussian VAR leading term proposed by Gallant and Tauchen (1989).

Assume that the conditional distribution of an \( N \times 1 \) vector \( y_t \) given the entire past depends only on a finite number \( L \) of lagged values of \( y_t \), denoted by \( x_{t-1} = (y'_{t-L}, y'_{t-L+1}, \ldots, y'_{t-1})' \) which is a vector of length \( LN \). The procedure consists of approximating the conditional density of \( y_t \) given \( x_{t-1} \) by a truncated Hermite expansion which has the form of a polynomial in \( z_t \) times the standard normal density, where \( z_t \) is the centered and scaled value of \( y_t \), \( z_t = R^{-1}(y_t - b_0 - Bx_{t-1}) \).
The truncated expansion is the semiparametric model. The conditional SNP density for $x_i$ given $x_{i-1}$ is approximated by

$$f(x_i | x_{i-1}) = \frac{\left[ \sum_{\alpha} \alpha \pi(x_i | x_{i-1}) \right]^{\frac{1}{2}}}{\int \left[ \sum_{\alpha} \alpha \pi(x_i | x_{i-1}) \right]^{\frac{1}{2}} \phi(u) du}, \quad (2.15)$$

where $\phi$ denotes the standard Gaussian density, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)'$, $x_i = \pi(x_i | x_{i-1})^\alpha$, which is of degree $|\alpha| = \sum_{i=1}^N |\alpha_i|$, $\alpha_\alpha = \sum_{\beta} \alpha_\beta x_i^\beta$, $\beta = (\beta_1, \beta_2, \ldots, \beta_N)'$, $|\beta| = \sum_{\beta} |\beta_i|$, $x_i^\beta = \pi(x_i | x_{i-1})^\beta$ and $K_\alpha$ and $K_\beta$ are positive integers. The conditional density of $y_i$ given $x_{i-1}$ is $h(y_i | x_{i-1}) = f[R^{-1}(y_i - \beta_0 - Bx_{i-1}) | x_{i-1}] / \det(R)$.

As pointed out by Gallant and Tauchen (1989), by increasing $K_\alpha$ and $K_\beta$ simultaneously, an SNP model will yield arbitrarily accurate approximations to a class of models which includes fat-tail distributions ($t$-like distributions) and skewed distributions. As the stationary distribution of the ARCH models is not known in closed form, one cannot say that the ARCH model belongs to the above class. However, the stationary distribution of the ARCH model has fat tails and only a finite number of moments as the $t$-distribution. Conditionally, the variances of ARCH and SNP models are polynomials in a finite number of lags. One might therefore expect that the conditional density of an ARCH model could be approximated arbitrarily closely by SNP for large $K_\alpha$ and $K_\beta$. For large $L$, this may also be true for GARCH models, of which the conditional variance is a polynomial in an infinite number of lags.

An alternative to using the ARCH framework is to assume the changing variance to follow some latent process. This leads to a stochastic variance or volatility (SV) model [see e.g. Ghijsels et al. (1995)]. Assuming for the sake of simplicity of exposition that the drift parameter is zero, a simple SV model for returns $y_i$ has been proposed by Taylor (1986)

$$y_i = \sigma_i \exp(\alpha_i/2), \sigma_i \sim NID(0,1), \quad (2.16)$$

$$\alpha_{i+1} = \alpha_0 + \phi \alpha_i + \eta_t, \eta_t \sim NID(0, \sigma_\eta^2),$$

where the random variables $\omega_t$ and $\eta_t$ are independent.

This model has been used by Hull and White (1987) for instance in pricing foreign currency options. Its time series properties are discussed by Taylor (1986, 1994). The statistical properties of SV models are documented in Taylor (1994) who denotes these models as autoregressive variance (ARV) models. A major difficulty arises with the estimation of SV models which are nonlinear and not conditionally Gaussian. Many estimation methods such as the generalized method of moments (GMM) or quasi maximum likelihood method (QML) used to estimate SV models are inefficient. But methods relying on simulation-based techniques make it possible to perform Bayesian estimation or classical likelihood analysis [see e.g. Kim and Shephard (1994)]. Currently, only few studies compare the performance of the GARCH and SV approaches to modeling volatility. Ruiz
(1993) compares the GARCH(1,1), EGARCH(1,0) and ARV(1) models when applied to daily exchange rates from 1/10/1981 to 28/6/1985 for the Pound sterling, Deutsche mark, Yen and Swiss franc vis-à-vis the U.S. dollar. Within sample performance of the three models is very similar. When the models are used to forecast out-of-sample volatility, the ARCH models exhibit severe biases which do not occur for the SV volatilities.

For daily and weekly returns on the S&P 500 index over the periods 7/3/1962 to 12/31/1987 and 7/11/1962 to 12/30/1992 respectively, Kim and Shephard (1994) conclude that a simple first order SV model fits the data as well as the popular ARCH models. For daily data on the S&P 500 index for the years 1980 to 1987, Danielsson (1994) finds that the EGARCH(2,1) model performs better than ARCH(5), GARCH(1,2), IGARCH(1,1,0) models. It also outperforms a simple SV model estimated by simulated maximum likelihood. The difference between a dynamic SV model and the EGARCH log-likelihood values is 25.5 in favor of the SV model with four parameters whereas the EGARCH model has five parameters.

2.4. Multivariate GARCH models

With the exception of the SNP model, the models presented in the Sections 2.2 and 2.3 are univariate. The analysis of many issues in asset pricing and portfolio allocation requires a multivariate framework.

Consider an \( N \times 1 \) vector stochastic process \( \{ y_t \} \) which we write as

\[
y_t = \Omega_t^{1/2} \epsilon_t, \tag{2.17}
\]

with \( \epsilon_t \) being an \( N \times 1 \) i.i.d. vector with \( \mathbb{E} \epsilon_t = 0 \) and \( \text{Var}(\epsilon_t) = I_N \) and \( \Omega_t \) being the \( N \times N \) covariance matrix of \( y_t \) conditional on information available at time \( t \).

In a multivariate linear GARCH\((p,q)\) model, Bollerslev, Engle and Woldridge (1988) assume that \( \Omega_t \) is given by a linear function of the lagged cross squared errors and lagged values of \( \Omega_t \)

\[
\text{vech}(\Omega_t) = \alpha_0 + z \sum_{i=1}^{q} A_i \text{vech}(\epsilon_{t-i} \epsilon_{t-i}^T) + \sum_{i=1}^{p} B_i \text{vech}(\Omega_{t-i}) \tag{2.18}
\]

where \( \text{vech}(\cdot) \) denotes the operator that stacks the lower portion of an \( N \times N \) matrix as an \( N(N + 1)/2 \) by 1 vector. In (2.18), \( \alpha_0 \) is an \( N(N + 1)/2 \) vector and the \( A_i \) and \( B_i \)'s are \( N(N + 1)/2 \) matrices. The number of unknown parameters in (2.18) equals \( N(N + 1)/2 + N(N + 1)(p + q)/2 \) and in practice some simplifying assumptions have to be imposed to achieve parsimony. For instance, Bollerslev et al. (1988) use the diagonal GARCH\((p,q)\) model assuming that the matrices \( A_i \) and \( B_i \) are diagonal. Other representations include the constant conditional correlation model used by Baillie and Bollerslev (1990) and Vlaar and Palm (1993) who assume the conditional variances to be GARCH processes.

