ARCH MODELS: PROPERTIES, ESTIMATION AND TESTING

Anil K. Bera

University of Illinois at Urbana-Champaign

and

Matthew L. Higgins

University of Wisconsin-Milwaukee

Abstract. The aim of this survey paper is to provide an account of some of the important developments in the autoregressive conditional heteroskedasticity (ARCH) model since its inception in a seminal paper by Engle (1982). This model takes account of many observed properties of asset prices, and therefore, various interpretations can be attributed to it. We start with the basic ARCH models and discuss their different interpretations. ARCH models have been generalized in different directions to accommodate more and more features of the real world. We provide a comprehensive treatment of many of the extensions of the original ARCH model. Next we discuss estimation and testing for ARCH models and note that these models lead to some interesting and unique problems. There have been numerous applications and we mention some of these as we present different models. The paper includes a glossary of the acronyms for the models we describe.

Keywords. ARCH; GARCH; nonlinearity; nonnormality; persistence; random coefficient model; volatility.

1. Introduction

The history of ARCH models is indeed a very short one, for it was introduced by Robert Engle only a decade ago. Within this brief period, however, the ARCH literature has grown in a spectacular fashion. The numerous applications of ARCH models defies observed trends in scientific advancements. Usually, applications lag theoretical developments, but Engle's original ARCH model and its various generalizations have been applied to numerous economic and financial data series of many countries, while it has seen relatively fewer theoretical advancements.

The concept of ARCH might be only a decade old, but its roots go far into the past, possibly as far as Bachelier (1900), who was the first to conduct a rigorous study of the behavior of speculative prices. There was then a period of long silence. Mandelbrot (1963a,b, 1967) revived the interest in the time series properties of asset prices with his theory that 'random variables with an infinite population variance are indispensable for a workable description of price changes' (cf. 1963b, p. 421). His observations, such as unconditional

0950–0804/93/04 305–62 JOURNAL OF ECONOMIC SURVEYS Vol. 7, No. 4

distributions have thick tails, variances change over time and large (small) changes tend to be followed by large (small) changes of either sign, are 'stylized facts' for many economic and financial variables. Figures 1, 2 and 3 present three typical data series on price changes. These are, respectively the weekly rate of return on the U.S. dollar/British pound exchange rate, changes in the three month treasury bill rate and the growth rate of the NYSE monthly composite index. The first noticeable thing is that for all three series, the means appear to be constant, while the variances change over time. In particular, for the treasury bill rate, there is a dramatic increase in the variance in the late seventies and early eighties. Sample statistics from these series overwhelmingly support Mandelbrot's other stylized facts.

Prior to the introduction of ARCH, researchers were very much aware of changes in variance, but used only informal procedures to take account of this. For example, Mandelbrot (1963a) used recursive estimates of the variance over time and Klien (1977) took five period moving variance estimates about a ten period moving sample mean. Engle's (1982) ARCH model was the first formal model which seemed to capture the stylized facts mentioned above.

The ARCH model is useful not only because it captures some stylized facts, but also because it has applications to numerous and diverse areas. For example, it has been used in asset pricing to test the CAAPM, the I-CAPM, the CCAPM and the APT; to develop volatility tests for market efficiency and to estimate the

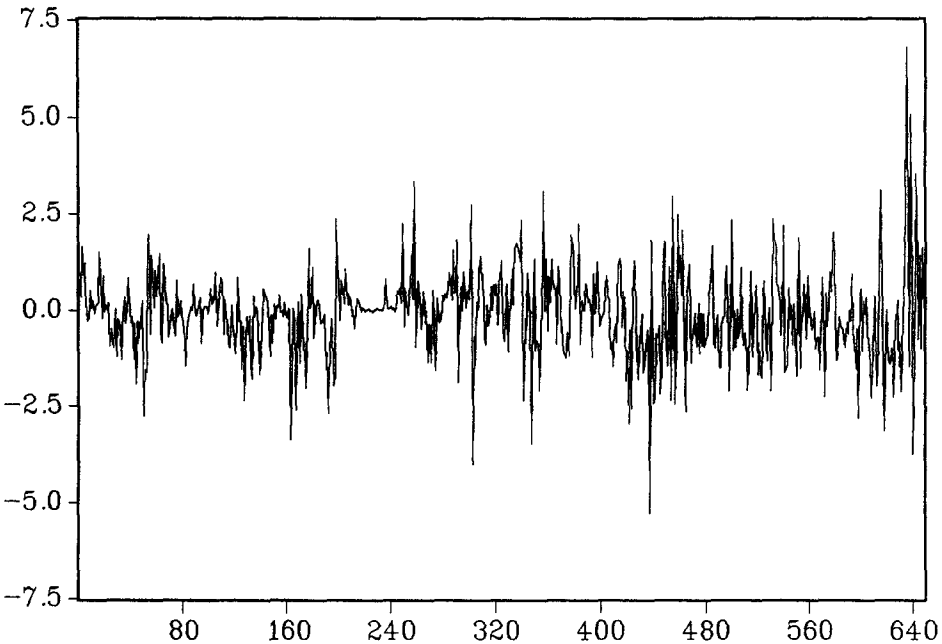

Figure 1. Weekly rate of return on the dollar/pound exchange rate.

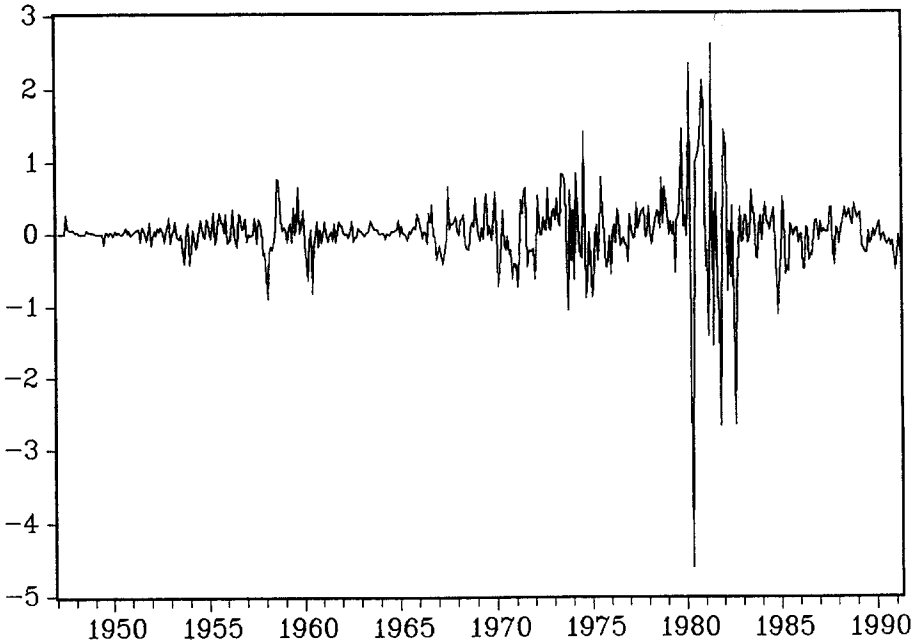

Figure 2. First difference of the 3 month treasury bill rate.

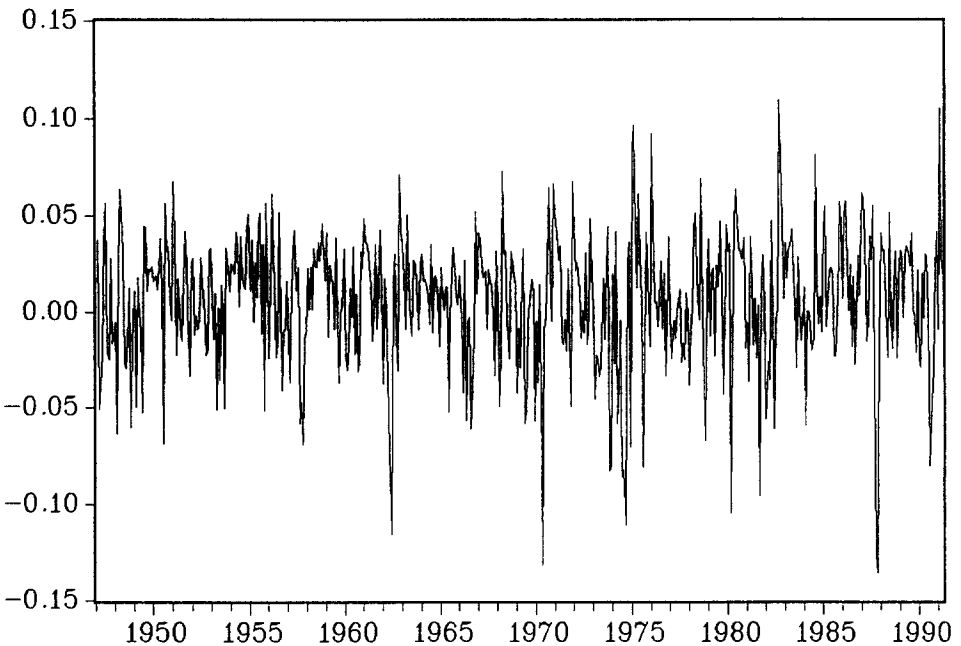

Figure 3. Monthly rate of return on the NYSE composite index.

time varying systematic risk in the context of the market model. It has been used to measure the term structure of interest rates; to develop optimal dynamic hedging strategies; to examine how information flows across countries, markets and assets; to price options and to model risk premia. In macroeconomics, it has been successfully used to construct debt portfolios of developing countries, to measure inflationary uncertainty, to examine the relationship between exchange rate uncertainty and trade, to study the effects of central bank interventions, and to characterize the relationship between the macroeconomy and the stock market.

The literature on ARCH is so vast, it is almost impossible to provide a comprehensive review. There are already a few survey papers on this topic. In particular, we would refer the readers to Engle and Bollerslev (1986) and Bollerslev, Chou and Kroner (1992). The latter paper noted several hundred papers that apply the ARCH methodology to various financial markets. Some recent references to the very rapidly growing bibliography include Bekaert (1992), Bollerslev and Hodrick (1992), Duffee (1992), Koedijk, Stork and De Vries (1992) and Ng and Pirrong (1992), just to name a few. The purpose of this review paper is rather modest. Our aim is to provide an informal account of recent theoretical advances and their impact on applied work. It should be mentioned that our use of the term 'ARCH' does not refer to Engle's original model. By ARCH, we mean the phenomena of conditional heteroskedasticity in general and all models to capture this phenomena.

The plan of the paper is as follows. The basic ARCH models are described in the next section. As these models capture various stylized facts, they can be given different interpretations, and these are discussed in section 3. It has been found that the basic ARCH models are unable to capture all observed phenomena, such as the leverage effect, excess kurtosis and the high degree of nonlinearity. Generalizations of the basic ARCH models to capture these phenomena are the subject matter of section 4. Forecasting with ARCH models is treated in section 5. The following sections, 6 and 7, review further generalizations, such as multivariate ARCH and ARCH-in-mean (ARCH-M) models. In sections 8 and 9, we discuss estimation and testing of ARCH models. The last section concludes the paper with a few remarks. At the end of the paper, we include a complete glossary of the acronyms for the ARCH models which we describe in the survey.

2. Autoregressive conditional heteroskedasticity

In this section, we introduce the original ARCH model of Engle (1982). We begin by defining the ARCH process, and heuristically describe its properties. We emphasize the properties of the ARCH model which make it appealing for modeling the volatility of economic time series. Subsequently, we introduce the generalized ARCH (GARCH) model of Bollerslev (1986), which provides a parsimonious parameterization for the conditional variance. The properties of the ARCH process are then formally characterized by describing its

unconditional moments. We also discuss how aggregating an ARCH process over time effects the moments of the process.

2.1. *Definition of the process*

An ARCH process can be defined in a variety of contexts. We will define it in terms of the distribution of the errors of a dynamic linear regression model. The dependent variable y_t is assumed to be generated by

$$y_t = x_t'\xi + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{2.1}$$

where x_t is a $k \times 1$ vector of exogenous variables, which may include lagged values of the dependent variable, and ξ is a $k \times 1$ vector of regression parameters. The ARCH model characterizes the distribution of the stochastic error ε_t conditional on the realized values of the set of variables $\Psi_{t-1} = \{y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \ldots\}$. Specifically, Engle's (1982) original ARCH model assumes

$$\varepsilon_t \mid \Psi_{t-1} \sim N(0, h_t) \tag{2.2}$$

where

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2, \tag{2.3}$$

with $\alpha_0 > 0$ and $\alpha_i \geqslant 0$, $i = 1, \ldots, q$, to ensure that the conditional variance is positive. Note that since $\varepsilon_{t-i} = y_{t-i} - x_{t-i}'\xi$, $i = 1, \ldots q$, h_t is clearly a function of the elements of Ψ_{t-1}.

The distinguishing feature of the model (2.2) and (2.3) is not simply that the conditional variance h_t is a function of the conditioning set Ψ_{t-1}, but rather it is the particular functional form that is specified. Episodes of volatility are generally characterized as the clustering of large shocks to the dependent variable. The conditional variance function (2.3) is formulated to mimic this phenomena. In the regression model, a large shock is represented by a large deviation of y_t from its conditional mean $x_t'\xi$, or equivalently, a large positive or negative value of ε_t. In the ARCH regression model, the variance of the current error ε_t, conditional on the realized values of the lagged errors ε_{t-i}, $i = 1, \ldots, q$, is an increasing function of the magnitude of the lagged errors, irrespective of their signs. Hence, large errors of either sign *tend* to be followed by a large error of either sign. And similarly, small errors of either sign *tend* to be followed by a small error of either sign. The order of the lag q determines the length of time for which a shock persists in conditioning the variance of subsequent errors. The larger the value of q, the longer the episodes of volatility will tend to be.

A linear function of lagged squared errors, of course, is not the only conditional variance function that will produce clustering of large deviations. Any monotonically increasing function of the absolute values of the lagged errors will lead to such clustering. However, since variance is expected squared deviation, a linear combination of lagged squared errors is a natural measure of the recent trend in variance to translate to the current conditional variance h_t.

Alternative formulations of the conditional variance function have been found to be useful and these formulations will be discussed in depth in section 4.1.

To illustrate the characteristic appearance of ARCH data, we generate artificial samples from (2.2). An explicit generating equation for an ARCH process is

$$\varepsilon_t = \eta_t \sqrt{h_t}, \tag{2.4}$$

where $\eta_t \sim \text{IID } N(0, 1)$ and h_t is given by (2.3). Since h_t is a function of the elements of Ψ_{t-1}, and therefore, is fixed when conditioning on Ψ_{t-1}, it is clear that ε_t as given in (2.4) will be conditionally normal with $E(\varepsilon_t \mid \Psi_{t-1}) = \sqrt{h_t} E(\eta_t \mid \Psi_{t-1}) = 0$ and $\text{Var}(\varepsilon_t \mid \Psi_{t-1}) = h_t \text{ Var}(\eta_t \mid \Psi_{t-1}) = h_t$. Hence, the process specified by (2.4) is identical to the ARCH process (2.2). The generating equation (2.4) reveals that ARCH rescales an underlying Gaussian innovation process η_t by multiplying it by the conditional standard deviation which is a function of the information set Ψ_{t-1}. First, for comparison when ARCH is not present in the data, in Figure 4 we present a plot of 500 realizations of $\varepsilon_t = \eta_t$, setting $h_t = 1$ by imposing $\alpha_0 = 1$ and $\alpha_i = 0$ for $i = 1, \ldots, q$. The displayed data is simply Gaussian white noise, the process usually assumed for the errors in a linear model. Then, using the same η_t's shown in Figure 4, Figures 5 and 6 are plots of $\varepsilon_t = \eta_t \sqrt{h_t}$ for which the h_t's are respectively

$$h_t = 0.1 + 0.9\varepsilon_{t-1}^2 \tag{2.5}$$

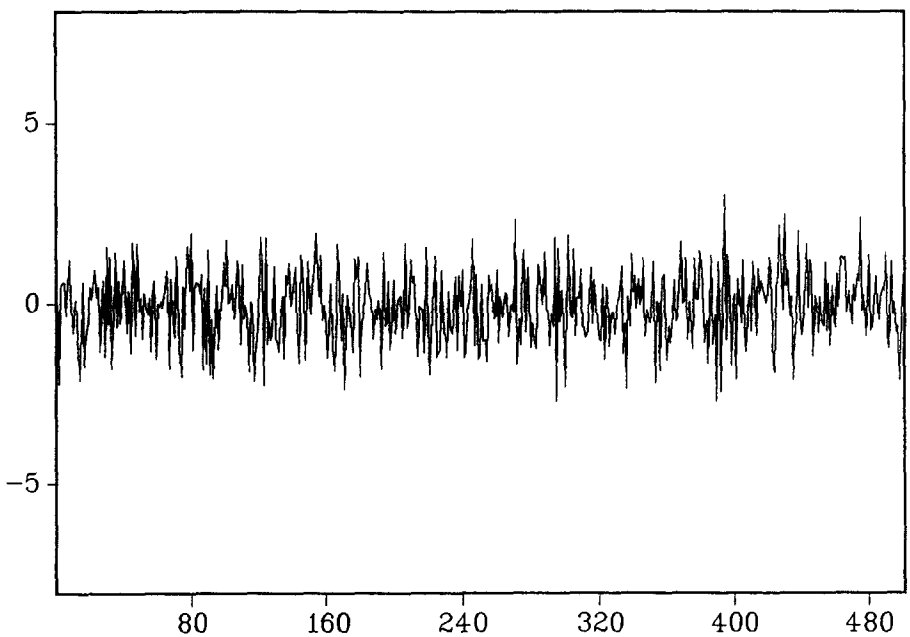

Figure 4. Simulated Gaussian white noise.