Conditions for the parametrization (2.18) to ensure that \( \Omega_t \) is positive definite for all values of \( \epsilon_t \) are difficult to check in practice. Engle and Kroner (1995)
propose a parametrization of the multivariate GARCH process to which they refer as the BEKK (Baba, Engle, Kraft and Kroner) representation

$$
\Omega_t = C_0^* C_0^* + \sum_{k=1}^K \sum_{i=1}^q A_{ik}^* e_{t-i} e_{t-i}^* A_{ik}^* + \sum_{k=1}^K \sum_{i=1}^q G_{ik}^* \Omega_{t-i} G_{ik}^*,
$$

(2.19)

where $C_0^*$, $A_{ik}^*$ and $G_{ik}^*$ are $N \times N$ parameter matrices with $C_0^*$ being triangular and the summation limit $K$ determines the generality of the process. The covariance matrix in (2.19) will be positive definite under weak conditions. Also this representation is sufficiently general that it includes all positive definite diagonal representations and most positive definite vec representations of the form (2.18). The representation (2.19) is usually more parsimonious in terms of numbers of parameters than (2.18). Given that the two parametrizations are found to be equivalent under quite general circumstances, the BEKK parametrization might be preferred because then positive definiteness is ensured quite easily.

Engle, Ng and Rothschild (1990) have proposed the factor-ARCH model as a parsimonious structure for the conditional covariance matrix of asset excess returns. These models incorporate the notion that risk on financial assets can be decomposed in a limited number of common factors $f_t$ and an asset specific (idiosyncratic) disturbance term. A factor structure arises from the Arbitrage Pricing Theory (APT) although APT does not imply that the number of factors is finite. The factor-ARCH model is used by Engle, Ng and Rothschild (1990) to model interest rate risk while in a companion paper, Ng et al. (1992) consider risk premia and anomalies to the capital asset pricing model (CAPM) on the U.S. stock market. Diebold and Nerlove (1989) apply a one factor model to exchange rates whereas King, Sentana and Wadhwani (1994) analyze the links between national stock markets using a factor model.

The factor model reads as follows

$$
y_t = \mu_t + Bf_t + \epsilon_t,
$$

(2.20)

with $y_t$ being an $N \times 1$ vector of returns, $\mu_t$ is an $N \times 1$ vector of expected returns, $B$ is a $N \times k$ matrix of factor loadings, $f_t$ is a $k \times 1$ vector of factors with conditional covariance matrix $A_t$ and $\epsilon_t$ denotes an $N \times 1$ vector of idiosyncratic shocks with conditional covariance matrix $\Psi_t$. The factors and the idiosyncratic shocks are uncorrelated. The conditional covariance matrix of $y_t$ is then given by

$$
\Omega_t = B A_t B' + \Psi_t.
$$

(2.21)

When $\Psi_t$ is constant and $A_t$ has constant (possibly zero) off-diagonal elements, the covariance matrix $\Omega_t$ can be expressed as

$$
\Omega_t = \sum_{i=1}^k b_i b_i^*/A_t + \Psi,
$$

(2.22)

where $b_i$ denotes the $i - th$ column of $B$ and $\Psi$ groups the off-diagonal elements of $A_t$ with the constant elements of the covariance matrix of $\epsilon_t$. As pointed out by
Engle et al. (1990) the model in (2.22) is observationally equivalent to a similar model with constant $\lambda$'s but time-varying $b$'s. An implication of the factor model (2.22) is that if $k < N$, we can construct $N - k$ portfolio's of assets, i.e. linear combinations of $y_t$, which have constant variance. There are $k$ portfolios which have $\lambda_t$ plus a constant as conditional variance.

The factor model (2.20) has to be completed by specifying processes for the factor variances. One could for instance assume that $\lambda_t$ is generated by a univariate GARCH process. Applying a one factor model to weekly data on the log differences for seven exchange rates vis à vis the US dollar for the period July 1973 to August 1985, Diebold and Nerlove (1989) assume that the single common factor has a variance $\lambda_t = \alpha_0 + \theta \sum_{i=1}^{15}(13 - i)\epsilon_{t-i}^2$. Notice that their covariance matrix is of dimension seven by seven but contains only nine unknown parameters, cf. those of $\bar{\Psi}$, $\bar{\alpha}_0$ and $\bar{\theta}$. By imposing a linearly decreasing pattern on the ARCH-coefficients, they achieve a substantial reduction of the number of parameters to estimate. A GARCH(1,1) specification would instead yield geometrically decreasing ARCH coefficients.

An alternative proposed by Engle et al. (1990) consists in assuming that the returns of each of the $k$ factor-representing portfolios follow a GARCH process. For $i = 1, \ldots, k$, the conditional variance of the $i-th$ portfolio is then given by

$$\phi_i \Omega \phi_j = \omega_i + \alpha_i (\phi_i \epsilon_{t-1})^2 + \beta_i \phi_i \Omega_{t-1} \phi_i,$$  \hspace{1cm} (2.23)

where for simplicity reason a GARCH(1, 1) model is assumed and $\phi_i$ is an $N \times 1$ vector of weights of the portfolio. The conditional variances of the portfolios differ from $\lambda_t$ by a constant term only, i.e. $\phi_i \Omega \phi_j = \lambda_t + \phi_i \bar{\Psi} \phi_j$, which together with (2.23) can be substituted into (2.22) so as to express the conditional covariance matrix $\Omega_t$ in (2.22) in terms of the conditional portfolio variances. Notice that $\phi_i \beta_i = 1$ and $\phi_i \beta_j = 0, j \neq i$. While the factor-GARCH model has theoretically appealing features, its estimation requires highly nonlinear methods. Maximum likelihood estimation has been considered among others by Lin (1992). Also, an identification issue has to be resolved when the factor portfolios are not directly observed before the model can be estimated [see Sentana (1992)]. In particular the factor representing portfolios have to be identified. In some instances, it is appropriate to assume that the factor representing portfolios are known and observed. For example, Engle et al. (1990) explain the monthly returns on Treasury bills with maturities ranging from one to twelve months and the value-weighted index of NYSE-AMSE stocks, for the period from August 1964 to November 1985. They select two factor-representing portfolios one of which having equal weights on each of the bills and zero weight on the stock index and the other having zero weights on the bills and all weight on the stock index. Models with observed factor-representing portfolios can be consistently estimated in two-steps. One can first estimate the univariate models for the portfolios. Using the estimates obtained in the first step, the factor loadings can be estimated consistently up to a sign as individual assets have a variance which is linear in the factor variances with coefficients that are equal to the squared factor loadings.
King et al. (1994) estimate a multivariate factor model as in (2.20) from monthly data on US dollar excess returns for 16 national stock markets for the period 1970.1 to 1988.10 using the maximum likelihood method. They assume that the risk premium \( \mu \), can be expressed as \( \mu_t = B A_t \xi_t \), with \( A_t \) being a diagonal matrix and \( \xi_t \) being a \( k \times 1 \) vector of constant parameters representing the price of risk for each factor. King et al. (1994) consider the model for \( k = 6 \) with 4 observed and 2 unobserved factors. The observable factors represent the unanticipated shocks to asset returns. These shocks are estimated as the common factors extracted from a four-factor model applied to the residuals from a vector autoregression for \( x_i \), a set of 10 observed macroeconomic variables. The variances of the common and idiosyncratic terms are assumed to follow univariate GARCH(1,1) processes in which the past squared values of the factors are replaced by their linear projection given some available information set. Notice that when the covariance matrix of the factor-GARCH model depends on prior unobservables, the return components have a conditionally stochastic volatility representation [see Anderson (1992), Harvey et al. (1992)].