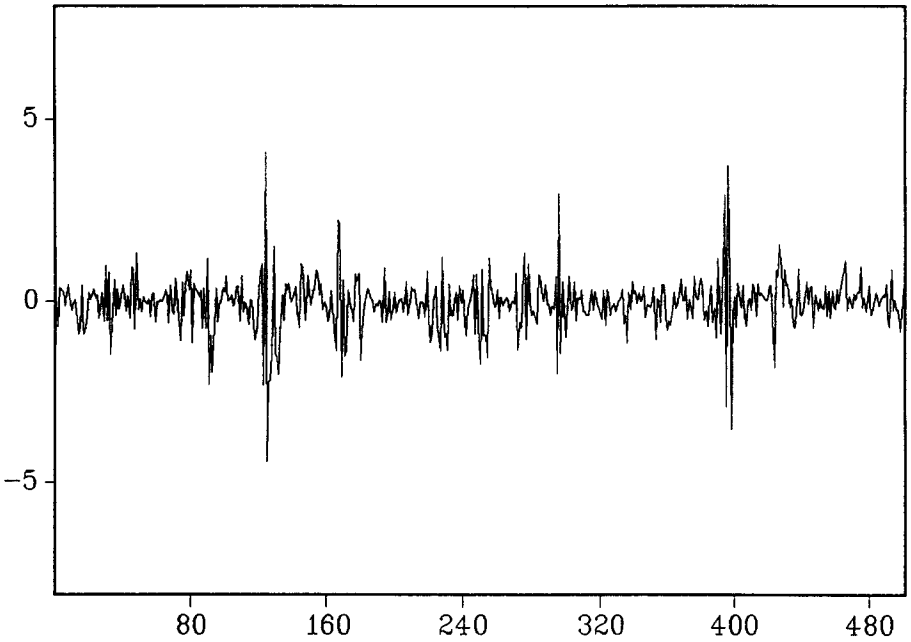

Figure 5. Simulated ARCH(1) data.

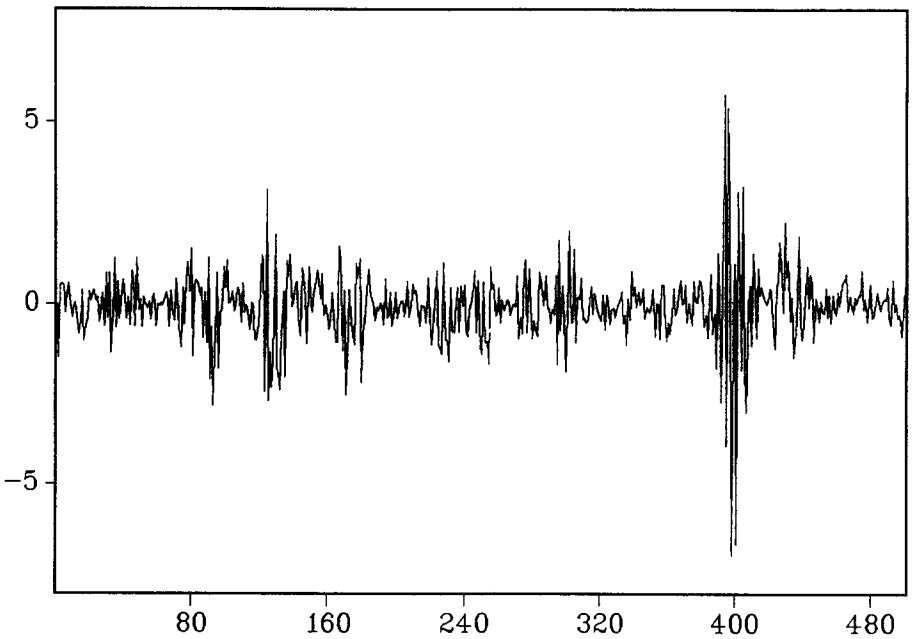

Figure 6. Simulated ARCH(4) data.

and

$$h_t = 0.1 + 0.36\varepsilon_{t-1}^2 + 0.27\varepsilon_{t-2}^2 + 0.18\varepsilon_{t-3}^2 + 0.09\varepsilon_{t-4}^2. \tag{2.6}$$

To make the scale of the data comparable in all three figures, the parameter values in (2.5) and (2.6) were chosen to make the unconditional variances of the ARCH processes equal to one. After section 2.3, it will be clear why the conditional variance functions (2.5) and (2.6) imply that the unconditional variances of the processes are one. We do not notice any clustering of the observations in Figure 4. Figures 5 and 6, however, have close resemblance to our earlier Figures 1, 2 and 3. In particular, the closeness of Figures 2 and 6 is quite striking. Comparing Figures 5 and 6, we also note that, as expected, the episodes of volatility are longer for ARCH(4).

2.2. *Generalized autoregressive conditional heteroskedasticity*

In the first empirical applications of ARCH to the relationship between the level and the volatility of inflation, Engle (1982, 1983) found that a large lag q was required in the conditional variance function. This would necessitate estimating a large number of parameters subject to inequality restrictions. To reduce the computational burden, Engle (1982, 1983) parameterized the conditional variance as

$$h_t = \alpha_0 + \alpha_1 \sum_{i=1}^{q} w_i \varepsilon_{t-i}^2,$$

where the weights

$$w_i = \frac{(q+1) - i}{\frac{1}{2}\, q(q+1)}$$

decline linearly and are constructed so that $\Sigma_{i=1}^{q} w_i = 1$. With this parameterization, a large lag can be specified and yet only two parameters are required to be estimated in the conditional variance function. Although linearly declining weights are plausible, the formulation does put undue restrictions on the dynamics of the ARCH process.

Bollerslev (1986) proposed an extension of the conditional variance function (2.3), which he termed generalized ARCH (GARCH), that has proven to be very useful in empirical work. The GARCH model was also independently proposed by Taylor (1986), who used a different acronym. They suggested that the conditional variance be specified as

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p}, \tag{2.7}$$

where the inequality restrictions

$$\begin{aligned} &\alpha_0 > 0 \\ &\alpha_i \geqslant 0 \quad \text{for } i = 1, \ldots, q \\ &\beta_i \geqslant 0 \quad \text{for } i = 1, \ldots, p \end{aligned} \tag{2.8}$$

are imposed to ensure that the conditional variance is strictly positive. A GARCH process with orders p and q is denoted as GARCH(p, q). The motivation of the GARCH process can be seen by expressing (2.7) as

$$h_t = \alpha_0 + \alpha(B)\varepsilon_t^2 + \beta(B)h_t,$$

where $\alpha(B) = \alpha_1 B + \cdots + \alpha_q B^q$ and $\beta(B) = \beta_1 B + \cdots + \beta_p B^p$ are polynomials in the backshift operator B. If the roots of $1 - \beta(Z)$ lie outside the unit circle, we can rewrite (2.7) as

$$\begin{aligned} h_t &= \frac{\alpha_0}{1-\beta(1)} + \frac{\alpha(B)}{1-\beta(B)}\varepsilon_t^2 \\ &= \alpha_0^* + \sum_{i=1}^{\infty} \delta_i \varepsilon_{t-i}^2 \end{aligned} \tag{2.9}$$

where $\alpha_0^* = \alpha_0/(1-\beta(1))$ and the coefficient δ_i is the coefficient of B^i in the expansion of $\alpha(B)[1-\beta(B)]^{-1}$. Hence, expression (2.9) reveals that a GARCH(p, q) process is an infinite order ARCH process with a rational lag structure imposed on the coefficients. The generalization of ARCH to GARCH is similar to the generalization of an MA process to an ARMA process. The intention is that GARCH can parsimoniously represent a high order ARCH process.

Although the restrictions (2.8) are sufficient to ensure that the conditional variance of a GARCH(p, q) process is strictly positive, Nelson and Cao (1992) demonstrated that weaker sufficient conditions can be found [see also Drost and Nijman (1992)]. They pointed out that from the *inverted* representation of h_t in (2.9),

$$\alpha_0^* > 0 \text{ and } \delta_i \geqslant 0, \quad i = 1, \ldots, \infty \tag{2.10}$$

are sufficient to ensure the conditional variance is strictly positive. Expressing α_0^* and the δ_i's in terms of the original parameters of the GARCH model, Nelson and Cao showed that (2.10) does not require that all the inequalities in (2.8) to hold. For example, in a GARCH(1, 2) process, $\alpha_0 > 0$, $\alpha_1 \geqslant 0$, $\beta_1 \geqslant 0$ and $\beta_1\alpha_1 + \alpha_2 \geqslant 0$ are sufficient to guarantee that $h_t > 0$. Therefore, in the GARCH(1, 2) model, α_2 may be negative. They presented general results for GARCH$(1, q)$ and GARCH$(2, q)$, but suggest a derivation for GARCH processes with $p \geqslant 3$ is difficult. Nelson and Cao cited several empirical studies, such as French, Schwert and Stambaugh (1987), Baillie and Bollerslev (1989), and Engle, Ito and Lin (1990), which reported negative coefficients and yet satisfy the conditions for a positive conditional variance based on (2.10). They concluded that the inequality restrictions (2.8) should not be imposed in estimation, as violation of these inequalities does not necessarily imply that the conditional variance function is misspecified.

2.3. *Unconditional moments of the ARCH*

Above, we described verbally the properties of ARCH and illustrated the visual appearance of ARCH with computer generated data. The unconditional moments of the ARCH process formally characterize these properties. Engle (1982) gave expressions for many of the moments, and stated necessary and sufficient conditions for the existence of the moments for the original linear ARCH process (2.3). Milhoj (1985) provided additional moments. Subsequently, Bollerslev (1986) extended these results to the GARCH process.

The derivation of the unconditional moments of the ARCH process is possible through extensive use of the following important probability result:

> *Law of Iterated Expectations*: Let Ω_1 and Ω_2 be two sets of random variables such that $\Omega_1 \subseteq \Omega_2$. Let y be a scalar random variable. Then $E(y \mid \Omega_1) = E[E(y \mid \Omega_2) \mid \Omega_1]$.

In the context of this paper, Ω_1 and Ω_2 are information sets available at different time periods. A special case of the law is frequently employed to find the moments of the ARCH process. If $\Omega_1 = \varnothing$ is the empty set, then $E(y) = E[E(y \mid \Omega_2)]$. This expression is useful because it relates an unconditional moment to a conditional moment. Since the ARCH model is specified in terms of its conditional moments, it provides a method for deriving unconditional moments.

Using the law of iterated expectations, we can easily derive the fundamental properties of an ARCH process. First, consider the unconditional mean of a GARCH(p, q) error ε_t with conditional variance (2.7). Applying the law of iterated expectations, $E(\varepsilon_t) = E[E(\varepsilon_t \mid \Psi_{t-1})]$. However, because the GARCH model specifies that $E(\varepsilon_t \mid \Psi_{t-1}) = 0$ for all realizations of Ψ_{t-1}, it immediately follows that $E(\varepsilon_t) = 0$. Thus, the GARCH process has mean zero.

Next, consider the unconditional variance of the GARCH(p, q) process. Although the variance of ε_t can be evaluated in general, for simple illustration, we consider the GARCH(1, 1) process. Using (2.7), with $p = q = 1$, and the law of iterated expectations

$$\begin{aligned} E(\varepsilon_t^2) &= E[E(\varepsilon_t^2 \mid \Psi_{t-1})] \\ &= E(h_t) \\ &= \alpha_0 + \alpha_1 E(\varepsilon_{t-1}^2) + \beta_1 E(h_{t-1}) \\ &= \alpha_0 + (\alpha_1 + \beta_1) E(\varepsilon_{t-1}^2), \end{aligned}$$

which is a linear difference equation for the sequence of variances. Assuming the process began infinitely far in the past with a finite initial variance, the sequence of variances converge to the constant

$$\sigma_\varepsilon^2 = E(\varepsilon_t^2) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

if $\alpha_1 + \beta_1 < 1$. For the general GARCH(p, q) process, Bollerslev (1986) gave the

necessary and sufficient condition

$$\alpha(1) + \beta(1) = \sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1 \tag{2.11}$$

for the existence of the variance. When this condition is satisfied, the variance is

$$\sigma_\varepsilon^2 = E(\varepsilon_t^2) = \frac{\alpha_0}{1 - \alpha(1) - \beta(1)}.$$

Although the variance of ε_t conditional on Ψ_{t-1} changes with the elements of the information set, unconditionally the ARCH process is *homoskedastic*. Considering Figures 5 and 6 again, the visual appearance of the generated data conveys the impression that the unconditional variance changes with time. This false perception results from the clustering of large deviations. A major contribution of the ARCH literature is the finding that apparent changes in the volatility of economic time series may be predictable and result from a specific type of nonlinear dependence rather than exogenous structural change in the variance.

The nature of the unconditional density of an ARCH process can be analyzed by the higher order moments. As ε_t is conditionally normal, for all odd integers m, $E(\varepsilon_t^m \mid \Psi_{t-1}) = 0$. The skewness coefficient is immediately seen to be zero. Since ε_t is continuous, this implies that the unconditional distribution is symmetric. Higher moments indicate further properties of the ARCH process. An expression for the fourth moment of a general GARCH(p, q) process is not available, but Engle (1982) gave it for the ARCH(1) process and Bollerslev (1986) generalized it to the GARCH(1,1) case. Engle's result for the ARCH(1) case requires that $3\alpha_1^2 < 1$ for the fourth moment to exist. Simple algebra then reveals that the kurtosis is

$$\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} = 3\left(\frac{1 - \alpha_1^2}{1 - 3\alpha_1^2}\right),$$

which is clearly greater than 3, the kurtosis coefficient of the normal distribution. Therefore, the ARCH(1) process has tails heavier than the normal distribution. This property makes the ARCH process attractive because the distributions of asset returns frequently display tails heavier than the normal distribution. Although no known closed form for the unconditional density function of an ARCH process exists, Nelson (1990b) demonstrated that under suitable conditions, as the time interval goes to zero, a GARCH(1,1) process approaches a continuous time process whose stationary unconditional distribution is a Student's t. Nelson's result indicates why heavy tailed distributions are so prevalent with high frequency financial data.

That the parameterization of the ARCH process does not a priori impose the existence of unconditional moments is an important characteristic of the model. It has long been suggested, at least as early as Mandelbrot (1963b), that the distribution of asset returns are such that the variance may not exist. In empirical applications of GARCH, estimated parameters frequently do not satisfy (2.11).

The fact that the ARCH model admits an infinite variance is desirable because such behaviour may be a characteristic of the data generating process that should be reflected in the estimated model. Also, fortunately, as will be noted in Section 8, even for GARCH models with infinite variances, standard results on consistency and asymptotic normality might still be valid.

Above we considered the univariate distribution of a single ε_t. The moments of the joint distribution of the ε_t's also reveal important properties of the ARCH process. For $k \geqslant 1$, the autocovariances of the GARCH(p, q) process are

$$\begin{aligned} E(\varepsilon_t \varepsilon_{t-k}) &= E[E(\varepsilon_t \varepsilon_{t-k} \mid \Psi_{t-1})] \\ &= E[\varepsilon_{t-k} E(\varepsilon_t \mid \Psi_{t-1})] \\ &= 0. \end{aligned}$$

Since the GARCH process is serially uncorrelated, with constant mean zero, the process is weakly stationary if the variance exists, that is if (2.11) holds. A remarkable property of a GARCH process, first demonstrated by Nelson (1990a) for GARCH(1,1), is that it may be strongly stationary without being weakly stationary. Bougerol and Picard (1992) extended Nelson's result to the GARCH(p, q) process and stated necessary and sufficient condition for strong stationarity. These conditions are very technical and will not be described here. That the GARCH process may be strongly stationary without being weakly stationary stems from the fact that weak stationarity requires that the mean, variance and autocovariances be *finite* and time invariant. Strong stationarity requires that the distribution function of any finite set of ε_t's is invariant under time translations. Finite moments are not required for strong stationarity. The results of Nelson (1990a) and Bougerol and Picard (1992) show that the unconditional variance may be infinite and yet the GARCH process may still be strongly stationary.

The lack of serial correlation is an important characteristic of the ARCH process which makes it suitable for modeling financial time series. The efficient market hypothesis asserts that past rates of return can not be used to improve the prediction of future rates of return. In (2.1), suppose the y_t is the rate of return on an asset and that $\xi = 0$ so that there is no regression component in the model. Then y_t is identical to ε_t and becomes a pure GARCH process. The optimal prediction of the return y_t is the expectation of the return conditional on any available information. But because the GARCH model specifies $E(y_t \mid \Psi_{t-1}) = E(y_t) = 0$, the past observations on y_t contained in Ψ_{t-1} do not alter the optimal prediction of the rate of return. Therefore, the presence of ARCH does not represent a violation of market efficiency.

Of course, the lack of serial correlation does not imply that the ε_t are independent. Above, we suggested that the qualitative appearance of data generated from an ARCH process arises from the particular type of dependence. Bollerslev (1986) gave a representation for the GARCH(p, q) process which reveals the nature of the dependence. Letting $\nu_t = \varepsilon_t^2 - h_t$, the squared error can

be written as

$$\begin{aligned}\varepsilon_t^2 &= h_t + \nu_t \\ &= \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i)\varepsilon_{t-i}^2 - \sum_{i=1}^{p} \beta_i(\varepsilon_{t-i}^2 - h_{t-i}) + \nu_t \\ &= \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i)\varepsilon_{t-i}^2 - \sum_{i=1}^{p} \beta_i \nu_{t-i} + \nu_t \qquad (2.12)\end{aligned}$$

where $m = \max(p, q)$, $\alpha_i = 0$ and $i > q$ and $\beta_i = 0$ for $i > p$. Because $E(\nu_t \mid \Psi_{t-1}) = 0$, the law of iterated expectations reveals that ν_t has mean zero and is serially uncorrelated. Therefore, from (2.12) we see that ε_t^2 has an ARMA(m,p) representation. The autocorrelation and partial autocorrelation functions of the squared process ε_t^2 will have the familiar patterns of those from an ARMA process. Bollerslev (1988) has suggested that these autocorrelation functions of ε_t^2 may be used to identify the orders p and q of the GARCH process. In practice, the identification of the order of a GARCH(p, q) has not posed much of a problem, at least in comparison with the earlier modeling experience with ARMA(p, q) processes. In applied work, it has been frequently demonstrated that the GARCH(1, 1) process is able to represent the majority of financial time series. A data set which requires a model of order greater than GARCH(1, 2) or GARCH(2, 1) is very rare.

2.4. *Illustrative example with the weekly dollar/pound exchange rate*

The ARCH model has been widely applied to the study of the dynamics of the rate of return on holding foreign currencies [see Bollerslev, Chou and Kroner (1992), pp. 37–46, for a survey of applications]. In this section, we illustrate the properties of conditionally heteroskedastic data by estimating ARCH and GARCH models for the weekly rate of return in the US/British currency exchange market. The data are the weekly spot exchange rate from January 1973 to June of 1985. There are 651 observations. Let s_t denote the spot price of the British pound in terms of the U.S dollar. We then analyze the continuously compounded percentage rate of return, $r_t = 100 \cdot \log(s_t/s_{t-1})$, from holding the British pound one week. This is the data plotted in Figure 1.

We begin by identifying and estimating an AR process for the mean of r_t. The autocorrelation and partial autocorrelation functions of r_t suggest the data can be represented by an AR(3) process. The estimated model is given by

$$\underset{(0.04)}{r_t = -0.07} + \underset{(0.04)}{0.27 r_{t-1}} - \underset{(0.04)}{0.08 r_{t-2}} + \underset{(0.04)}{0.10 r_{t-3}} \qquad l(\hat{\theta}) = -971.70,$$

where the standard errors are shown in parentheses and $l(\hat{\theta})$ is the value of the maximized log likelihood function assuming the data are normally distributed. Box-Pierce statistics computed from the residuals indicate that the AR(3) process adequately accounts for the serial correlation in the data. The higher order

moments of the residuals, however, reveal that nonlinearity is present in the data and that the unconditional distribution is nonnormal. In Table 1, we present the skewness and kurtosis coefficients of the residuals, and the autocorrelations of the squared residuals. If the errors of the AR process are independent, the autocorrelations of the squared residuals should be approximately zero. From Table 1, the autocorrelations at lags 1, 2, 3, 4 and 7 exceed twice their asymptotic standard errors, suggesting the presence of nonlinear dependence in the data. The skewness coefficient conveys some evidence of asymmetry in the unconditional distribution. The kurtosis coefficient is significantly greater than 3, which indicates that the unconditional distribution of the data has much heavier tails than a normal distribution.

As emphasized in section 2.3, nonlinear dependence and a heavy-tailed unconditional distribution are characteristic of conditionally heteroskedastic data. We maintain the AR(3) specification for the conditional mean of r_t, but now specify the error as an ARCH(q) process. The autocorrelations of the squares of the AR residuals suggest dependence through order 7. Therefore, we initially estimated an ARCH(7) model, but found α_7 to be insignificant. We respecify the errors as ARCH(6) and estimate the model by the maximum likelihood method to obtain

$$\underset{(0.03)}{r_t = -0.06} + \underset{(0.05)}{0.27r_{t-1}} + \underset{(0.05)}{0.03r_{t-2}} + \underset{(0.04)}{0.07r_{t-3}} \qquad l(\hat{\theta}) = -919.72$$

$$\underset{(0.06)}{h_t = 0.42} + \underset{(0.06)}{0.23\varepsilon_{t-1}^2} + \underset{(0.06)}{0.21\varepsilon_{t-2}^2} + \underset{(0.04)}{0.05\varepsilon_{t-3}^2}$$

$$+ \underset{(0.04)}{0.05\varepsilon_{t-4}^2} + \underset{(0.04)}{0.07\varepsilon_{t-5}^2} + \underset{(0.05)}{0.12\varepsilon_{t-6}^2}.$$

The ARCH parameters α_1, α_2 and α_6 are highly significant. The ARCH model also produces a significant increase in the value of the log likelihood. A likelihood ratio test easily rejects the null of an AR(3) process with independent Gaussian errors against the alternative of AR(3) process with conditionally normal ARCH(6) errors. Notice that the coefficient of r_{t-2} loses its significance. Frequently, after ARCH is accounted for, the initial specification of the mean must be reevaluated.

The ARCH(6) model can apparently explain the nonlinear dependence in the residuals. In Table 1, we present the autocorrelations of the squared standardized residuals $\hat{\eta}_t^2 = \hat{\varepsilon}_t^2/\hat{h}_t$. None of the first eight autocorrelations are significant at any reasonable significance level. The skewness coefficient of the standardized residuals is different in sign from the AR residuals and larger in magnitude, but still not excessively big. The sample kurtosis coefficient of the standardized residuals is smaller than the coefficient for the AR residuals, but is still significantly greater than 3. This suggests that the unconditional distribution of the conditionally normal ARCH process is not sufficiently heavy tailed to account for the excess kurtosis in the data. The rejection of the conditional normality assumption is frequently encountered in applications of the ARCH

Table 1. Summary statistics for the standardized residuals from AR, ARCH and GARCH models for the rate of return on the weekly US/British exchange rate

	Autocorrelations of squared residuals									
	1	2	3	4	5	6	7	8	skewness	kurtosis
AR(3)	0.16	0.14	0.22	0.15	0.06	0.05	0.11	0.00	0.22	6.78
AR(3) + ARCH(6)	0.01	−0.02	−0.01	0.01	−0.04	−0.02	−0.02	0.02	−0.45	6.40
AR(3) + GARCH(1, 1)	0.02	0.00	−0.02	0.00	−0.05	0.01	−0.03	−0.02	−0.41	6.21

Note:
(1) The asymptotic standard error of the autocorrelations of the squared standardized residuals is $1/\sqrt{T} = 0.04$.
(2) The asymptotic standard errors of the skewness and kurtosis coefficients are respectively 0.096 and 0.192.

model. As will be discussed in section 4.2, there are ways to take account of this excess kurtosis.

In section 2.2, we demonstrated how the GARCH model can provide a parsimonious parameterization of a high order ARCH process. To illustrate this, we estimate an AR(3) model for r_t with the conditional variance of the errors specified as GARCH(1, 1). Maintaining the conditional normality assumption, the estimated model is

$$\begin{aligned} r_t &= \underset{(0.04)}{-0.05} + \underset{(0.05)}{0.27}r_{t-1} - \underset{(0.05)}{0.003}r_{t-2} + \underset{(0.04)}{0.08}r_{t-3} \qquad l(\hat{\theta}) = -920.02 \\ h_t &= \underset{(0.03)}{0.09} + \underset{(0.04)}{0.17}\varepsilon_{t-1}^2 + \underset{(0.05)}{0.77}h_{t-1}. \end{aligned}$$

The estimates of the AR parameters are similar to the estimates for ARCH(6) errors, with only the coefficient of r_{t-2} changing sign and becoming even less significant. The autocorrelations of the squares of the GARCH(1, 1) standardized residuals, shown in Table 1, are insignificant and similar in magnitude to those for the ARCH(6) standardized residuals. This indicates that the GARCH(1, 1), which requires estimating only three conditional variance parameters, can account for the nonlinear dependence as well as the ARCH(6) model, which requires estimating seven conditional variance parameters. The

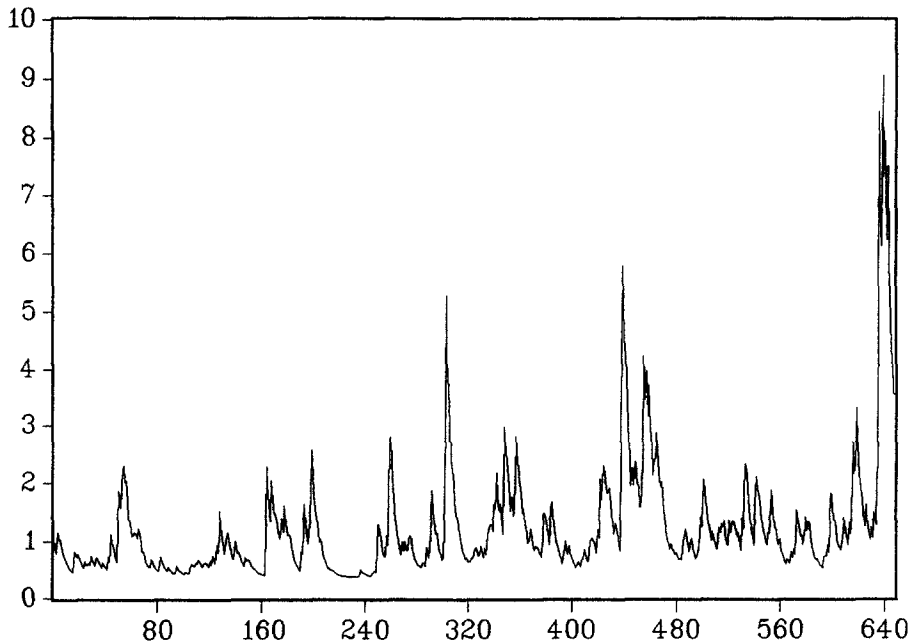

Figure 7. Conditional variance of the dollar/pound weekly return.

skewness and kurtosis coefficients of the standardized GARCH(1, 1) residuals, also given in Table 1, are almost identical to the coefficients for the standardized ARCH(6) residuals. The value of the maximized ARCH(6) log likelihood is marginally greater than the value of the GARCH(1, 1) log likelihood. But any model selection criteria, such as AIC or BIC, which penalizes a model for additional parameters, would select the GARCH(1, 1) specification over the ARCH(6) specification. Finally, in Figure 7 we present a plot of the estimates of the conditional variances, h_t, from the GARCH(1, 1) model. The conditional variances show considerable variation over time. Comparing the plot of the weekly returns in Figure 1 with the plot of the conditional variances in Figure 7, it is clear that a clustering of large deviations, of either sign, in the returns is associated with a rise in the conditional variance.

2.5. *Temporal aggregation of ARCH processes*

One of the important issues in time series modeling is temporal aggregation. It is well known that a high frequency (e.g., fitted to daily data) ARMA process aggregates to a low frequency (fitted to say, weekly data) ARMA process. A natural question is whether ARCH models also possess this property. Drost and Nijman (1992) considered this issue in detail and we follow their analysis. Let us consider the ARCH model (2.2) and (2.3) with $q = 1$, i.e.,

$$\varepsilon_t \mid \Psi_{t-1} \sim N(0, h_t) \tag{2.13}$$

where

$$h_t = E(\varepsilon_t^2 \mid \Psi_{t-1}) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad t = 1, 2, \ldots, T. \tag{2.14}$$

Suppose we want to find the corresponding model for ε_t, when $t = 2, 4, \ldots, T$. The information set will consist of only $\{y_{t-2}, x_{t-2}, y_{t-4}, x_{t-4}, \ldots\}$ and we will denote it by $\Psi_{t-(2)}$. Drost and Nijman showed that

$$\begin{aligned} E(\varepsilon_t \mid \Psi_{t-(2)}) &= 0 \\ E(\varepsilon_t^2 \mid \Psi_{t-(2)}) &= \alpha_0(1 + \alpha_1) + \alpha_1^2 \varepsilon_{t-2}^2, \quad t = 2, 4, \ldots, T \\ &= h_{t(2)}, \text{ say.} \end{aligned}$$

In general, if we consider $t = m, 2m, \ldots, T$, then

$$\begin{aligned} E(\varepsilon_t \mid \Psi_{t-(m)}) &= 0 \\ E(\varepsilon_t^2 \mid \Psi_{t-(m)}) &= \alpha_0 \frac{1 - \alpha_1^m}{1 - \alpha_1} + \alpha_1^m \varepsilon_{t-m}^2. \end{aligned}$$

Therefore, in terms of the first two moments, an ARCH process is closed under temporal aggregation and we have an algebraic relationship between the parameters corresponding to high and low frequency data. It is interesting to note that as $m \to \infty$, $E(\varepsilon_t^2 \mid \Psi_{t-(m)}) \to \alpha_0/(1 - \alpha_1)$, so that in the limit, the aggregate process behaves like a conditional homoskedastic model as pointed out by Diebold (1988). If we consider the reverse operation of going from a low

frequency model to a higher one, in the limit the process will have an integrated ARCH structure as noted by Nelson (1990b).

Now let us consider the distributional part of the specification (2.13), which can be stated as

$$\varepsilon_t/\sqrt{h_t} \mid \Psi_{t-1} \sim \text{IID } N(0,1) \qquad t = 1, 2, \ldots, T. \tag{2.15}$$

We need to check the conditional distribution of $\varepsilon_t^* = \varepsilon_t/\sqrt{h_{t(2)}}$, $t = 2, 4, \ldots, T$. Drost and Nijman (1992) showed that

$$E(\varepsilon_t^{*4} \mid \Psi_{t-(2)}) = 3 + 6\left[\frac{\alpha_0}{h_{t(2)}} - 1\right]^2.$$

Therefore, the conditional moments of ε_t^* depends on the information set and hence the conditional distribution of ε_t^* does not have the IID structure (2.15). Also the distribution is no longer normal. Therefore, from a distributional point of view, an ARCH process is not closed under aggregation.