A major finding is that only a small proportion of the covariances between national stock markets and their time-variation can be explained by observed factors. Conditional second moments are explained to a large extent by unobserved factors. This finding underlines the usefulness of models allowing for unobservable factors in explaining volatility within markets and volatility spillovers between markets. The application in King et al. (1994) also illustrates the appropriateness and feasibility of the use of factor models to explain the time-dependence in second order moments of a multivariate time series of dimension 16. While it was possible to jointly estimate the factor model with some 200 parameters, the authors had to estimate the vector autoregression for \( x_i \) separately in a first step. Given that the dimension of the parameter space of multivariate factor-GARCH models will usually be high, two-step estimation procedures will be a feasible alternative to fully joint estimation procedures based on the likelihood principle.

2.5. Persistence in the conditional variance

For high-frequency time series data, the conditional variance estimated using a GARCH\((p,q)\) process (2.2) often exhibits persistence, that is \( \sum_{i=1}^p \beta_i + \sum_{i=1}^q \alpha_i \) is close to one. When this sum is equal to one, the IGARCH model arises. This means that current information remains of importance when forecasting the conditional variance for all horizons. The unconditional variance does not exist in that case. Bollerslev (1986) has shown that under normality, the GARCH process (2.2) is wide sense stationary with unconditional variance \( \text{var}(y_t) = \alpha_0 (1- \sum_{i=1}^p \beta_i - \sum_{i=1}^q \alpha_i)^{-1} \) and \( \text{cov}(y_t, y_s) = 0 \) for \( t \neq s \) if and only if \( \sum_{i=1}^p \beta_i + \sum_{i=1}^q \alpha_i < 1 \). Nelson (1990a) and Bougerol and Picard (1992) prove that the IGARCH model is strictly stationary and ergodic but not covariance stationary.

Similarly, as shown in Bollerslev and Engle (1993), the multivariate GARCH \((p,q)\) process (2.18) is covariance stationary if and only if the roots of the
characteristic polynomial \( \det[I - A(\lambda^{-1}) - B(\lambda^{-1})] = 0 \) lie inside the unit circle. In that case, there will be no persistence in the variance. On the other hand, if some eigenvalues lie on the unit circle, shocks to the conditional covariance matrix remain important for forecasts of all horizons. If the eigenvalues are outside the unit circle, the effect of a shock to the covariance matrix will explode over time. Notice that the above conditions on the roots of the characteristic polynomial also apply to the BEKK model (2.19), as shown by Engle and Kroner (1995). In many empirical studies of financial data using univariate GARCH\((p, q)\) models, the estimated parameters are found to have a sum close to one. A detailed survey of the literature can be found in Bollerslev, Chou and Kroner (1992). The multivariate \( k \) factor model (2.20) with a GARCH\((p, q)\) process of the form (2.23) for the factors will be covariance stationary if the portfolios and \( \epsilon_t \) are covariance stationary.

In line with the concept of cointegration between a set of variables, Bollerslev and Engle (1993) put forward a definition of co-persistence in variance. The basic idea is that several time series may show persistence in the variance while at the same time some linear combinations of the variables may exhibit no persistence in the variance. Bollerslev and Engle (1993) derive necessary and sufficient conditions for co-persistence in the variances of a multivariate GARCH\((p, q)\) process. In practice, co-persistence in the variances allows one to construct portfolios with stationary volatilities from the assets which have nonstationary return volatilities.

The finding of unit roots in multivariate GARCH models has led to new developments in factor-ARCH models. Engle and Lee (1993) formulate a factor model of the form of the King et al. (1994) within which they allow for permanent IGARCH\((1, 0, 1)\) and transitory GARCH\((1, 1)\) components in the volatilities.

Engle and Lee (1993) apply several variants of the component model to daily returns on the CRSP value-weighted index and fourteen individual stocks of large U.S. companies for a sample period from July 1, 1962 to December 31, 1991. Their major empirical finding is that the persistence of individual return volatilities is due to the persistence of both market volatility (assumed to be a common factor) and idiosyncratic volatilities of individual stocks. These results imply that the hypothesis that stock return volatility is co-persistent with market volatility is rejected when market shocks are assumed not to affect idiosyncratic volatility.

Using a factor-component-GARCH model with observed factors, Palm and Urbain (1995) also find significant persistence in the common and idiosyncratic factors volatilities using daily observations on returns of stock price indices for Europe, the Far-East and North-America for the period February 1982–August 1995.

While the use of factor-component-GARCH models is still in its infancy, the empirical finding of persistence in return volatilities [see e.g. French, Schwert and Stambaugh (1987), Chou (1988), Pagan and Schwert (1990), Ding et al. (1993) and Engle and Gonzalez-Rivera (1991)], common factor and/or idiosyncratic factor volatilities raises a number of important questions. For instance is the finding of persistence in volatilities in agreement with the stationarity assumption for asset returns which has often been made in the literature? Would finance
theory not predict that a nonstationarity in the volatility leads to a non-
stationarity in asset returns? What is the precise form of the persistence in vol-
atilities and in the return series? Should it be modeled as a unit root in the
permanent component of the conditional variances or should one allow for fra-
ctional integration or should it be modeled as regimes switches as e.g. in Cai
(1994) or in Hamilton and Susmel (1994)? There is increasing evidence that return
series exhibit fractional integration [see e.g. Baillie (1994)]. The difficulty of em-
pirically distinguishing between persistence arising from unit roots or from fra-
ctional differencing is due to the low power of many existing testing pro-
cedures.