For practical purposes, if we specify an ARCH model only in terms of moments, it is possible to estimate the low frequency parameters from the estimation of a high frequency model and vice versa. Drost and Nijman (1992) demonstrated this using the empirical results of Baillie and Bollerslev (1989), who fitted a GARCH(1, 1) model to several exchange rates. For the Swiss franc, the estimates of α_1 and β_1 from the daily data were 0.073 and 0.907. Using the relationship between the parameters of high and low frequency data, Drost and Nijman showed that the implied weekly estimates are 0.112 and 0.792. Baillie and Bollerslev's estimates using the actual weekly data were 0.121 and 0.781. Except for the Japanese yen, Drost and Nijman found that direct estimates were very close to the implied weekly estimates.

3. Interpretations of ARCH

Apart from their simplicity, the main reason for the success of ARCH models is that they take account of the many observed features of the data, such as thick tails of the distribution, clustering of large and small observations, nonlinearity and changes in our ability to forecast future values. Therefore, it is not surprising that these models can be interpreted in a number of ways, and we discuss some of these interpretations in this section.

3.1. *Random coefficient interpretation*

In the last section, we noted that ARCH takes account of the clustering of large and small errors and fatness of the tail part of the distribution (excess kurtosis) as observed in many financial data series. One of the major considerations for introducing ARCH by Engle (1982, p. 989) was that econometricians' ability to predict the future varies from one period to another. Predictions are usually done by using a conditional mean model. Uncertainty about the conditional mean can be expressed by a random coefficient formulation. Consider a random

coefficient AR(1) process

$$y_t = \phi_t y_{t-1} + \varepsilon_t$$

where $\phi_t \sim (\phi, \alpha_1)$, and $\varepsilon_t \sim (0, \alpha_0)$ are independent. Then $E(y_t \mid \Psi_{t-1}) = \phi y_{t-1}$ as with the fixed AR(1) process; however, now $\mathrm{Var}(y_t \mid \Psi_{t-1}) = \alpha_0 + \alpha_1 y_{t-1}^2$, which has the same form as (2.3). To obtain a general ARCH(q) model in our regression context from a random coefficient framework, we need to start with the following set up:

$$y_t = x_t'\xi + \varepsilon_t \tag{3.1}$$

$$\varepsilon_t = \sum_{i=1}^{q} \phi_{it}\varepsilon_{t-i} + u_t$$

$$= \sum_{i=1}^{q} (\phi_i + \eta_{it})\varepsilon_{t-i} + u_t \tag{3.2}$$

where $\eta_t = (\eta_{1t}, \ldots, \eta_{qt})' \sim (0, A_{q\times q})$ and $u_t \sim (0, \sigma_u^2)$ are independent. It immediately follows that

$$E(\varepsilon_t \mid \Psi_{t-1}) = \phi'\underline{\varepsilon}_{t-1},$$

where $\phi = (\phi_1, \ldots, \phi_q)'$ and $\underline{\varepsilon}_{t-1} = (\varepsilon_{t-1}, \ldots, \varepsilon_{t-q})'$, and that

$$\mathrm{Var}(\varepsilon_t \mid \Psi_{t-1}) = \underline{\varepsilon}_{t-1}' A \underline{\varepsilon}_{t-1} + \sigma_u^2. \tag{3.3}$$

If $A = ((\alpha_{ij}))$ is a diagonal matrix with $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_q)$ and $\sigma_u^2 = \alpha_0$, then

$$\mathrm{Var}(\varepsilon_t \mid \Psi_{t-1}) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$$

as we have in (2.3). A non-diagonal 'A' specifies an ARCH process with additional cross-product terms between the past errors. The intuition behind the inclusion of the cross-product terms is that they take account of the effect of the interaction between the lagged residuals on the conditional variance. White's (1980) test for heteroskedasticity has a similar feature which includes the cross-products of the regressors as the test variables while the operational form of the Breusch and Pagan (1979) test does not. The model (3.3) was discussed in detail by Bera, Higgins and Lee (1992), which they called the augmented ARCH (AARCH) model [see also Tsay (1987)]. If we add linear terms of $\underline{\varepsilon}_{t-1}$ in (3.3), we obtain the quadratic ARCH (QARCH) model of Sentana (1991). Bera, Higgins and Lee (1990) extended the framework (3.1) and (3.2) to give the GARCH(p, q) model a random coefficient interpretation. It is immediately seen that unlike ARCH, AARCH is not symmetric in the sense that the conditional variance depends on the sign of the individual lagged ε_t's.

In their empirical analysis of exchange rate data, Cheung and Pauly (1990) found that many of the off-diagonal elements of 'A' were significantly different from zero and concluded that a random coefficient formulation provided a richer formulation of time varying volatility than did the standard ARCH characterization. Bera, Higgins and Lee (1990) also noted similar results when

they reconsidered Engle's (1983) model for measuring variability of U.S. inflation. They found that the estimate of the coefficient of the augmented ARCH term $\varepsilon_{t-4}\varepsilon_{t-7}$ was -0.195 with a t-statistic of 4.89. Their resulting specification of an AR-AARCH model passed the specification tests and diagnostic checks they performed, while Engle's original ARCH model had some unexplained serial correlation and conditional heteroskedasticity. An empirical application in Sentana (1991) with a century of daily U.S. stock returns provided support for his QARCH model. Coefficients of all of the cross-product terms were highly significant.

Bera and Lee (1993) established the connection between random coefficients and ARCH in a somewhat indirect way. They applied White's (1982) information matrix (IM) test to a linear regression model with autocorrelated errors. The IM test had six distinct components and a special case of one component, which corresponded to the autocorrelation parameter ϕ, was found to be identical to Engle's (1982) Lagrange multiplier (LM) test for ARCH. Given Chesher's (1984) interpretation of the IM test as a test for parameter variation, it can be said that as far as the test is concerned, the presence of ARCH is 'equivalent' to random variation in the autocorrelation coefficient. In our above analysis, we noted that both ARCH and the random coefficient model lead to the same first two conditional moments. Under the additional assumption of conditional normality, all the moments, and hence the two processes themselves, will be identical.

One byproduct of the random coefficient representation of the ARCH model is that standard results from the time series literature can be used to derive the necessary and sufficient conditions for stationarity. Andel (1976), Nicholls and Quinn (1982) and Ray (1983) stated simple conditions for second order stationarity of the AR process with random coefficients. In section 2.3, we noted that the stationarity condition for an ARCH(q) process in the absence of autocorrelation is $\Sigma_{i=1}^{q}\alpha_i < 1$. As demonstrated in Bera, Higgins and Lee (1990), presence of autocorrelation leads to a different stationarity condition. For example, the stationarity condition for an ARCH(q) process in the presence of first order serial correlation is

$$\frac{1}{1-\phi_1^2}\sum_{i=1}^{q}\alpha_i < 1.$$

In the absence of autocorrelation, $\Sigma_{i=1}^{q}\alpha_i < 1$ is sufficient for weak stationarity. This clearly demonstrates that the presence of autocorrelation can make a stationary ARCH process non-stationary.

3.2. *A conditional mixture model interpretation*

Following the work of Clark (1973) and Tauchen and Pitts (1983), Gallant, Hsieh and Tauchen (1991) provided an interesting rationale for the presence of conditional heteroskedasticity and heterogeneity in the higher order moments of

asset prices. Let us write the observed price change y_t as

$$y_t = \mu_t + \sum_{i=1}^{I_t} \zeta_i. \tag{3.4}$$

where $\zeta_i \sim \text{IID } N(0, \tau^2)$. Here μ_t can be viewed as the forecastable component, the ζ_i's are the incremental changes and I_t is the number of times new information comes to the market in period t. I_t is a serially dependent unobservable random variable and is independent of $\{\zeta_i\}$. Because of the randomness of I_t, y_t is not normally distributed; it is in fact a mixture of normal distributions. Here we can view y_t as a subordinated stochastic process, where $y_t - \mu_t$ is subordinate to ζ_i, and I_t is the directing process. Equation (3.4) can be written as

$$y_t = \mu_t + \tau I_t^{1/2} \nu_t \tag{3.4a}$$

with $\nu_t \sim N(0, 1)$. Then, conditional on the information set Ψ_{t-1} and I_t, we have the conditional heteroskedastic normal distribution

$$y_t \mid \Psi_{t-1}, I_t \sim N(\mu_t, \tau^2 I_t). \tag{3.5}$$

Since I_t is not observable, in practice we can work only with the conditional distribution $y_t \mid \Psi_{t-1}$. From the general result that if a random variable is conditionally (on I_t) normal, unconditionally it must be nonnormal, a realistic distribution for $y_t \mid \Psi_{t-1}$ would be conditionally heteroskedastic and nonnormal.

Framework (3.5) is very general, and a variety of interesting cases can be derived from this. When I_t is a constant c, we have

$$y_t \mid \Psi_{t-1} \sim N(\mu_t, c\tau^2),$$

which is our standard homoskedastic model. If our information set Ψ_{t-1} also includes I_t, then

$$y_t \mid \Psi_{t-1} \sim N(\mu_t, \tau^2 I_t).$$

This is a conditional heteroskedastic-normal model. However, the assumption about the knowledge of I_t is not realistic. For the general case (3.5) the first four moments are

$$\begin{aligned}
E[(y_t - \mu_t) \mid \Psi_{t-1}] &= 0 \\
E[(y_t - \mu_t)^2 \mid \Psi_{t-1}] &= \tau^2 E[I_t \mid \Psi_{t-1}] \\
E[(y_t - \mu_t)^3 \mid \Psi_{t-1}] &= 0 \\
E[(y_t - \mu_t)^4 \mid \Psi_{t-1}] &= \tau^4 3 E[I_t^2 \mid \Psi_{t-1}].
\end{aligned}$$

Hence, the conditional kurtosis

$$\frac{E[(y_t - \mu_t)^4 \mid \Psi_{t-1}]}{E[(y_t - \mu_t)^2 \mid \Psi_{t-1}]^2} = \frac{3E[I_t^2 \mid \Psi_{t-1}]}{E[I_t \mid \Psi_{t-1}]^2} \tag{3.6}$$

exceeds 3. Therefore, it is not surprising that in many empirical studies the normal-ARCH model could not capture most of the excess kurtosis in the data,

while a conditional t or some nonnormal ARCH models worked somewhat better [see, for example, Engle and Bollerslev (1986), Baillie and Bollerslev (1989), Bollerslev (1987), Hsieh (1989), Gallant, Hsieh and Tauchen (1991), Gallant and Tauchen (1989), Lee and Tse (1991)]. Conditional t or other nonnormal distributions do not of course solve all the problems, since the quantity in (3.6) is not necessarily time invariant. The conditional t distribution, for example, although it allows kurtosis to exceed 3, assumes constant conditional kurtosis. Note that the kurtosis in (3.6) will be time invariant if I_t and Ψ_{t-1} are independent. To take account of the time varying higher moments, Hansen (1992) generalized the conditional t model by expressing the corresponding shape parameter (the degrees of freedom) as a function of the information set. We will discuss this model in section 4.2.

Bera and Zuo (1991) suggested a specification test for ARCH models, which examines the constancy of the kurtosis of the standardized residuals of the estimated ARCH model. They call it a test for heterokurtosis. The test is derived using the information matrix test principle and hence is a test for heterogeneity of the ARCH parameters. As we discussed earlier, conditional heteroskedasticity can be viewed as a randomness of the AR parameters. Conditional heterokurtosis is related to the heterogeneity of the ARCH parameters. Mizrach (1990) used a generalization of the ARCH model which allowed for time varying coefficients in the conditional variance equation, and found the model to perform better than the standard GARCH model in an exchange rate application.

At this point a question could be raised: why in many empirical applications do ARCH models work remarkably well? To explain this, we again follow Gallant, Hsieh and Tauchen (1991). As noted before, the conditional variance is

$$E[(y_t - \mu_t)^2 \mid \Psi_{t-1}] = \tau^2 E[I_t \mid \Psi_{t-1}].$$

Denoting $y_t - \mu_t = \tau I_t^{1/2} \nu_t$ as the error ε_t, we have

$$\begin{aligned} \mathrm{Cov}(\varepsilon_t^2, \varepsilon_{t-j}^2) &= \tau^4 \ \mathrm{Cov}(I_t \nu_t^2, I_{t-j} \nu_{t-j}^2) \\ &= \tau^4 \ \mathrm{Cov}(I_t, I_{t-j}). \end{aligned}$$

If the I_t's are serially dependent, which seems plausible a priori, that will introduce correlation in the squared errors. The ARCH methodology tries to capture this correlation.

Using U.S. daily stock return data, Lamoureux and Lastrapes (1990a) provided empirical evidence in support of the hypothesis that ARCH is a manifestation of the time dependence in the rate of information arrival to the market. They assumed that I_t in (3.4) is serially correlated and expressed it as

$$I_t = \gamma_0 + \gamma(B) I_{t-1} + u_t, \tag{3.7}$$

where γ_0 is a constant, $\gamma(B)$ is a lag polynomial and u_t is white noise. Defining $\Omega_t = E[(y_t - \mu_t)^2 \mid I_t] = \tau^2 I_t$ and using (3.7) we have

$$\Omega_t = \tau^2 \gamma_0 + \gamma(B) \Omega_{t-1} + \tau^2 u_t,$$

which has a similar structure to that of a GARCH model. Since I_t is not observable, Lamoureux and Lastrapes used daily trading volume, V_t, as a proxy for the daily information that flows into the market. When V_t was included as an extra variable in the GARCH(1, 1) model (2.7), its coefficient was highly significant for all of the 20 stocks they considered. Also, inclusion of V_t in h_t made the ARCH effects (coefficients α_1 and β_1) become negligible for most of the stocks. To summarize, this empirical work supports the view that ARCH in daily stock returns is an outcome of the time dependence in the news that flows into the market.

To evaluate the role of news in the determination of volatility in the foreign exchange markets, Engle, Ito and Lin (1990) provided a test of two hypotheses — heat waves and meteor showers. The heat wave hypothesis states that the major sources of disturbances come from *within* a market, while the meteor shower hypothesis states that disturbances come from spillovers *between* markets. They used the intra-daily yen/dollar exchange rate in the Tokyo, European, New York and Pacific markets. To test the two hypotheses, they included the squared innovations from the other markets in the specification of each h_t. Coefficients of all of these variables were found to be highly significant, thus lending support to the meteor shower hypothesis. In fact, they found that the foreign news was more important than the past domestic news. In particular, Japanese news had the greatest impact on the volatility of all markets except the Tokyo market.

3.3. *Nonlinear model interpretation*

It is clear that one of the essential features of the ARCH model is $\mathrm{Cov}(\varepsilon_t^2, \varepsilon_{t-j}^2) \neq 0$, although $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$, for $j \neq 0$. In other words, ARCH postulates a nonlinear relationship between ε_t and its past values. There are many nonlinear time series models such as the bilinear, threshold autoregressive, exponential autoregressive and nonlinear moving average models that can also exhibit this property [see Tong (1990)]. For simplicity, we concentrate on the bilinear model, and its relation to the ARCH model. A time series $\{\varepsilon_t\}$ is said to follow a bilinear model if it satisfies [see Granger and Andersen (1978) and Tong (1990)]

$$\varepsilon_t = \sum_{i=1}^{p} \phi_i \varepsilon_{t-i} + \sum_{j=1}^{r} \sum_{k=1}^{s} b_{jk} \varepsilon_{t-j} u_{t-k} + u_t, \tag{3.8}$$

where u_t is a sequence of IID $(0, \sigma_u^2)$ variables. The first two conditional moments for this process are

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

$$\mathrm{Var}(\varepsilon_t \mid \Psi_{t-1}) = \sigma_u^2.$$

These conditional moments contrast with those of an ARCH process in which

the conditional mean is, in general, a constant, but the conditional variance is time varying. Their unconditional moments, however, might be similar. For example, the bilinear model

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

has $E(\varepsilon_t) = 0$ and $\mathrm{Cov}(\varepsilon_t^2, \varepsilon_{t-2}^2) = b_{21}^2\sigma_u^2$. As this process is autocorrelated in squares, it will exhibit temporal clustering of large and small deviations like an ARCH process. In fact, a bilinear model is quite similar to an ARCH model, in that it can also be represented as a varying coefficient model. Equation (3.8) can be written as

$$\begin{aligned}\varepsilon_t &= \sum_{j=1}^{m} [\phi_j + A_j(t)]\,\varepsilon_{t-j} + u_t \\ &= \sum_{j=1}^{m} \phi_{jt}\varepsilon_{t-j} + u_t, \text{ say,}\end{aligned} \tag{3.9}$$

where $m = \max(p, r)$, and $A_j(t) = \Sigma_{k=1}^{s} b_{jk}u_{t-k}$ with $\phi_i = 0$, $i \geqslant p+1$, $b_{jk} = 0$, $j \geqslant r+1$ [see Tong (1990, p. 114)]. The basic difference between (3.2) and (3.9) is that in the former, the coefficients are purely random, whereas in (3.9), the varying coefficient part $A_j(t)$ has a structure which is a linear function of the lagged innovations u_t.

There is yet another way of looking at the similarities and differences between ARCH and bilinear models. Although both models take account of nonlinear dependence, ARCH represents the dependence in a multiplicative fashion,

$$\begin{aligned}\varepsilon_t &= u_t \cdot f_1(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots; u_{t-1}, u_{t-2}, \ldots) \\ &= u_t \cdot f_{1t}, \text{ say,}\end{aligned} \tag{3.10}$$

while a bilinear model postulates an additive structure

$$\begin{aligned}\varepsilon_t &= f_2(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots; u_{t-1}, u_{t-2}, \ldots) + u_t \\ &= f_{2t} + u_t, \text{ say,}\end{aligned} \tag{3.11}$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are some well defined nonlinear functions. Hsieh (1989) exploited these differences to discriminate between the two types of nonlinearities. Higgins and Bera (1991) suggested a Cox nonnested procedure to test these two models against each other.

From a practical point of view, these models have different implications. Using a bilinear model we can improve the point forecast over standard ARMA modeling, but cannot assess the accuracy of the forecast interval. On the other hand, the ARCH specification makes it possible to forecast the conditional variance without any additional gain in point forecastability. It is quite possible that the data may be represented by a joint ARCH-bilinear model such as the one suggested by Weiss (1986c). Higgins and Bera (1989, 1991) developed simple procedures for detecting the joint presence of ARCH and bilinearity.

The empirical results on this topic are somewhat mixed. Hsieh (1989) finds that the ARCH model is able to account for the nonlinearities in the daily German

mark, Canadian dollar and Swiss franc, but not in the British pound nor the Japanese yen. The ARCH standardized residuals exhibited substantial nonlinearity for the latter two currencies and, for the British pound, more nonnormality (excess kurtosis) than the raw data. Diebold and Nason (1990) addressed the issue of whether conditional heteroskedasticity actually exists in exchange rate data or whether it is just a reflection of some misspecification in the conditional mean of the model. They tackled the problem by estimating the conditional mean through a nonparametric regression and testing the residuals for the presence of ARCH. ARCH was found in the nonparametric residuals, implying that conditional heteroskedasticity was not due to missepecification of the mean. Higgins and Bera (1991) applied the Cox test to six weekly exchange rates. For the Canadian dollar, the GARCH model was not rejected and for the British pound, the bilinear was not rejected. For the other currencies, the French franc, the German mark, the Japanese yen and the Swiss franc, both of the models were found to be inadequate.

Lastly, we should mention an inherent problem in using a nonlinear conditional mean specification to model financial data. For a nonlinear conditional mean model to explain the sort of volatility observed in practice, the variation in the conditional first moment would have to be enormous, implying huge unexploited profit opportunities for the traders. Possibly, due to this reason, models which are nonlinear in the mean have not become as popular in analyzing financial data. The ARCH models do not have this drawback because changes in volatility are represented by changes in the conditional variance, linking volatility to a natural measure of risk.

3.4. *Other interpretation*

Continuing with the question of why ARCH is so prevalent in empirical studies, there are a number of other interesting explanations, such as Mizrach's (1990) learning model and Stock's (1988) time deformation hypothesis. Mizrach (1990) developed a model of asset pricing and learning in which ARCH disturbances evolve out of the decision problem of economic agents. He showed that errors made by the agents during the learning process are highly persistent, and that the current errors are dependent on all past errors. This leads the conditional variance to have an ARCH like structure with a long lag.

Stock (1988) established the link between time deformation and ARCH models. Any economic variable, in general, evolves on an 'operational' time scale, while in practice it is measured on a 'calendar' time scale. And this inappropriate use of a calendar time scale may lead to volatility clustering since relative to the calendar time, the variable may evolve more quickly or slowly [see Diebold (1986a)]. Stock (1988) showed that a time deformation model of a random variable ε_t can be approximated by

$$\varepsilon_t = \rho_t \varepsilon_{t-1} + \nu_t, \qquad \nu_t \mid \Psi_{t-1} \sim N(0, h_t),$$

where $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$. Stock also established that when a relatively long

segment of operational time has elapsed during a unit of calendar time, ρ_t is small and h_t is large, i.e., the time varying autoregressive parameter is inversely related to the conditional variance.

A number of researchers investigated the empirical relationship between autocorrelation and volatility, see for example Kim (1989), Sentana and Wadhwani (1990), Oedegaard (1991) and LeBaron (1992). Oedegaard found that the first order autocorrelation of the Standard and Poor's (S & P) 500 daily index decreased over time, which he attributed to the introduction of new financial markets, such as options and futures on the index. However, when ARCH was explicitly introduced into the model, the evidence of time varying autocorrelation became very weak. The other papers detected the simultaneous presence of autocorrelation and ARCH, and found them to be inversely related. LeBaron (1992) used the following model

$$\begin{aligned} y_t &= a + f(h_t)y_{t-1} + \varepsilon_t \\ \varepsilon_t \mid \Psi_{t-1} &\sim N(0, h_t) \\ f(h_t) &= b_0 + b_1 e^{-h_t/b_2}, \end{aligned} \tag{3.12}$$

where h_t was specified as a GARCH(1, 1) model. The function $f(\cdot)$ took account of the changing autocorrelation parameter. For estimation, LeBaron set b_2 to the sample variances of the various series he considered. Since,

$$\frac{\mathrm{d}f(h_t)}{\mathrm{d}h_t} = -\frac{b_1}{b_2}\,\mathrm{e}^{-h_t/b_2},$$

the coefficient b_1 measures the influence of volatility on autocorrelation. For the S & P 500 composite daily index from January 1928 to May 1990, the estimate of b_1 was 0.36 with a t-value of 11.70. When the sample was divided into three subsamples, the estimate of b_1 did not change very much. For other data series, he used the weekly return for the S & P 500 index, the Center for Research and Securities Prices (CRSP) value weighted index, the Dow index and IBM returns. The general result was that lower correlations were connected with periods of high volatility. As possible explanations, LeBaron mentioned nontrading and the accumulation of news. Some stocks do not trade close to the end of the day and information arriving during that period is reflected on the next day's trading. This induces serial correlation. At the same time, nontrading results in overall lower trade volume, which has a strong positive relationship with volatility. When new information reaches the market very slowly, for traders the optimal action is to do nothing until enough information is accumulated. This leads to low trade volume and high correlation. Finding the exact causes of serial correlation and its relationship with volatility is still an open empirical problem. The relationship noted in (3.12) requires further investigation and some other models need to be examined.

4. Extensions of the model

In the original exposition of the ARCH model, it was natural for Engle (1982) to assume that the conditional variance function was linear in the squared errors and that the conditional distribution was normal. He acknowledged, however, that the linearity and conditional normality assumptions may not be appropriate in particular applications. Subsequent empirical work has born this out. In this section, we survey alternative formulations of the conditional variance function and conditional distribution which have proven useful in applied research.

4.1. *Nonlinear conditional variance*

One of the first difficulties encountered with the linear ARCH model was that the estimated α_i coefficients were frequently found to be negative. To avoid this problem Geweke (1986) and Milhoj (1987a) suggested the log ARCH model [see also Pantula (1986)]

$$\log(h_t) = \alpha_0 + \alpha_1 \log(\varepsilon_{t-i}^2) + \cdots + \alpha_q \log(\varepsilon_{t-q}^2). \tag{4.1}$$

Taking the exponential of both sides of (4.1), $h_t = e^{(\cdot)}$ is strictly positive, and therefore, no inequality restrictions are required for the α_i's to ensure that the conditional variance is strictly positive. To determine whether the linear model (2.3) or the logarithmic model (4.1) provided a better fit to actual data, Higgins and Bera (1992) proposed a nonlinear ARCH (NARCH) model, which still requires non-negativity restrictions, but includes linear ARCH as a special case and log ARCH as a limiting case. They specified the conditional variance as

$$h_t = [\phi_0(\sigma^2)^\delta + \phi_1(\varepsilon_{t-1}^2)^\delta + \cdots + \phi_q(\varepsilon_{t-q}^2)^\delta]^{1/\delta}, \tag{4.2}$$

where $\sigma^2 > 0$, $\phi_i \geqslant 0$, $\delta > 0$ and the ϕ_i's are such that $\Sigma_{i=0}^q \phi_i = 1$. The motivation of the NARCH model can be seen by rearranging (4.2) to give

$$\frac{h_t^\delta - 1}{\delta} = \phi_0 \frac{(\sigma^2)^\delta - 1}{\delta} + \phi_1 \frac{(\varepsilon_{t-1}^2)^\delta - 1}{\delta} + \cdots + \phi_q \frac{(\varepsilon_{t-q}^2)^\delta - 1}{\delta}, \tag{4.3}$$

from which it is evident that the NARCH model is a Box-Cox power transformation of both sides of the linear ARCH model. It is apparent that when $\delta = 1$, (4.3) is equivalent to the linear ARCH model and that as $\delta \to 0$, (4.3) approaches the log ARCH model (4.1). Higgins and Bera (1992) estimated (4.2) with weekly exchange rates and found that δ was typically significantly less than one and much closer to zero, indicating that the data favored the logarithmic rather than the linear ARCH model. Extensions of the above functional forms to the GARCH process are straightforward.

A possible limitation of the functional forms described above is that the conditional variance function h_t is symmetric in the lagged ε_t's. Nelson (1991) suggested that a symmetric conditional variance function may be inappropriate for modeling the volatility of returns on stocks because it cannot represent a phenomena known as the 'leverage effect', which is the negative correlation

between volatility and past returns. In a symmetric ARCH model, h_t is not affected by the sign of ε_{t-i}, and therefore h_t is uncorrelated with past errors. To rectify this, Nelson began by defining $\varepsilon_t = \eta_t\sqrt{h_t}$, where η_t is independent and identically distributed with $E(\eta_t) = 0$ and $\mathrm{Var}(\eta_t) = 1$. He suggested that in the general ARCH formulation

$$h_t = h(\eta_{t-1}, \ldots, \eta_{t-q}, h_{t-1}, \ldots, h_{t-p}), \tag{4.4}$$

h_t can be viewed as a stochastic process in which η_t serves as the forcing variable for both the conditional variance and the error. He then chose $h(\cdot)$ in (4.4) to produce the desired dependence. To avoid nonnegativity restrictions on parameters, Nelson maintained the logarithmic specification (4.1) and proposed

$$\log(h_t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i g(\eta_{t-i}) + \sum_{i=1}^{p} \beta_i \log(h_{t-i}), \tag{4.5}$$

where

$$g(\eta_t) = \theta\eta_t + \gamma[\,|\eta_t| - E\,|\eta_t|\,]. \tag{4.6}$$

The conditional variance (4.5), with (4.6), is known as exponential GARCH (EGARCH). It is easy to see that the sequence $g(\eta_t)$ is independent with mean zero and constant, if finite, variance. Therefore, (4.5) represents a linear ARMA model for $\log(h_t)$ with innovation $g(\eta_t)$. The properties of the EGARCH model are determined by the careful construction of the function (4.6). These properties are:

(1) The innovation to the conditional variance is piecewise linear in η_t, with slopes $\alpha_i(\theta + \gamma)$ when η_t is positive and $\alpha_i(\theta - \gamma)$ when η_t is negative. This produces the asymmetry in the conditional variance.
(2) The first term in (4.6) allows for correlation between the error and future conditional variances. For example, suppose that $\gamma = 0$ and that $\theta < 0$. Then a negative η_t will cause the error to be negative and the current innovation to the variance process to be positive.
(3) The second term in (4.6) produces the ARCH effect. Suppose that $\theta = 0$ and $\gamma > 0$. Whenever the absolute magnitude of η_t exceeds its expected value, the innovation $g(\eta_t)$ is positive. Therefore, large shocks increase the conditional variance.

Nelson (1991) fitted the EGARCH model to the excess daily return on the CRSP value-weighted stock market index from July 1962 to December 1987. The estimate of θ was -0.118 and had a standard error of 0.008, confirming a highly significant negative correlation between the excess return and subsequent volatility. For other applications of the EGARCH model see, for example, Pagan and Schwert (1990) and Taylor (1990).

Building on the success of the EGARCH model to represent asymmetric responses in the conditional variance to positive and negative errors, a series of papers have proposed other ARCH models which allow a very general shape in the conditional variance function. Although these models are parametric, and

estimated by maximum likelihood, they are nonparametric in spirit because the shape of the conditional variance function is largely determined by the data itself. Glosten, Jagannathan and Runkle (1991) and Zakoian (1990) independently suggested a conditional standard deviation of the form

$$\sqrt{h_t} = \alpha_0 + \sum_{i=1}^{q} \alpha_i^+ \varepsilon_{t-i}^+ - \sum_{i=1}^{q} \alpha_i^- \varepsilon_{t-i}^-, \tag{4.7}$$

where $\varepsilon_t^+ = \max\{\varepsilon_t, 0\}$ and $\varepsilon_t^- = \min\{\varepsilon_t, 0\}$ [see also Rabemananjara and Zakoian (1993)]. The parameters are constrained by $\alpha_0 > 0$, $\alpha_i^+ \geqslant 0$, and $\alpha_i^- \geqslant 0$ for $i = 1, \ldots, q$, to ensure that the conditional standard deviation is positive. Zakoian referred to this formulation as a threshold ARCH (TARCH) model because the coefficient of ε_{t-i} changes when ε_{t-i} crosses the *threshold* of zero. When $\varepsilon_{t-i} > 0$, the conditional standard deviation is linear in ε_{t-i} with slope α_i^+ and when $\varepsilon_{t-i} < 0$, the conditional standard deviation is linear in ε_{t-i} with slope $-\alpha_i^-$. This allows for asymmetry in the conditional variance in the fashion of EGARCH.

Gourieroux and Monfort (1992) proposed that a step function over the support of the conditioning error vector $\underline{\varepsilon}_{t-1} = (\varepsilon_{t-1}, \ldots, \varepsilon_{t-q})'$ can approximate a highly nonlinear conditional variance function. Let $A_1, \ldots, A_m$ be a partition of the support of ε_t. Gourieroux and Monfort considered a conditional variance of the form

$$h_t = \alpha_0 + \sum_{i=1}^{m} \sum_{j=1}^{q} \alpha_{ij} 1_{A_i}(\varepsilon_{t-j}), \tag{4.8}$$

where $1_A(\varepsilon_t)$ is the indicator function of the set A, which takes the value one when $\varepsilon_t \in A$ and zero otherwise. They describe (4.8) as a qualitative TARCH (QTARCH) model because the conditional variance is determined by the region in R^q in which $\underline{\varepsilon}_{t-1}$ lies, rather than by the continuous values of the elements of $\underline{\varepsilon}_{t-1}$.