3. Statistical inference

3.1. Estimation and testing

GARCH models are usually estimated by the method of maximum likelihood
(ML) or quasi-maximum likelihood (QML). In some applications, the general-
ized method of moments (GMM) has been used [see e.g. Glosten et al. (1993)]. Stoc-
astic volatility models were usually estimated by GMM. More recently indirect
inference methods [see e.g. Gouriéroux and Monfort (1993) and Gallant et al.
(1994)] have been advocated and used to estimate stochastic volatility models.
Bayesian methods have been developed for volatility models [see e.g. Jacquier et al.
(1994) for the estimation of stochastic volatility models and Geweke (1994) for the
estimation of stochastic volatility and GARCH models]. For simplicity reason, we
discuss ML estimation of the GARCH(1,1) model (2.1) and (2.2) under the as-
sumption that $\epsilon_t$ is distributed as $\mathcal{N}(0,1)$. The log-likelihood function $L$ for $T$
observations on $y_t$, denoted by $y = (y_1, y_2, ..., y_T)'$, can be written as

$$L(y | \theta) = \sum_{t=1}^{T} l_t,$$  \hspace{1cm} (3.1)

where $l_t = c - \frac{1}{2} \ln h_t - \frac{1}{2} y_t^2 / h_t$ with $\theta = (a_0, a_1, \beta_1)'$, $h_1 = \sigma^2 = a_0 / (1 - a_1 - \beta_1)$
and $h_t$ given by (2.2) for $t > 1$.

Given initial values for the parameter vector $\theta$, the log-likelihood function (3.1)
can be evaluated by computing $h_t$, $t = 1, 2, ..., T$ recursively and substituting the
values in (3.1). Standard numerical algorithms can be used to compute the
maximum of (3.1). As is well-known, under regularity conditions given for in-
stance in Crowder (1976), the value of $\theta$ which maximizes $L$, $\hat{\theta}_{ML}$, is consistent,
 asymptotically normally distributed and efficient

$$\sqrt{T} (\hat{\theta}_{ML} - \theta) \overset{d}{\sim} \mathcal{N}(0, \text{Var}(\hat{\theta}_{ML})), \hspace{1cm} (3.2)$$

where $\text{Var}(\hat{\theta}_{ML}) = - [T^{-1} \sum_{t=1}^{T} \mathbb{E} \partial^2 L / \partial \theta \partial \theta']^{-1}$. The asymptotic covariance
matrix of $\hat{\theta}_{ML}$ can be consistently estimated by the inverse of the Hessian matrix
associated with (3.1), evaluated at $\hat{\theta}_{ML}$. A proof of the consistency and asymptotic
normality of the ML-estimator in GARCH(1,1) and IGARCH(1,1) models is given by Lumsdaine (1992) under the condition that $E[\ln(\varepsilon_t^2 + \beta_1)] < 0$. The existence of finite fourth moments of $\varepsilon_t$ is not required. Unlike models with a unit root in the conditional mean, the ML estimator in models with and without a unit root in the conditional variance have the same limiting distribution.

As shown by Weiss (1986) for time series models with ARCH errors, by Bollerslev and Wooldridge (1992) and Gouriéroux (1992) for GARCH processes, the quasi-ML estimator or the pseudo-ML estimator of $\theta$ is obtained by maximizing the normal log-likelihood function (3.1) although the true probability density function is non-normal. Under regularity conditions the QML-estimator has the following asymptotic distribution

$$\sqrt{T}(\hat{\theta}_{\text{QML}} - \theta) \overset{D}{\sim} N(0, B^{-1}AB^{-1})$$  \hspace{1cm} (3.3)

where $A = E_0[\partial L_t/\partial \theta \cdot \partial L_t/\partial \theta']$ is the covariance matrix of the score vector of $L$ and $B = -E_0[\partial^2 L_t/\partial \theta \partial \theta']$ where $E_0$ denotes the expectation conditional on the true probability density function for the data. Of course, if the latter is the normal distribution, the asymptotic distributions in (3.2) and (3.3) will be identical. Lee and Hansen (1994) prove consistency and asymptotic normality of the QML estimator of the Gaussian GARCH(1,1) model. The disturbance scaled by its conditional standard deviation need not be normally distributed nor independent over time. The GARCH process may be integrated $a_t + \beta_1 = 1$ and even explosive $a_t + \beta_1 > 1$ provided the conditional fourth moment of the scaled disturbance is bounded. In finite samples, for symmetric departures from conditional normality the QML has been found close to the exact ML-estimator in a simulation study by Bollerslev and Wooldridge (1992). For non-symmetric conditional true distributions, both in small and large samples the loss of efficiency of QML compared to exact ML can be quite substantial. Semi-parametric density estimation as proposed by Engle and Gonzalez-Rivera (1991) using a linear spline with smoothness priors will then be an attractive alternative to QML.

With respect to ML and QML methods to estimate GARCH models, some comments can be made. First, although GARCH generates fat tails in the unconditional distribution, when combined with conditional normality, it does not fully account for excess-kurtosis present in many financial data. The student $t$-distribution with the number of degrees of freedom to be estimated has been used by several authors. Other densities which have been used in the estimation of GARCH models are the normal-Poisson mixture [see e.g. Jorion (1988), Nieuwland et al. (1991)], the normal-lognormal mixture distribution [e.g. Hsieh (1989)] and the generalized error distribution [see e.g. Nelson (1991)] and the Bernoulli-normal mixture [Vlaar and Palm (1993)]. De Vries (1991) proposes to use a GARCH-like process with conditional stable distribution which models the clustering of volatility, has fat tails and an unconditional stable distribution.

Second, for some models such as the regression model under conditional normal ARCH-disturbances, the information matrix is block-diagonal [see e.g. Engle (1982)]. The implications are important in that the regression coefficients...
and the ARCH parameters can be estimated separately without loss of asymptotic efficiency. Also, their variances can be obtained separately. These results have been generalized by Linton (1993) who shows that the parameters of the conditional mean are adaptive in the sense of Bickel when the errors follow a stationary ARCH(q) process with an unknown conditional density which is symmetric about zero. In other words, estimating the unknown score function using the kernel method based on the normal density function yields parameter estimates of the conditional mean which have the same asymptotic distribution as the ML estimator based on the true distribution. This block-diagonality does not hold for the ARCH-M model as there the conditional mean of a series depends on parameters of the conditional variance process. Also for an EGARCH disturbance process, the block-diagonality of the information matrix fails to hold.

Indirect inference put forward by Gourieroux and Monfort (1993) and the efficient method of moments by Gallant et al. (1994) will be attractive when it is difficult to apply QML or ML but it is possible to estimate some function of the parameters of interest from the data.