Engle and Ng (1991) provided a summary of asymmetric ARCH models and introduced several new models of their own. They concentrated on the GARCH(1, 1) process and the functional relationship $h_t = h(\varepsilon_{t-1})$, which they term the 'news impact curve'. They proposed the parametric models

$$h_t = \alpha_0 + \alpha_1(\varepsilon_{t-1} + \gamma)^2 + \beta h_{t-1} \tag{4.9}$$

$$h_t = \alpha_0 + \alpha_1(\varepsilon_{t-1}/h_{t-1}^{1/2} + \gamma)^2 + \beta h_{t-1} \tag{4.10}$$

$$h_t = \alpha_0 + \alpha_1(\varepsilon_{t-1} + \gamma h_{t-1}^{1/2})^2 + \beta h_{t-1}. \tag{4.11}$$

In the standard GARCH(1, 1) model, while holding h_{t-1} constant, h_t is a parabola in ε_{t-1} that takes its minimum at $\varepsilon_{t-1} = 0$. In the conditional variance function (4.9), the introduction of the parameter γ shifts the parabola horizontally so that the minimum occurs at $\varepsilon_{t-1} = -\gamma$. This produces asymmetry because if, for example, $\gamma < 0$, then $h_t = h(-\varepsilon_{t-1})$ exceeds $h_t = h(\varepsilon_{t-1})$ for $\varepsilon_{t-1} > 0$. The model (4.10) is similar to (4.9), except that the conditional variance is quadratic in the standardized error $\varepsilon_{t-1}/h_{t-1}^{1/2}$. In (4.11), the minimum of h_t

occurs at $-\gamma h_{t-1}^{1/2}$, which varies with the information set. Engle and Ng (1991) also proposed a very flexible functional form, which is similar to the QTARCH model but is piecewise linear over the support of ε_{t-1} rather than a step function as in (4.8). They characterized this model as 'partially nonparametric' (PNP). They partitioned the support of ε_{t-1} into intervals, where the boundaries of the intervals are $\{\tau_{m^-}, \ldots, \tau_{-1}, 0, \tau_1, \ldots, \tau_{m^+}\}$, and m^- is the number of intervals below zero and m^+ is the number of intervals above zero. Engle and Ng then specified

$$h_t = \alpha + \sum_{i=0}^{m^+} \theta_i P_{it}(\varepsilon_{t-1} - \tau_i) + \sum_{i=0}^{m^-} \delta_i N_{it}(\varepsilon_{t-1} - \tau_{-i}) + \beta h_{t-1} \qquad (4.12)$$

where the variables P_{it} and N_{it} are defined as

$$P_{it} = \begin{cases} 1 & \text{if } \varepsilon_{t-1} > \tau_i \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad N_{it} = \begin{cases} 1 & \text{if } \varepsilon_{t-1} < \tau_{-i} \\ 0 & \text{otherwise} \end{cases}.$$

From (4.12), h_t will be linear with a different slope over each interval. For example, if ε_{t-1} is positive and lies in the interval (τ_i, τ_{i+1}), then the slope coefficient is $\theta_1 + \cdots + \theta_i$. Engle and Ng chose the τ_i's to be multiples of the unconditional standard deviation of the series.

Engle and Ng (1991) also conducted an experiment to compare the ability of asymmetric ARCH models to represent the conditional variance of stock returns. Using daily observations on the Japanese TOPIX stock index from January 1980 to September 1987, Engle and Ng fitted GARCH(1, 1) versions of the EGARCH and TARCH models, and the models given by (4.9), (4.10) and (4.11). All of the fitted models confirmed the presence of the leverage effect. But using a series of diagnostic tests, which we describe in section 9, Engle and Ng concluded that the simple parametric models (4.9), (4.10) and (4.11) significantly underestimated the volatility produced by large negative errors. The EGARCH and TARCH models, however, adequately represented this 'negative size' effect. Engle and Ng also estimated the PNP model and used the fitted conditional variance function as a baseline by which to compare the other asymmetric ARCH models. Relative to the prediction of the PNP model, the three models given in (4.9), (4.10) and (4.11), again underpredicted volatility for large negative ε_{t-1} and overpredicted volatility for large positive ε_{t-1}. The fitted conditional variance functions of the EGARCH and TARCH model were very close to the PNP's, but the EGARCH significantly overstated volatility for extremely large negative ε_{t-1}. Although based on only one data set, Engle and Ng's results indicate that for a parsimonious and highly parametric model, EGARCH can represent an asymmetric conditional variance remarkably well. Whether any inadequacies in the EGARCH functional form for representing the volatility of stock returns justifies the additional computational effort of estimating a more flexible model like the TARCH, QTARCH or PNP models, may largely depend on the peculiarities of the individual data set and the ultimate purpose of the empirical analysis.

In the context of estimating risk primia, Pagan and Hong (1991) suggested that no parametric functional form is sufficiently general to represent the diverse types of data which display conditional heteroskedasticity. Using data from French, Schwert and Stambaugh (1987) and Engle, Lilien and Robins (1987), Pagan and Hong (1991) used a nonparametric kernel estimator of the conditional variance and demonstrated that the nonparametric estimators give different conclusions about the effect of the risk premium on asset returns than do the standard parametric ARCH models. In Section 8, we will briefly discuss the nonparametric approach suggested in Pagan and Hong (1991). Undoubtedly, as research on ARCH phenomena continues, new empirical regularities of conditional heteroskedasticty will be discovered and new functional forms will be put forward to model these regularities.

4.2. *Nonnormal conditional distribution*

As described in section 2.3, an attractive feature of the ARCH process is that even though the conditional distribution of the error is normal, the unconditional distribution is nonnormal with tails thicker than the normal distribution. In spite of this property, early empirical work with ARCH models for daily exchange rates indicated that the implied unconditional distributions of estimated ARCH models were not sufficiently leptokurtic to represent the distribution of returns. In the linear regression model with conditionally normal ARCH errors, suppose that $\hat{\varepsilon}_t$ and $\hat{h}_t$ are estimates of the error and conditional variance. Then the standardized residuals $\hat{\varepsilon}_t/\hat{h}_t^{1/2}$ should be approximately $N(0, 1)$. Hsieh (1988, 1989), McCurdy and Morgan (1988) and Milhøj (1987b), however, demonstrated for a variety of currencies that the sample kurtosis coefficient of the standardized residuals often exceeded three.

The frequent inability of the conditionally normal ARCH model to pass this simple diagnostic test has led to the use of conditional distributions more general than the normal distribution. Let $\eta_t = \varepsilon_t/h_t^{1/2} = (y_t - x_t'\xi)/h_t^{1/2}$ be the standardized error. In this approach, the conditional distribution of η_t is specified as

$$\eta_t \mid \Psi_{t-1} \sim f(\eta, \theta), \tag{4.13}$$

where θ is a low dimension parameter vector whose value determines the shape of the conditional distribution of η_t. In the conditionally normal ARCH model, θ is absent and $f(\eta)$ is simply the $N(0, 1)$ density. Bollerslev (1987) was the first to adopt this approach and specified $f(\eta, \theta)$ as a conditional t distribution, where θ, a scalar, is the degrees of freedom of the distribution. The conditional t distribution allows for heavier tails than the normal distribution and, as the degrees of freedom goes to infinity, includes the normal distribution as a limiting case. Bollerslev suggested that a test for conditional normality could be conducted by testing that the reciprocal of the degrees of freedom equals zero. Using the daily rate of return in the spot market for the German mark and the British pound from March 1980 to January 1985, Bollerslev estimated

GARCH(1, 1) models with conditional t distributions and rejected the hypothesis of conditional normality. The sample kurtosis coefficients of the standardized residuals were very close to the kurtosis coefficients of the t distribution evaluated at the estimated parameters. With the German mark, for example, the sample kurtosis coefficient of the standardized residuals $\hat{\varepsilon}_t/\hat{h}_t^{1/2}$ was 4.63, while the implied kurtosis of the fitted t distribution was 4.45, suggesting that the conditional t distribution adequately accounted for the excess kurtosis in the unconditional distribution. Bollerslev presented similar results for the daily rate of return on five S & P 500 stock indexes. Engle and Bollerslev (1986), Baillie and Bollerslev (1989) and Hsieh (1989) also found that employing a conditional t distribution helped account for the excess kurtosis in daily exchange rates. Spanos (1991) demonstrated that if the observed data is assumed to have an uncorrelated multivariate t distribution, the conditional distribution of the error also has a t distribution, with an ARCH structure for the variance.

Other specifications of the conditional distribution of the ARCH process have been suggested. Nelson (1991) employed a generalized error distribution (GED) with his EGARCH model. The GED encompasses distributions with tails both thicker and thinner than the normal, and includes the normal as a special case. For a stock price index, Nelson found evidence of nonnormality in the conditional distribution, but concluded that tails of the estimated GED were not sufficiently thick to account for a large number of outliers in the data. Lee and Tse (1991) suggested that not only may the conditional distribution be leptokurtotic, but also asymmetric. They argued that for rates of return which cannot be negative, such as nominal interest rates, the conditional distribution should be skewed to the right. They used a distribution based on the first three terms of the Gram-Charlier series, that allows for both thick tails and skewness. Using interest rates from the Singapore Asian dollar market, Lee and Tse estimated their model but failed to find any evidence of skewness.

As with parametric specifications of the conditional variance function, no single parametric specification of the conditional density (4.13) appears to be suitable for all conditionally heteroskedastic data. Applications in which none of the above conditional distributions appear to be appropriate are often encountered. For example, Hsieh (1989) found that a GARCH(1, 1) model with either a conditional t or a conditional GED distribution could not adequately represent daily returns on the British pound nor the Japanese yen. Hansen (1992) recently suggested an approach to allow more flexibility in the conditional distribution within a parametric framework. While conventional ARCH models allow the mean and variance to be time varying, Hansen argues that other properties of the conditional distribution, such as skewness and kurtosis, should also be time varying and a function of the current information set. More formally, Hansen proposed that the conditional distribution (4.13) should be generalized to

$$\eta_t \mid \Psi_{t-1} \sim f(\eta, \theta_t), \tag{4.14}$$

where the parameters θ_t, which determine the shape of the conditional density,

are themselves a function of the elements of the information set Ψ_{t-1}. Hansen refers to (4.14) as an autoregressive conditional density (ARCD) model.

To illustrate the use of an ARCD model, Hansen estimated a GARCH model with a conditional t distribution and time varying degrees of freedom for the monthly excess holding yield on short-term U.S. Treasury securities. To allow the tail thickness of the conditional distribution to be determined by the information set, the degrees of freedom were parameterized as logistic transformation of a quadratic function of the lagged error and the difference between the one-month yield and the instantaneous yield. A likelihood ratio test rejected a conditional t distribution with constant degrees of freedom in favor of the ARCD model. A time plot of the estimated degrees of freedom revealed that the degrees of freedom varied considerably over time, with a mean of about 5, but frequently reaching 30 and 2.1, the upper and lower bounds imposed by the logistic transformation.

5. Forecasting with ARCH models

A very important use of ARCH models is the evaluation of the accuracy of forecasts. In standard time series methodology which uses conditionally homoskedastic ARMA processes, the variance of the forecast error does not depend on the current information set. If the series being forecasted displays ARCH, the current information set can indicate the accuracy by which the series can be forecasted. Below, we demonstrate how this is possible. Engle and Kraft (1983) were the first to consider the effect of ARCH on forecasting. Baillie and Bollerslev (1992) extended many of their results. The discussion below draws heavily from these two papers.

5.1. *Measurement of forecast uncertainty*

We illustrate the effects of ARCH on the measurement of forecast uncertainty in the context of predicting a univariate linear time series. Consider the ARMA(k, l) process

$$\phi(B)y_t = \theta(B)\varepsilon_t \tag{5.1}$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_k B^k$, $\theta(B) = 1 + \theta_1 B + \cdots + \theta_l B^l$, B is the backshift operator and ε_t is a GARCH(p, q) process. We consider forecasting the value of the process s periods from an origin t, which is given by

$$y_{t+s} = \sum_{i=1}^{k} \phi_i y_{t+s-i} + \sum_{i=1}^{l} \theta_i \varepsilon_{t+s-i} + \varepsilon_{t+s}.$$

The optimal predictor is the mean of y_{t+s} conditional on the available information up to period t, Ψ_t. Because $E(\varepsilon_{t+s} \mid \Psi_t) = 0$, the optimal predictor is

$$E(y_{t+s} \mid \Psi_t) = \sum_{i=1}^{k} \phi_i E(y_{t+s-i} \mid \Psi_t) + \sum_{i=1}^{l} \theta_i E(\varepsilon_{t+s-i} \mid \Psi_t), \tag{5.2}$$

where:

(a) $E(y_{t+s-i} \mid \Psi_t)$, for $i < s$, is given recursively by (5.2)
(b) $E(y_{t+s-i} \mid \Psi_t) = y_{t+s-i}$, for $i \geqslant s$
(c) $E(\varepsilon_{t+s-i} \mid \Psi_t) = 0$ for, $i < s$,
(d) $E(\varepsilon_{t+s-i} \mid \Psi_t) = \varepsilon_{t+s-i}$, for $i \geqslant s$.

Expression (5.2) is the standard recursive relation for the optimal point forecast of the conventional ARMA process, which can be found for example in Box and Jenkins (1976, p. 129). Therefore, the presence of ARCH does not affect the way in which the point forecast is constructed. This is because ARCH introduces dependence in high order moments and only affects the uncertainty in the point forecast.

To consider the effect of ARCH on the uncertainty of the point forecast, we require an expression for the forecast error. Assuming the roots of $\phi(B) = 1 - \phi_1 B - \cdots - \phi_k B^k$ lie outside the unit circle, the ARMA process (5.1) can be inverted to give

$$y_{t+s} = \sum_{i=0}^{\infty} \gamma_i \varepsilon_{t+s-i}, \tag{5.3}$$

where γ_i is the coefficient of B^i in the expansion of $\phi(B)^{-1}\theta(B)$. Using the moving average representation, the optimal predictor is

$$E(y_{t+s} \mid \Psi_t) = \sum_{i=s}^{\infty} \gamma_i \varepsilon_{t+s-i}. \tag{5.4}$$

Let $e_{t,s}$ be the forecast error from origin t with forecast horizon s. Subtracting (5.4) from (5.3), the forecast error

$$e_{t,s} = y_{t+s} - E(y_{t+s} \mid \Psi_t) = \sum_{i=0}^{s-1} \gamma_i \varepsilon_{t+s-i} \tag{5.5}$$

is seen to be a linear combination of the innovations from $t+1$ to the horizon $t+s$. The uncertainty in a forecast can be measured by the variance of the forecast error conditional on the information Ψ_t used to construct the forecast. Using (5.5), the conditional variance of the forecast error is

$$\mathrm{Var}(e_{t,s} \mid \Psi_t) = \sum_{i=0}^{s-1} \gamma_i^2 E(\varepsilon_{t+s-i}^2 \mid \Psi_t). \tag{5.6}$$

Expression (5.6) reveals how ARCH affects the conditional variance of the forecast error. When ARCH is present, $E(\varepsilon_{t+s-i}^2 \mid \Psi_t)$ will depend on the elements of Ψ_t and will, in general, be time varying. In contrast, for a conditionally homoskedastic model, in which $E(\varepsilon_{t+s-i}^2 \mid \Psi_t) = \sigma_\varepsilon^2$, the variance of the forecast error reduces to

$$\mathrm{Var}(e_{t,s} \mid \Psi_t) = \sigma_\varepsilon^2 \sum_{i=0}^{s-1} \gamma_i^2.$$

In this case, the variance of the forecast error does not depend upon the elements of the information set Ψ_t, but only on the length of the forecast horizon s.

To make (5.6) operational, for constructing prediction intervals for example, it is necessary to evaluate the expectations $E(\varepsilon_{t+s-i}^2 \mid \Psi_t)$. This can be done by using the ARMA(m, p) representation of the square of a GARCH(p, q) process [see equation (2.12)]:

$$\varepsilon_{t+s}^2 = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i)\varepsilon_{t+s-i}^2 - \sum_{i=1}^{p} \beta_i \nu_{t+s-i} + \nu_{t+s}.$$

The conditional expectation is then seen to be

$$E(\varepsilon_{t+s}^2 \mid \Psi_t) = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i)E(\varepsilon_{t+s-i}^2 \mid \Psi_t) - \sum_{i=1}^{p} \beta_i E(\nu_{t+s-i} \mid \Psi_t), \quad (5.7)$$

where:

(a) $E(\varepsilon_{t+s-i}^2 \mid \Psi_t)$, for $i < s$, is given recursively by (5.7)
(b) $E(\varepsilon_{t+s-i}^2 \mid \Psi_t) = \varepsilon_{t+s-i}^2$, for $i \geqslant s$
(c) $E(\nu_{t+s-i} \mid \Psi_t) = 0$, for $i < s$
(d) $E(\nu_{t+s-i} \mid \Psi) = \nu_{t+s-i}$, for $i \geqslant s$.

The expression for $E(\varepsilon_{t+s}^2 \mid \Psi_t)$ in (5.7) is completely analogous to the optimal predictor $E(y_{t+s} \mid \Psi_t)$ in (5.2).