The indirect estimator has been used by Engle and Lee (1994) to estimate diffusion models of stochastic volatility. As a starting point, they estimate GARCH(1,1) models from daily returns on the S&P 500 Index for the period 1991,1–1990,9. The resulting QML estimates for $\hat{\theta}$ are used to estimate the parameters of the underlying diffusion model for the asset price $p_t$ and its conditional variance $\sigma_t^2$

\begin{align}
    (a) & \quad y_t = \mu \, dt + \sigma_t \, dw_{yt} \\
    (b) & \quad d\sigma_t^2 = \phi (\sigma_t^2 - \delta^2) \, dt + \xi \sigma_t^2 \, dw_{\sigma t} \\
    (c) & \quad \text{correl}(dw_{yt}, dw_{\sigma t}) = \rho
\end{align}

with $y_t = d \ln p_t$, $dw_y$ and $dw_{\sigma}$ being Wiener processes, using the relationships which match the first and second order conditional moments of the GARCH model and the diffusion model (see Nelson (1990b)): $\phi = s_0, \phi = (1 - \alpha(1 - \beta_1)) dt$, $\xi = \alpha \sqrt{\kappa - 1} dt$, $\delta = 1$ with $\kappa$ being the conditional kurtosis of the shocks of the GARCH model. Indirect estimation based on estimates of a discrete time GARCH model appears to be an appropriate way to estimate the parameters of the underlying diffusion process.

To estimate stochastic volatility models, Gallant et al. (1994) use an indirect method based on the score of two auxiliary models. Both auxiliary models assume an SNP density as given in (2.15). When the SNP density is in the form of an ARCH model with conditionally homogeneous non-Gaussian innovations, it is termed nonparametric ARCH model because it is similar to the nonparametric ARCH process considered by Engle and Gonzalez-Rivera (1991). In the second model, the homogeneity constraint is dropped and the model is called the fully nonparametric specification. The SNP models are estimated by QML.

Gallant et al. (1994) use daily observations on the S&P Composite Index for the period 1928–1987 to estimate a univariate model and daily observations for the period 1977–1992 to estimate a trivariate model for the S&P NYSE Index, the
DM/$ exchange rate and the tree month Eurodollar interest rate. The stochastic volatility model is found to be able to match the ARCH part of the nonparametric ARCH score for stock prices and interest rates. However it does not match the moments of the distribution of the innovations. For the exchange rate series, the stochastic volatility model fails to fit the ARCH part.

Testing for the presence of ARCH($q$) has also been extensively considered in the literature. A simple and frequently used test of the hypothesis $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_q = 0$ against the alternative $H_1 : \alpha_1 \geq 0, ... , \alpha_q \geq 0$ with at least one strict inequality is the Lagrange multiplier (LM) test proposed by Engle

$$LM = \frac{1}{2} f_0' \sigma^2(z)^{-1} z' f_0,$$

where $z_t = (1, y_{t-1}^2, ..., y_{t-q}^2)'$, $z = (z_1, ..., z_T)'$ and $f_0$ is the column vector of $(y^2 / \sigma_0^2 - 1)$.

An asymptotically equivalent statistic is $LM = TR^2$, where $R^2$ is the squared multiple correlation between $f_0$ and $z$ and $T$ is the sample size. This is also the $R^2$ of a regression of $y_t^2$ on an intercept and $q$ lagged values of $y_t^2$. As shown by Engle (1982), a two-sided LM test has an asymptotic $\chi^2$-distribution with $q$ degrees of freedom. Demos and Sentana (1991) report critical values for the one-sided LM test which are robust to non-normality. A difficulty in constructing LM tests for GARCH disturbances is that the block of the information matrix whose inverse is required, is singular, as pointed out by Bollerslev (1986). This is due to the fact that under the null hypothesis, $\beta_1$ in the GARCH(1,1) model is not identified. Lee (1991) has shown how this difficulty can be avoided and that the LM tests for ARCH and GARCH errors are identical.

Lee and King (1993) derive a locally most mean powerful (LMP)-based score (LBS) test for the presence of ARCH and GARCH disturbances. The test is based on the sum of the scores evaluated at the null hypothesis and nuisance parameters replaced by their ML estimates. In the absence of nuisance parameters, the test is LMP. The sum of the scores is then standardized by dividing it by its large sample standard error. The resulting test statistic has an asymptotic $N(0,1)$ distribution. The test statistics used to test against an ARCH($q$) process can also be used to test against a GARCH($p,q$) process. In small samples, the LBS test appears to have better power than the LM-test and its asymptotic critical values were found to be at least as accurate.

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 statistics associated with $H_0 : \alpha_1 = 0$ and $\beta_1 = 0$ against $H_1 : \alpha_1 \geq 0$ or $\beta_1 \geq 0$ with at least one strict inequality do not have a $\chi^2$-distribution with two degrees of freedom as the standard assumption that the true parameter value under $H_0$ does not lie on the border of the parameter space does not hold. A LR test which uses a $\chi^2$-distribution with two degrees of freedom can be shown to be conservative [see e.g. Kode and Palm (1986)]. Also, the problem of lack of identification of some parameters mentioned above can lead to a break down of standard Wald and LR testing procedures. These ARCH statistics test for specific forms of conditional
heteroskedasticity. Many tests however have been designed to test for general
departures from independently, identically distributed random variables. For
instance, the BDS test put forward by Brock, Dechert and Scheinkman (1987)
tests for general nonlinear dependence. Its power against ARCH alternatives is
similar to that of the LM-ARCH test [see e.g. Brock, Hsieh and LeBaron (1991)].
For other alternatives, the power of the BDS test may be higher. The application
by Bera and Lee (1993) of the White Information Matrix (IM) criterion to the
linear regression model with autoregressive disturbances lead to a generalization
of Engle’s LM test for ARCH where ARCH processes are specified as random
coefficient autoregressive models. Several authors have noted that ARCH can be
given a random coefficient interpretation [see e.g. Tsay (1987)]. Bera, Lee and
Higgins (1992) point out the dangers of tackling specification problems one at a
time rather than considering them jointly and provide a framework for analyzing
autocorrelation and ARCH simultaneously. That such a framework is needed has
been illustrated by e.g. Diebold (1987) in a convincing way by showing that in the
presence of ARCH, standard tests for serial correlation will lead to over-rejection
of the null hypothesis. Notice that the presence of ARCH could be interpreted in
several ways such as nonnormality (excess kurtosis, skewness for asymmetric
ARCH) [see e.g. Engle (1982)] and nonlinearity [see e.g. Higgins and Bera (1992)].

Recently Bollerslev and Wooldridge (1992) have developed robust LM tests for
the adequacy of the jointly parametrized mean and variance. Their test is based
on the gradient of the log-likelihood function evaluated at the constrained QML-
estimator and can be computed from simple auxiliary regressions. Only first
derivatives of the conditional mean and variance functions are required. The
authors present simulation results revealing that in most cases, the robust test
statistics compare favorably to nonrobust (standard) Wald and LM tests.

This conclusion is in line with findings by Lumsdaine (1995) who compares
GARCH(1,1) and IGARCH(1,1) models in a simulation study of the finite-
sample properties of the ML estimator and related test statistics. While the
asymptotic distribution is found to be well approximated by the estimated
$\hat{t}$-statistics, parameter estimators are skewed for finite sample size, Wald tests have
the best size, the standard LM test is highly oversized but versions that are robust
to possible nonnormality perform better.

Various model diagnostics have been proposed in the literature. For instance,
Li and Mak (1994) examine the asymptotic distribution of the squared standard-
dized residual autocorrelations from a Gaussian process with time-dependent
conditional mean and variance estimated by ML. The residuals are then stan-
dardized by dividing them by their conditional standard deviation and sub-
tracting their sample mean. The conditional mean and variance of the process
can be nonlinear functions of the information available at time $t$. These functions
are assumed to have continuous second order derivatives. When the data gener-
ating process is ARCH($q$), a Box-Pierce type portmanteau test based on aut-
correlations of squared standardized residuals of order $r$ up to $M$ will have an
asymptotic $\chi^2$-distribution with $M - r$ degrees of freedom when $r > q$. These
types of diagnostics are very useful for checking the adequacy of the model.
Specific kinds of hypotheses can arise in multivariate GARCH models. For instance, GARCH can be a common feature to several time series. Engle and Kozicki (1993) define a feature that is present in a group of time series as common to those series if there exists a nonzero linear combination of the series that does not have the feature. As an example, consider the bivariate version of the factorARCH model in (2.20) with one factor and constant idiosyncratic factor covariance matrix. If the variance of \( f_t \) follows a GARCH process, the series \( y_{it} \) will also be GARCH, but the linear combination \( y_{it} - b_1 y_{it} b_2 y_{it} \) will have a constant conditional variance. In this example, the series \( y_{1t} \) and \( y_{2t} \) share a common feature of the form of a common factor with a time-varying conditional variance. Engle and Kozicki (1993) put forward tests for common features. Engle and Susmel (1993) apply the procedure to test for ARCH as common feature in international equity markets. The approach is as follows. First, test for the presence of ARCH in the individual time series. Second, if the ARCH effects are significant in both series, consider the linear combination \( y_{1t} - \delta y_{2t} \) and regress its squared value on lagged squared values and lagged cross products of the series \( y_{it} \) up to lag \( q \) and minimize \( TR^2(\delta) \) over the coefficient \( \delta \). If instead of two series, a set of \( k \) series is considered, \( \delta \) becomes a \( (k-1) \times 1 \) vector. As shown by Engle and Kozicki (1993) the test statistic which minimizes \( TR^2(\delta) \) with respect to \( \delta \) has a \( \chi^2 \)-distribution with degrees of freedom given by the number of lagged squared values included in the regressions minus \( (k-1) \). Engle and Susmel (1993) applied the test to weekly returns on stock market indexes for 18 major stock markets in the world over the period January 1980 to January 1990. They found two groups of countries, one of European countries and one of Far East countries which show similar time-varying volatility. The common feature tests therefore confirm the existence of a common factor-ARCH structure for each group.

4. Statistical properties

In this section, we shall summarize the main results about the statistical properties of GARCH models and give appropriate references to the literature.

4.1. Moments

Bollerslev (1986) has shown that under conditional normality, the GARCH process (2.2) is wide sense stationary with \( \text{E}y_t = 0 \) and \( \text{var}(y_t) = a_0[1 - \alpha(1) - \beta(1)]^{-1} \) and \( \text{cov}(y_t, y_s) = 0 \) for \( t \neq s \) if and only if \( \alpha(1) + \beta(1) < 1 \). For the GARCH(1,1) model given in (2.2), a necessary and sufficient condition for the existence of the \( 2 \ r-th \) moment is \( \sum_{j=0}^{\infty} \gamma_j \alpha_j \beta_j^{-1} < 1 \) when \( a_0 = 1 \) and \( \alpha_j = \pi_{1m}^{2j-1} \). Bollerslev (1986) also provides a recursive formula for even moments of \( y_t \) when \( p = q = 1 \). The fourth moment of a conditionally normal GARCH(1,1) variable will be \( \text{E}\gamma_t^4 = 3(\text{E}\gamma_t^2)^2[1 - (\beta_1 + \alpha_1)^2]/[1 - (\beta_1 + \alpha_1)^2 - 2\alpha_1^2] \) if it exists. As a result of the symmetry of the normal distribution, odd moments are zero if they exist. These results extend results for the ARCH(q) process given in Engle (1982). The condition given above is sufficient for strict stationarity but not necessary.
As shown in Krengel (1985), strict stationarity of a vector ARCH process \( y_t \) is equivalent to the conditions that \( \Omega_t = \Omega(y_{t-1}, y_{t-2}, ...) \) being measurable and trace \( \Omega_t \Omega_t' < \infty \) a.s. [see also Bollerslev et al. (1994)]. Moment boundedness i.e. \( E[\text{trace} (\Omega_t \Omega_t')] \) being finite for some \( r > 0 \) implies trace \( \Omega_t \Omega_t' < \infty \) a.s. Nelson (1990a) has shown that for the GARCH(1,1) model (2.2), \( y_t \) is strictly stationary if and only if \( E[\ln(\beta_1 + \alpha_i \varepsilon_t^2)] < 0 \) with \( \varepsilon_t \) being i.i.d. (not necessarily conditional normal) and \( y_t^2 \) nondegenerate. This requirement is much weaker than \( \alpha_1 + \beta_1 < 1 \). He also has shown that the IGARCH(1,1) model without drift converges almost surely to zero, while in the presence of a positive drift it is strictly stationary and ergodic. Extensions to general univariate GARCH\((p,q)\) processes have been obtained by Bougerol and Picard (1992).

4.2. GARCH and continuous time models

GARCH models are nonlinear stochastic difference equations which can be estimated more easily than the stochastic differential equations used in the theoretical finance literature to model time-varying volatility. In practice, observations are usually recorded at discrete points in time so that a discrete time model or a discrete time approximation to a continuous model will have to be used in statistical inference. Nelson (1990b) derives conditions for the convergence of stochastic difference equations, among which ARCH processes, to stochastic differential equations as the length of the interval between observations \( h \) goes to zero. He applies these results to the GARCH(1,1) and the EGARCH model. Nelson (1992) investigates the properties of estimates of the conditional covariance matrix generated by a misspecified ARCH model. When a diffusion process is observed at discrete time intervals of length \( h \), the difference between an estimate of its conditional instantaneous covariance matrix based on a GARCH(1,1) model or on an EGARCH model and the true value converges to zero in probability as \( h \downarrow 0 \). The required regularity conditions are that the distribution does not have fat tails and that the conditional covariance matrix moves smoothly over time. Using high-frequency data, misspecified ARCH models can yield accurate estimates of volatility. In a way, the GARCH model which averages squared values of variables can be interpreted as a nonparametric estimate of the conditional variance at time \( t \). Discrete time models can also be approximated by continuous time diffusion models. Different ARCH models will in general have different diffusion limits. As shown by Nelson (1990b), the continuous limit may yield convenient approximations for forecast and other moments when a discrete time model leads to intractable distributions.