As an example of constructing estimates of the variance of the forecast, consider the stationary AR(1) process

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \ |\phi_1| < 1,$$

where ε_t is a GARCH(1, 1) process. The optimal point forecasts follow the recursion

$$E(y_{t+s} \mid \Psi_t) = \phi_1 E(y_{t+s-1} \mid \Psi_{t-1})$$

where the first period forecast is $E(y_{t+1} \mid \Psi_t) = \phi_1 y_t$. Inverting the AR(1) process, the coefficients in (5.4) are seen to be $\gamma_i = \phi_1^i$. Therefore, from (5.6), the variance of the forecast error is

$$\mathrm{Var}(e_{t,s} \mid \Psi_t) = \sum_{i=0}^{s-1} \phi_1^{2i} E(\varepsilon_{t+s-i}^2 \mid \Psi_t), \ s \geqslant 1,$$

where the expectations can be computed recursively by

$$E(\varepsilon_{t+s}^2 \mid \Psi_t) = \alpha_0 + (\alpha_1 + \beta_1)E(\varepsilon_{t+s-1}^2 \mid \Psi_t), \quad s > 1,$$

with the initial expectation $E(\varepsilon_{t+1}^2 \mid \Psi_t) = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 h_t$.

To further demonstrate the effect of ARCH on the construction of forecast intervals, in Figure 8 we present prediction intervals for the generated ARCH(4) data that was displayed in Figure 6. Since the process has a constant conditional mean of zero, the optimal point forecast of the series is simply zero. The prediction intervals are then $\pm 2E(\varepsilon_{t+s}^2 \mid \Psi_t)^{1/2}$, where $E(\varepsilon_{t+s}^2 \mid \Psi_t)$ is given in (5.7). The information sets Ψ_{100} and Ψ_{400} on which the intervals are based were

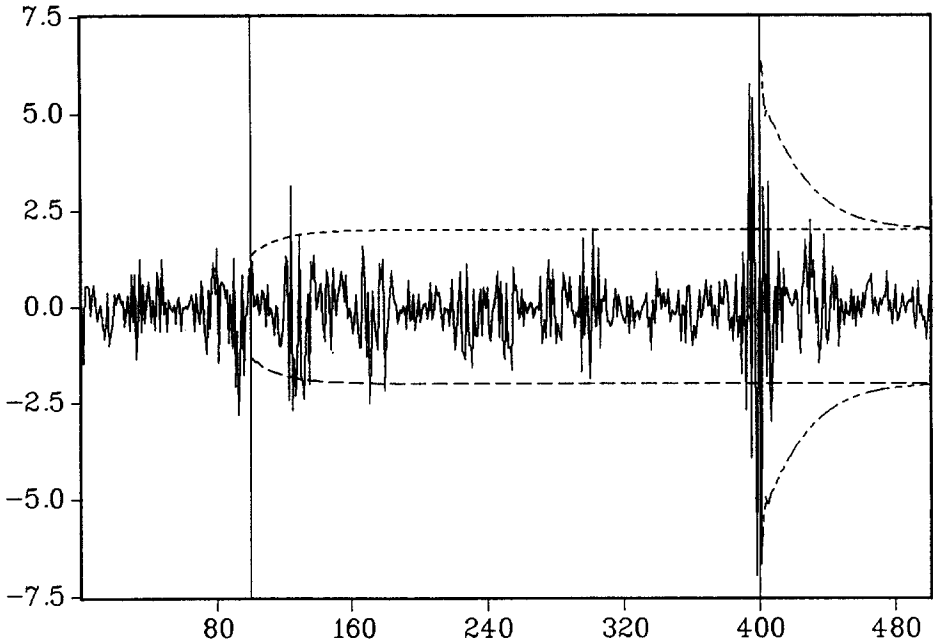

Figure 8. Two standard error prediction intervals.

chosen because $t = 100$ was a tranquil period for the series and $t = 400$ was a volatile period. Notice that for Ψ_{100}, the prediction intervals increase monotonically, indicating that uncertainty increases with the forecast horizon, while for Ψ_{400}, the intervals decrease, indicating that certainty in the point forecast increases with the forecast horizon. Although at first sight it may seem peculiar that the accuracy of a forecast can increase as we forecast further into the future, this phenomena is very plausible in the context of ARCH models. If the forecast is constructed in a highly volatile period, an ARCH model will convey that volatility is likely to persist for several periods. But as the forecast horizon increases, the volatility is likely to return to its typical level, and therefore, the expected accuracy of the point forecast actually increases as we forecast further ahead. Also notice that for both information sets, the intervals converge to $\pm 2\sigma_\varepsilon$, where σ_ε is the unconditional standard deviation of the process. In an important class of ARCH models, the conditional variances of the forecast errors may not converge to the unconditional variance of the process. We characterize this class of models in the next section.

5.2. *Persistence in variance*

When ARCH is present, current information is useful for assessing the accuracy by which a process can be forecasted. It is interesting to consider how the available information Ψ_t affects the forecast uncertainty as the forecast horizon

s increases. For $s > p$, the conditional variance (5.7) of the innovation to the forecast error reduces to

$$E(\varepsilon_{t+s}^2 \mid \Psi_t) = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i) E(\varepsilon_{t+s-i}^2 \mid \Psi_t), \tag{5.8}$$

which is a linear difference equation for the sequence $\{E(\varepsilon_{t+s}^2 \mid \Psi_t)\}_{s=p+1}^{\infty}$. If the roots of $1 - (\alpha_1 + \beta_1)Z - \cdots - (\alpha_m + \beta_m)Z^m = 1 - \alpha(Z) - \beta(Z)$ lie outside the unit circle, the solution sequence of (5.8) converges to

$$\lim_{s \to \infty} E(\varepsilon_{t+s}^2 \mid \Psi_{t-1}) = \frac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_q - \beta_1 - \cdots - \beta_p},$$

which is the unconditional variance of the innovation. In this case, as the forecast horizon becomes very large, the conditioning set provides no information about the variance of ε_{t+s}. If, however, the roots of $1 - \alpha(Z) - \beta(Z)$ lie on or inside the unit circle, this will not be the case. For example, consider a GARCH(1, 1) process with $1 - \alpha(Z) - \beta(Z)$ having a unit root, implying $\alpha_1 + \beta_1 = 1$. Then (5.8) reduces to

$$E(\varepsilon_{t+s}^2 \mid \Psi_t) = \alpha_0 + E(\varepsilon_{t+s-1}^2 \mid \Psi_t).$$

which has the solution

$$E(\varepsilon_{t+s}^2 \mid \Psi_t) = s\alpha_0 + E(\varepsilon_t^2 \mid \Psi_t).$$

Therefore, when $\alpha_1 + \beta_1 = 1$, the conditional variance grows linearly with the forecast horizon and the dependence on the information set persists through $E(\varepsilon_t^2 \mid \Psi_t)$.

Engle and Bollerslev (1986) were the first to consider GARCH processes with $\alpha(1) + \beta(1) = 1$ as a distinct class of models, which they termed integrated GARCH (IGARCH). They pointed out the similarity between IGARCH processes and processes that are integrated in the mean. For a process that is integrated in the mean, that is one that must be differenced to induce stationarity, a shock in the current period affects the level of the series into the indefinite future. In an IGARCH process, a current shock persists indefinitely in conditioning the future variances. The IGARCH model is important because a remarkable empirical regularity, repeatedly observed in applied work, is that the estimated coefficients of a GARCH conditional variance sum close to one. For example, Baillie and Bollerslev (1989) estimated GARCH(1, 1) models for six U.S. exchange rates and found $\hat{\alpha}_1 + \hat{\beta}_1$ ranging between 0.94 and 0.99 for the six series. Bollerslev and Engle (1989) considered multivariate IGARCH processes and defined a concept of co-integration in variance which they termed *co-persistence*. A set of univariate IGARCH processes are co-persistent if there exists a linear combination of the processes which is *not* integrated in variance. Nelson (1990) has cautioned that drawing an analogy with processes that are integrated in the mean, however, may be somewhat misleading. As described in section 2.3, Nelson (1990a) demonstrated that although IGARCH models are not weakly stationary, because they have infinite variances, they can be strongly

stationary. Processes that are integrated in the mean are not stationary in any sense.

The consistent finding of very large persistence in variance in financial time series is perplexing because currently no theory predicts that this should be the case. Lamoureux and Lastrapes (1990b) argued that large persistence may actually represent misspecification of the variance and result from structural change in the unconditional variance of the process, as represented by changes in α_0 in (2.7). A discrete change in the unconditional variance of a process produces clustering of large and small deviations which may show up as persistence in a fitted ARCH model. To illustrate this possibility, Lamoureux and Lastrapes used 17 years of daily returns on the stocks of 30 randomly selected companies and estimated GARCH(1, 1) models holding α_0 constant and allowing α_0 to change discretely over sub-periods of the sample. For the restricted model, in which α_0 is constant, the average estimate of $\alpha_1 + \beta_1$ for the 30 companies was 0.978, while for the unrestricted model, in which α_0 is allowed to change, the average estimate fell to 0.817. Lamoureux and Lastrapes also present Monte Carlo evidence which demonstrated that the MLE of $\alpha_1 + \beta_1$ has a large positive bias when changes in the unconditional variance are ignored.

6. Multivariate ARCH models

As economic variables are inter-related, generalization of univariate models to the multivariate and simultaneous set-up is quite natural — this is more so for ARCH models. Apart from possible gains in efficiency in parameter estimation, estimation of a number of financial 'coefficients' such as the systematic risk (beta coefficient) and the hedge ratio, requires sample values of covariances between relevant variables. The motivation for multivariate ARCH also stems from the fact that many economic variables react to the same information, and hence, have nonzero covariances conditional on the information set. For simplicity, we concentrate on two variables, and using our earlier notation as in (3.4a), let

$$\begin{aligned} y_{1t} &= \mu_{1t} + \tau_1 I_t^{1/2} \nu_{1t} \\ y_{2t} &= \mu_{2t} + \tau_2 I_t^{1/2} \nu_{2t}, \end{aligned} \tag{6.1}$$

where y_{1t} and y_{2t} are two time series, driven by the same directing process I_t, and

$$\begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & c_{12} \\ c_{12} & 1 \end{pmatrix}\right].$$

Then we have

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} \Bigg| \, \Psi_{t-1}, \; I_t \sim N\left[\begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix}, \; I_t \begin{pmatrix} \tau_1^2 & c_{12}\tau_1\tau_2 \\ c_{12}\tau_1\tau_2 & \tau_2^2 \end{pmatrix}\right]. \tag{6.2}$$

This is the bivariate counterpart of (3.5) and provides a rationale behind higher dimensional ARCH processes. As discussed in section 3.2, several special cases can be derived from (6.2).

Let us now consider an $N \times 1$ vector time series $y_t = (y_{1t}, \ldots, y_{Nt})'$. We can express a general form of the multivariate GARCH model as

$$y_t \mid \Psi_{t-1} \sim N(\mu_t, H_t),$$

where μ_t is an $N \times 1$ vector and H_t is an $N \times N$ matrix. Of course, the conditional distribution could be something other than normal. As in the univariate case, one main problem is the specification of H_t. In fact we will soon realize, the problem is more serious here. Even if we confine ourselves to linear specifications for multivariate ARCH, there are many choices.

To express H_t in a vector form, we use the 'vech' notation which stacks the lower triangular elements of a symmetric matrix in a column. A somewhat general form of H_t can be written as

$$\operatorname{vech}(H_t) = \operatorname{vech}(\Sigma) + \sum_{i=1}^{q} A_i \operatorname{vech}(\varepsilon_{t-i}\varepsilon_{t-i}') + \sum_{i=1}^{p} B_i \operatorname{vech}(H_{t-i}), \tag{6.3}$$

where $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$, Σ is an $N \times N$ positive definite matrix and A_i and B_i are $N(N+1)/2 \times N(N+1)/2$ matrices. This is a direct generalization of our earlier univariate GARCH(p, q) model given in equation (2.7). Representation (6.3) is called the 'vech representation' of a multivariate ARCH model. For $N = 2$ and $p = q = 1$, (6.3) takes the form

$$\operatorname{vech}(H_t) = \begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix}. \tag{6.4}$$

The two main problems concerning the specification of H_t are that it should be positive definite for all possible realizations and some exclusion restrictions should be imposed so that the number of parameters to be estimated is not very large. Formulation (6.3) will be difficult to estimate, for it has $[N(N+1)/2]\,[1 + [N(N+1)/2](p+q)]$ parameters, which for the special bivariate case (6.4) amounts to 21 parameters — still too large.

Engle, Granger and Kraft (1984) published the first paper on multivariate ARCH models. They considered a bivariate ARCH model which was (6.4) without the lagged h_t components. For that model, they showed necessary conditions for H_t to be positive definite are

$$\begin{gathered} \sigma_{11} > 0,\ \sigma_{22} > 0,\ \sigma_{11}\sigma_{22} - \sigma_{12}^2 > 0, \\ a_{11} \geqslant 0,\ a_{13} \geqslant 0,\ a_{31} \geqslant 0,\ a_{33} \geqslant 0, \\ a_{11}a_{33} - a_{22}^2 \geqslant 0, \\ a_{11}a_{13} - \tfrac{1}{4}\, a_{12}^2 \geqslant 0,\ a_{11}a_{31} - a_{21}^2 \geqslant 0, \\ a_{31}a_{33} - \tfrac{1}{4}\, a_{32}^2 \geqslant 0,\ a_{13}a_{33} - a_{23}^2 \geqslant 0. \end{gathered} \tag{6.5}$$

Note that in (6.3) and (6.4), each $h_{ij,t}$ depends on lagged squared residuals and past variances of all the variables in the system. One simple assumption that could be made to reduce the number of parameters is to specify that a conditional variance depends only on its own lagged squared residuals and lagged values. The assumption amounts to taking A_i and B_i to be diagonal matrices. In that case, conditions in (6.5) reduce to

$$\begin{aligned} &\sigma_{11} > 0,\ \sigma_{22} > 0,\ \sigma_{11}\sigma_{22} - \sigma_{12}^2 > 0 \\ &a_{11} \geqslant 0,\ a_{33} \geqslant 0,\ a_{11}a_{33} - a_{22}^2 \geqslant 0. \end{aligned} \tag{6.6}$$

From (6.3), the 'diagonal representation' for $p = q = 1$ can be expressed as

$$h_{ij,t} = \sigma_{ij} + a_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + b_{ij}h_{ij,t-1} \qquad i, j = 1, 2, \ldots, N \tag{6.7}$$

This form was used by Bollerslev, Engle and Wooldridge (1988) for their analysis of returns on bills, bonds and stocks, and by Baillie and Myers (1991) and Bera, Garcia and Roh (1991) for hedge ratio estimation in commodity markets.

The diagonal representation appears to be too restrictive, and at the same time, positive definiteness of the resulting H_t, in general, is not easy to check and also difficult to impose at the estimation stage [see (6.6)]. Baba, Engle, Kraft and Kroner (1990) suggested the following parameterization, known as the 'BEKK representation', which is almost guaranteed to be positive definite

$$H_t = \Sigma + \sum_{i=1}^{q} A_i^{*\prime}\varepsilon_{t-i}\varepsilon_{t-i}'A_i^* + \sum_{i=1}^{p} B_i^{*\prime}H_{t-i}B_i^*, \tag{6.8}$$

where A_i^* and B_i^* are $N \times N$ matrices. If Σ is positive definite, then so is H_t. For $N = 2$ and $p = q = 1$, (6.8) will have only 11 parameters compared to the 21 parameters of the vech representation (6.4), as (6.8) now takes the form

$$\begin{aligned} \begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{bmatrix} &= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} + \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \\ &\quad \times \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix} + \begin{bmatrix} b_{11}^* & b_{12}^* \\ b_{21}^* & b_{22}^* \end{bmatrix}' \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{12,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} b_{11}^* & b_{12}^* \\ b_{21}^* & b_{22}^* \end{bmatrix}. \end{aligned} \tag{6.9}$$

By taking the vech of (6.8), it can be shown that under certain nonlinear restrictions on A_i^*, B_i^*, A_i and B_i, (6.3) and (6.8) are equivalent [see Baba, Engle, Kraft and Kroner (1990).] The relationship is easily seen by comparing the special cases (6.4) and (6.9).