Nelson and Foster (1994) examine the issue of selecting an ARCH process to consistently and efficiently estimate the conditional variance of the diffusion process generating the data. They obtain the approximate distribution of the measurement error resulting from the use of an approximate ARCH filter. Their result allows to compare the efficiency of various ARCH filters and to characterize asymptotically optimal ARCH conditional variance estimates. They derive optimal ARCH filters for three diffusion models and examine the filtering
properties of several GARCH models. For instance, if the data generating process is given by the diffusion equations (3.4) with independent Brownian motions \( \mu = 0 \) and \( \delta = 1 \), the asymptotically optimal filter for \( \sigma_t^2 \) sets the drift for \( \gamma_t \) \( \tilde{w} = \mu \) and the conditional variance

\[
\tilde{\sigma}_{t+h}^2 = w_h + (1 - \phi h - \alpha h^{1/2})\tilde{\sigma}_t^2 + h^{1/2} \alpha \varepsilon_{t+h}^2
\]

(4.1)

with \( \varepsilon_{y,t+h} = h^{-1/2}[y_{t+h} - y_t - \mathbb{E}(y_{t+h} - y_t)] \), \( w = \kappa \phi \) and \( \zeta = \xi / \sqrt{2} \).

The asymptotically optimal filter for (3.4) with independent Brownian motions therefore is the GARCH(1,1) model. When \( w_y \) and \( w_x \) are correlated, the GARCH(1,1) model (4.1) is no longer optimal. Nelson and Foster (1994) show that the nonlinear asymmetric GARCH model proposed by Engle and Ng (1993) fulfills the optimality conditions in this case. Nelson and Foster (1994) also study the properties of various ARCH filters when the data are generated by a discrete time near-diffusion process. Their findings have important implications for the choice of a functional form for the ARCH filter in empirical research.

The use of continuous record asymptotics has greatly enhanced our understanding of the relationship between continuous time stochastic differential equations and discrete time ARCH models as the sampling frequency increases. Similarly, issues of temporal aggregation play an important role in modeling time-varying volatilities, in particular when an investigator has the choice between using data observed with a high frequency or using observations sampled less frequently.

More efficient parameter estimates may be obtained from the high frequency data. On other occasions, an investigator may be interested in the parameters of the high frequency model while only low frequency observations are available.

The temporal aggregation problem has been addressed by Diebold (1988) who has shown that the conditional heteroskedasticity disappears in the limit as the sampling frequency decreases and that in the case of flow variables the marginal distribution of the low frequency observations converges to the normal distribution.

Drost and Nijman (1993) study the question whether the class of GARCH processes is closed under temporal aggregation when either stock or flow variables are modeled.

The question can be answered if some qualifications are made. Three definitions of GARCH are adopted. The sequence of variables \( y_t \) in (2.2) is defined to be generated by a strong GARCH process if \( \alpha_t, \alpha_i, i = 1, 2, \ldots, q \) and \( \beta_t, i = 1, 2, \ldots, p \) can be chosen such that \( \varepsilon_t = y_t h_t^{-1/2} \) is i.i.d. with mean zero and variance 1. The sequence \( y_t \) is said to be semi-strong GARCH if \( \mathbb{E}(y_t | y_{t-1}, y_{t-2}, \ldots) = 0 \) and \( \mathbb{E}[\varepsilon_t^2 | y_{t-1}, y_{t-2}, \ldots] = h_t \) whereas it is weakly GARCH if \( \mathbb{E}(y_t | y_{t-1}, y_{t-2}, \ldots) = 0 \) and \( \mathbb{E}[\varepsilon_t^2 | y_{t-1}, y_{t-2}, \ldots] = h_t \) where \( h_t \) denotes the best linear predictor in terms of a constant, \( y_{t-1}, y_{t-2}, \ldots, y_{t-2}^2, \ldots \).

The main finding of Drost and Nijman (1993) is that the class of symmetric weak GARCH processes for either stock or flow variables is closed under tem-
poral aggregation. This means that if the high frequency process is symmetric (weak) GARCH, the low frequency process will also be symmetric weak GARCH. The parameters of the conditional variance of the low frequency process depend upon the mean, variance and kurtosis of the corresponding high frequency process. The conditional heteroskedasticity disappears as the sampling frequency increases for GARCH processes with $\sum_{i=1}^{\pi} a_i + \sum_{i=1}^{\mu} \beta_i < 1$. The class of strong or semi-strong GARCH processes is generally not closed under temporal aggregation suggesting that strong or semi-strong GARCH processes will often be approximations only to the data generating process if the observation frequency does not exactly correspond with the frequency of the data generating process.

In a companion paper, Drost and Werker (1995) study the properties of a continuous time GARCH process, i.e. a process of which the increments $X_{t+h} - X_t, t \in hN$ are weak GARCH for each fixed time interval $h > 0$. Obviously in the light of the results by Drost and Nijman (1993) a continuous time GARCH process cannot be strong or semi-strong GARCH as the classes of these processes are not closed under temporal aggregation.

The assumption of an underlying continuous time GARCH process leads to a kurtosis in excess of three for the associated discrete GARCH models, implying thick tails. Drost and Werker (1995) show how the parameters of the continuous time diffusion process can be identified from the discrete time GARCH parameters. The relations between the parameters of the continuous and discrete time models can be used to estimate the diffusion model from discrete time observations in a fairly straightforward way.

Nijman and Sentana (1993) complement the results of Drost and Nijman (1993) by showing that contemporaneous aggregation of independent univariate GARCH processes yields a weak GARCH process. Then they generalize this finding by showing that a linear combination of variables generated by a multivariate GARCH process will also be weak GARCH. The marginal processes of multivariate GARCH models will be weak GARCH as well. Finally, from simulation experiments the authors conclude that in many instances, estimators which are ML under the assumption that the process is strong GARCH with conditional normal distribution converge to values close to the weak GARCH parameters as the sample size increases.

The findings on temporal and contemporaneous aggregation of GARCH processes indicate that linear transformations of GARCH processes are generally only weak GARCH.

4.3. Forecasting volatility

Time series models are often built to generate out-of-sample forecasts. The issue of forecasting in models with time-dependent conditional heteroskedasticity has been investigated by several authors. Engle and Kraft (1983) and Engle and Bollerslev (1986) obtain expressions for the multi-step forecast error variance for time series models with ARCH and GARCH errors respectively. Bollerslev
(1986), Granger, White and Kamstra (1989) are concerned with the construction of one-step-ahead forecast intervals with time-varying variances. Baillie and Bollerslev (1992) consider a single equation regression model with ARMA-GARCH disturbances, for which they derive the minimum MSE forecast. They also derive the moments of the forecast error distribution for the dynamic model with GARCH(1,1) disturbances. These moments are used in the construction of forecast intervals using the Cornish-Fisher asymptotic expansion. Geweke (1989) obtains the multi-step ahead forecast error density for linear models with ARCH disturbances by numerical integration within a Bayesian context.

Nelson and Foster (1995) derive conditions under which for data observed at high frequency a misspecified ARCH model performs well in forecasting of a time series process and its volatility. In line with the conditions for successful filtering obtained by Nelson and Foster (1994), the basic requirement is that the ARCH model correctly specifies the functional form of the first two conditional moments of all state variables.

To illustrate the construction of estimates of the forecast error variance, consider a stationary AR(1) process

$$y_t = \phi y_{t-1} + u_t,$$

where $u_t = e_t h_t^{1/2}$ is a GARCH(1,1) process as in (2.2). The minimum MSE forecast of $y_{t+s}$ at period $t$ is $E_t(y_{t+s}) = \phi^s y_t$. The forecast error $w_t = y_{t+s} - \phi^s y_t$ can be expressed as $w_t = u_{t+s} + \phi u_{t+s-1} + \ldots + \phi^{s-1} u_{t+1}$. Its conditional variance at time $t$

$$\text{Var}(w_t) = \sum_{i=0}^{s-1} \phi^i E_t(u_{t+s-i}^2), s > 0,$$

can be computed recursively. The GARCH(1,1) process for $u_t$ leads to an ARMA representation for $u_t^2$ [see Bollerslev (1986)]

$$u_t^2 = \alpha_0 + (\alpha_1 + \beta_1) u_{t-1}^2 - \beta_1 u_{t-1} + v_t,$$

with $v_t = u_t^2 - h_t$. The expectations on the r.h.s. of (4.3) can be readily obtained from expression (4.4)

$$E_t(h_{t+s}) = E_t(u_{t+s}^2) = \alpha_0 + (\alpha_1 + \beta_1) E_t(u_{t+s-1}^2), s > 1,$$

as shown by Engle and Bollerslev (1986). As the forecast horizon increases, the optimal forecast converges monotonically to the unconditional variance $\alpha_0/(1 - \alpha_1 - \beta_1)$. For the IGARCH(1,1) model, shocks to the conditional variance are persistent and $E_t(h_{t+s}) = \alpha_0(s - 1) + h_t$. The expression (4.5) can be used as a forecast of future volatility. Baillie and Bollerslev (1992) derive an expression for the conditional MSE of $E_t(h_{t+s})$ as a forecast of the conditional variance at period $t + s$. 
5. Conclusions

In this paper, we have surveyed the literature on modeling time-varying volatility using GARCH processes. In reviewing the vast number of contributions we have put most emphasis on recent developments.

In less than fifteen years since the path-breaking publication of Engle (1982) much progress has been made in understanding GARCH models and in applying them to economic time series. This progress has drastically changed the way in which empirical time series research is carried out. At the same time, the statistical properties of time series, in particular financial time series which were not accounted for by existing models have led to new developments in the field of volatility modeling. The finding of skewness and skewed correlations defined as $[(\sum_i y_i^2 y_{i+k})/(Te^\lambda \sqrt{k})]$ fostered the development of asymmetric GARCH model. The presence of excess kurtosis in GARCH models with conditional normally distributed innovations has led to the use of student-GARCH models and GARCH-jump models. Persistence in conditional variances was modeled using variance component models with a stochastic trend component.

The finding of time-variation in conditional covariances and correlations resulted in the development of multivariate GARCH and factor-GARCH models. Factor-GARCH models have several attractive features. First, they can be easily interpreted in terms of economic theory (factor models like the arbitrage pricing theory have been used extensively in finance). Second, they allow for a parsimonious representation of time-varying variances and covariances for a high-dimensional vector of variables. Third, they can account for both observed and unobserved factors. Fourth, they have interesting implications for common features of the variables. These common features can be tested in a straightforward way. Fifth, they have appeared to fit well in several instances.

As has become apparent in Section 2, the functional forms of time-varying volatility has attracted a lot of attention by researchers to an extent where one wonders whether the returns from designing new GARCH specification are still positive. While some specifications are close if not perfect substitutes for others, the results by Nelson and Foster on the use of GARCH as filters to estimate the conditional variance of an underlying diffusion model put the issue of choosing a functional form for the GARCH model in a new perspective. For a given diffusion process some GARCH model will be an optimal (efficient) filter whereas others with similar properties might not be optimal. The research by Nelson and Foster (1994) suggests that prior knowledge about the form of the underlying diffusion process will be useful when choosing the functional form for the GARCH model.

As shown by Anderson (1992, 1994) GARCH processes belong to the class of deterministic, conditionally heteroskedastic volatility processes. The ease of evaluating the GARCH likelihood function and the ability of the GARCH specification to accommodate the time-varying volatility, in particular to yield a flexible, parsimonious representation of the correlation found for the squared values of many series (comparable to the parsimonious representation of condi-
tional means using ARMA schemes) has led to the widespread use of GARCH models. The history of the stochastic volatility model is brief. This model has been put forward as a parsimoniously parameterized alternative to GARCH models. While one of its attractive features is the low number of parameters needed to fit the time-variation of volatility of many time series, likelihood-based inference of stochastic volatility models requires numerical integration or the use of the Kalman filter. As mentioned in Section 3, many of these problems have by now been resolved. The statistical properties of GARCH models and stochastic volatility models differ. Comparisons of these models [see for instance Danielson (1994), Hsieh (1991), Jacquier et al. (1995) and Ruiz (1993)] on the basis of financial time series led to the conclusion that these models put different weights on various moments functions. The choice among these models will very often be an empirical question.

In other instances, a GARCH model will be preferred because it yields an optimal filter of the variance of the underlying diffusion model. Factor-GARCH models with unobserved factors will lead to stochastic volatility components when one has to condition on the latent factors. The borders between the two classes of volatility models are expected to lose sharpness.

Results on temporal aggregation of GARCH processes indicate that weak GARCH is the most common case. For reasons of aggregation, models relying on strong GARCH are at best approximations to the data generating process, a situation in which a pragmatic view of using data information to select the model might be the most appropriate.

Topics for future research are improving our understanding and the modeling of relationships between volatilities of different series and markets. Multivariate GARCH, factor-GARCH and stochastic volatility models will be used and extended. Questions regarding the nature and the transmission of persistence in volatility from one series to another, the transmission of persistence in volatility into the conditional expected return will have to receive more attention in the future. Finally, statistical methods for testing and estimating volatility models and for forecasting volatility will be on the research agenda for a while. In particular, nonparametric and semiparametric methods appear to open up new perspectives to modeling time-variation in conditional distributions of economic time series.

References


Lumsdaine, R. L. (1992). Asymptotic properties of the quasi-maximum likelihood estimator in GARCH(1,1) and 1GARCH(1,1) models. Unpublished manuscript, Department of Economics, Princeton University.