Bollerslev (1990) introduced an attractive way to simplify H_t. He assumed that the conditional correlation matrix of $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ is constant and expressed H_t as

$$H_t = \operatorname{diag}(\sqrt{h_{11,t}}, \ldots, \sqrt{h_{NN,t}})R \operatorname{diag}(\sqrt{h_{11,t}}, \ldots, \sqrt{h_{NN,t}}), \tag{6.10}$$

where R is the time invariant correlation matrix. When $N = 2$, this representation takes the form

$$H_t = \begin{bmatrix} \sqrt{h_{11,t}} & 0 \\ 0 & \sqrt{h_{22,t}} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \sqrt{h_{11,t}} & 0 \\ 0 & \sqrt{h_{22,t}} \end{bmatrix}, \tag{6.11}$$

where $|\rho| < 1$ is the correlation coefficient between ε_{1t} and ε_{2t}, and the individual variances $h_{11,t}$ and $h_{22,t}$ are assumed to be standard univariate ARCH(p, q) processes, for example

$$h_{11,t} = \sigma_{11} + \sum_{i=1}^{q} \alpha_{1i}\varepsilon_{1,t-i}^2 + \sum_{i=1}^{p} \beta_{1i}h_{11,t-i}. \tag{6.12}$$

For positive definiteness of H_t in this constant correlation representation, we need $\sigma_{ii} > 0$, $\alpha_{ij} \geqslant 0$, $\beta_{ik} \geqslant 0$, $i = 1, \ldots, N$, $j = 1, \ldots, q$, $k = 1, \ldots, p$. Many of the recent applications of bivariate ARCH use this representation [see, for example, Baillie and Bollerslev (1990); Baillie and Myers (1991); Bera, Garcia and Roh (1990); Bollerslev (1990); Kroner and Claessens (1991) and Kroner and Sultan (1991)]. However, it is quite obvious that constant correlation is a strong assumption. Bera and Roh (1991) suggested a test for the constant correlation hypothesis and found that the null hypothesis is rejected for many financial data series.

None of the above forms take account of the motivation behind multivariate ARCH discussed earlier. Diebold and Nerlove (1989) were the first to exploit the theory that only a few factors influence all the variables $(y_1, \ldots, y_N)$ and their conditional variances. They suggested an one factor multivariate ARCH model represented as

$$y_t = \lambda F_t + \eta_t, \tag{6.13}$$

where $\eta_t = (\eta_{1t}, \ldots, \eta_{Nt})$, $\eta_{it} \sim (0, \sigma_{ii})$, $i = 1, \ldots, N$ and the unobservable factor F_t is conditionally distributed as $F_t \mid \Psi_{t-1} \sim N(0, h_t)$. Then

$$\mathrm{Var}(y_t \mid \Psi_{t-1}) = h_t\lambda\lambda' + \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{NN}) \tag{6.14}$$

and we can specify a univariate GARCH process for h_t. The effect of the common factor F_t on y_i is measured by λ_i $(i = 1, \ldots, N)$. Their application of this model to seven weekly exchange rate series gave superior results compared to seven separate univariate ARCH models.

Harvey, Ruiz and Sentana (1992) presented a more general unobserved component model, that includes (6.13) as a special case, and allows for richer dynamics in the mean of y_t. They consider the model

$$y_t = Z_t\alpha_t + \Lambda\eta_t + \eta_t^*$$

in which α_t is a $m \times 1$ state vector that evolves according to the transition equation

$$\alpha_t = T_t\alpha_{t-1} + \Gamma\varepsilon_t + \varepsilon_t^*,$$

where T_t and Z_t are observable matrices, and the $m \times 1$ vector ε_t^* and the $N \times 1$ vector η_t^* are conditionally homoskedastic. Conditional heteroskedasticity is introduced through the scalar processes ε_t and η_t, which are assumed to follow univariate ARCH processes. The state space formulation provides a convenient representation for estimation and prediction by means of a Kalman filter. Higgins and Majin (1992) applied a univariate version of this model to measure

the time-varying volatility of both the latent ex ante real interest rate and the market's forecast errors of inflation from the observable ex post real interest rate.

To discuss Engle's (1987) multivariate ARCH model with a k-factor structure, we start with a slight generalization of the BEKK representation (6.8), namely

$$H_t = \Sigma + \sum_{j=1}^{k} \left[\sum_{i=1}^{q} A_{ij}^{*\prime} \varepsilon_{t-i} \varepsilon_{t-i}' A_{ij}^{*} + \sum_{i=1}^{p} B_{ij}^{*\prime} H_{t-i} B_{ij}^{*} \right], \tag{6.15}$$

where A_{ij}^{*} and B_{ij}^{*} are $N \times N$ matrices and $k \ll N$. Engle obtained a very parsimonious structure for H_t by restricting the rank of A_{ij}^{*} and B_{ij}^{*} to one [see also Lin (1992)]. More specifically, he assumed that these matrices have the same left and right eigenvectors, g_j and f_j, i.e.,

$$A_{ij}^{*} = \alpha_{ij} f_j g_j' \text{ and } B_{ij}^{*} = \beta_{ij} f_j g_j'$$

with

$$f_j g_l' = \begin{cases} 0 & \text{for } j \neq l \\ 1 & \text{for } j = l, \end{cases}$$

where f_j and g_i are $N \times 1$ vectors, $j, l = 1, 2, \ldots, N$. Using these expressions for A_{ij}^{*} and B_{ij}^{*} in (6.15), we have the k-factor GARCH(p, q) model

$$H_t = \Sigma + \sum_{j=1}^{k} g_j g_j' \left[\sum_{i=1}^{q} \alpha_{ij}^2 f_j' \varepsilon_{t-i} \varepsilon_{t-i}' f_j + \sum_{i=1}^{p} \beta_{ij}^2 f_j' H_{t-i} f_j \right]. \tag{6.16}$$

In the Diebold and Nerlove model, the factor is an unobserved latent variable, while in the k-factor GARCH model, the j-th factor, F_{jt}, is a linear combination of the residuals, namely $F_{jt} = f_j' \varepsilon_t$. Therefore,

$$\operatorname{Var}(F_{jt} \mid \Psi_{t-1}) = f_j' H_t f_j = h_{jt}^{*} \text{ (say)}. \tag{6.17}$$

Substituting (6.16) into (6.17), we have

$$h_{jt}^{*} = f_j' \Sigma f_j + \sum_{i=1}^{q} \alpha_{ij}^2 F_{j,t-i}^2 + \sum_{i=1}^{p} \beta_{ij}^2 h_{j,t-i}^{*}.$$

Therefore, each h_{jt}^{*} has a GARCH(p, q) structure, $j = 1, \ldots, k$. This enables us to express (6.16) as

$$\begin{aligned} H_t &= \Sigma + \sum_{j=1}^{k} g_j g_j' [h_{jt}^{*} - f_j' \Sigma f_j] \\ &= \Sigma^{*} + \sum_{j=1}^{k} g_j g_j' h_{jt}^{*}, \end{aligned} \tag{6.18}$$

where

$$\Sigma^{*} = \Sigma - \sum_{j=1}^{k} g_j g_j' f_j' \Sigma f_j.$$

Expression (6.18) demonstrates that the conditional variance of ε_t is regulated

completely by the conditional variance of the k factors. Also for $k = 1$, we can see the similarities between (6.14) and (6.18). Engle (1987), Kroner (1988), Lin (1992) and Engle, Ng and Rothschild (1990) discussed other interesting properties of the k-factor GARCH model. Two notable applications of this model are Engle, Ng and Rothschild (1990) to explain the excess return for treasury bills and Ng, Engle and Rothschild (1992) to study the behavior of stock returns.

Lin (1992) examined the finite sample properties of various estimators, such as maximum likelihood and two stage estimators, for the factor GARCH(1, 1) model through simulation. Estimators were found to be, in general, unbiased. And, as predicted by asymptotic theory, the maximum likelihood estimators were most efficient. The major problems in estimation are devising methods for finding the number of factors and the factor weights.

7. ARCH-M models

It is reasonable to expect that the mean and variance of a return move in the same direction. Denoting the mean by μ_t, we can express this idea as

$$\mu_t = \xi_0 + \delta g(h_t),$$

where $g(h_t)$ is a monotonic function of the conditional variance h_t, with $g(\alpha_0) = 0$. In finance models, $\delta g(h_t)$ represents the risk premium, that is, the increase in the expected rate of return due to an increase in the variance of the return. Existence of risk premia in foreign exchange markets and the term structure of interest rates have been studied extensively. Most of the earlier studies concentrated on detecting a *constant* risk premium. ARCH in the mean (ARCH-M) models, first proposed by Engle, Lilien and Robbins (1987), provide a new approach by which we can test for and estimate a *time varying* risk premium. In the regression set-up, an ARCH-M model is specified as

$$y_t = x_t'\xi + \delta g(h_t) + \varepsilon_t, \tag{7.1}$$

where

$$\varepsilon_t \mid \Psi_{t-1} \sim N(0, h_t)$$

and h_t is an ARCH or GARCH process. The presence of h_t in the conditional mean is the distinctive feature of this model.

To examine the properties of the ARCH-M model, we consider a simple version of (7.1), namely

$$y_t = \delta h_t + \varepsilon_t,$$

where $\varepsilon_t \mid \Psi_{t-1} \sim N(0, h_t)$ and $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$. We can then write

$$y_t = \delta\alpha_0 + \delta\alpha_1 \varepsilon_{t-1}^2 + \varepsilon_t,$$

where ε_t follows an ARCH(1) process. From this expression, and using

$E(\varepsilon_{t-1}^2) = \alpha_0/(1-\alpha_1)$, it immediately follows that

$$E(y_t) = \delta\alpha_0\left(1 + \frac{\alpha_0}{1-\alpha_1}\right),$$

which can be viewed in finance models as the unconditional expected return for holding a risky asset. Similarly,

$$\mathrm{Var}(y_t) = \frac{\alpha_0}{1-\alpha_1} + \frac{(\delta\alpha_1)^2 2\alpha_0^2}{(1-\alpha_1)^2(1-3\alpha_1^2)}.$$

In the absence of a risk premium $\mathrm{Var}(y_t) = \alpha_0/(1-\alpha_1)$. Therefore, the second component of $\mathrm{Var}(y_t)$ is due to the presence of a risk premium which makes y_t more dispersed. Finally, the ARCH-M effect makes y_t serially correlated, since [see Hong (1991)]

$$\rho_1 = \mathrm{Corr}(y_t, y_{t-1}) = \frac{2\alpha_1^3\delta^2\alpha_0}{2\alpha_1^2\delta^2\alpha_0 + (1-\alpha_1)(1-3\alpha_1^2)}$$

$$\rho_k = \mathrm{Corr}(y_t, y_{t-k}) = \alpha_1^{k-1}\rho_1, \quad k = 2, 3, \ldots.$$

From the expressions for ρ_1 and ρ_2, it is easily seen that the admissible region for (ρ_1, ρ_2) will be very restrictive. Bollerslev (1988) obtained similar results for the GARCH process. ARCH-M models introduce some interesting problems in terms of estimation and testing which will be discussed in the following sections.

In most applications, $g(h_t) = \sqrt{h_t}$ has been used [see, for example, Domowitz and Hakkio (1985) and Bollerslev, Engle and Wooldridge (1988)], although Engle, Lilien and Robins (1987) found that $g(h_t) = \log(h_t)$ worked better in their estimation of the time varying risk premia in the term structure. Pagan and Hong (1991) commented that the use of $\log(h_t)$ is problematic since for $h_t < 1$, $g(h_t)$ will be negative and also when $h_t \to 0$, the effect on y_t will be infinite.

8. Estimation

The most commonly used estimation procedure for ARCH models has been the maximum likelihood approach. The log likelihood function of the standard ARCH regression model

$$y_t \mid \Psi_{t-1} \sim N(x_t'\xi, h_t)$$

is given by

$$l(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\theta),$$

where

$$l_t(\theta) = \text{const.} - \frac{1}{2}\log(h_t) - \frac{\varepsilon_t^2}{2h_t^2} \tag{8.1}$$

and $\theta = (\xi', \gamma')'$. Here ξ and γ denote the conditional mean and conditional variance parameters respectively. One attractive feature of this normal likelihood

function is that the information matrix is block diagonal between the parameters ξ and γ. To see this, note that the (i,j)th element of the off-diagonal block of the information matrix can be written as

$$\frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial^2 l_t}{\partial\xi_i\partial\gamma_j}\right] = \frac{1}{T}\sum_{t=1}^{T} E\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\xi_i}\frac{\partial h_t}{\partial\gamma_j}\right]. \tag{8.2}$$

If h_t is a symmetric function of the lagged errors in the sense of Engle (1982), then the last expression in square brackets is anti-symmetric and, therefore, has expectation zero. The ARCH, GARCH, log ARCH and NARCH models given in (2.3), (2.7), (4.1) and (4.2) respectively, are all symmetric according to the definition of Engle (1982).

The advantage of having this block diagonality is that, under the likelihood framework, estimation and testing for the mean and variance parameters can be carried out separately [see Engle (1982, p. 996), Bollerslev (1986, p. 317) and Higgins and Bera (1992, p. 996)]. Most of the applied work on ARCH models use the Berndt, Hall, Hall and Hausman (1974) algorithm (BHHH) to maximize $l(\theta)$. Starting from estimates of the r'th iteration, the $(r+1)$'th step of the BHHH algorithm can be written as

$$\xi^{(r+1)} = \xi^{(r)} + \left[\sum_{t=1}^{T}\left(\frac{\partial l_t}{\partial\xi}\right)\left(\frac{\partial l_t}{\partial\xi}\right)'\right]^{-1}\sum_{t=1}^{T}\frac{\partial l_t}{\partial\xi}$$

and

$$\gamma^{(r+1)} = \gamma^{(r)} + \left[\sum_{t=1}^{T}\left(\frac{\partial l_t}{\partial\gamma}\right)\left(\frac{\partial l_t}{\partial\gamma}\right)'\right]^{-1}\sum_{t=1}^{T}\frac{\partial l_t}{\partial\gamma},$$

where the derivatives are evaluated at $\xi^{(r)}$ and $\gamma^{(r)}$. The block diagonality of the information matrix no longer holds for the ARCH-M model in (7.1) and the asymmetric models like AARCH in (3.3) and EGARCH in (4.5). For these models, the BHHH algorithm needs to be carried out jointly for both the conditional mean and variance parameters.

For most applications, it is very difficult to justify the conditional normality assumption in (8.1). Therefore, the log likelihood function $l(\theta)$ may be misspecified. However, we can still obtain estimates of ξ and γ by maximizing $l(\theta)$ and such estimators are called quasi maximum likelihood estimators (QMLE). Weiss (1986a) was the first to study the asymptotic properties of the QMLE of ARCH models. His important finding was that as long as the first two conditional moments are correctly specified, ξ and γ will be consistently estimated even if the conditional normality assumption is violated. To state the asymptotic distribution of the QMLE $\hat{\theta} = (\hat{\xi}', \hat{\gamma}')'$, let us denote

$$A = -\frac{1}{T}E\left[\frac{\partial^2 l(\theta_0)}{\partial\theta\,\partial\theta'}\right] \text{ and } B = \frac{1}{T}E\left[\left(\frac{\partial l(\theta_0)}{\partial\theta}\right)\left(\frac{\partial l(\theta_0)}{\partial\theta}\right)'\right], \tag{8.3}$$

where θ_0 is the true value of the parameter. Then under certain regularity

conditions

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{L} N(0, A^{-1}BA^{-1})$$

and consistent estimators of A and B are given by

$$\hat{A} = -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 l_t(\hat{\theta})}{\partial\theta\,\partial\theta'} \text{ and } \hat{B} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial l_t(\hat{\theta})}{\partial\theta}\right)\left(\frac{\partial l_t(\hat{\theta})}{\partial\theta}\right)'.$$

Robust inference about θ can be achieved using this result. If the normality assumption is correct, $A = B$ and valid inference can be drawn using either $\hat{A}^{-1}$ or $\hat{B}^{-1}$ as the covariance matrix estimator. Bollerslev and Wooldridge (1992) generalized the univariate ARCH results of Weiss (1986a) to the multivariate GARCH case under a different set of regularity conditions. Although the specification of a univariate ARCH model in Weiss (1986a) was very general, he assumed a finite fourth moment of the error term. As an example, for the ARCH(2) model this condition requires [see Bollerslev (1986)]

$$3\alpha_1^2 + 3\alpha_2^2 - 3\alpha_2^3 + 3\alpha_1^2\alpha_2 + \alpha_2 < 1,$$

which might be difficult to justify in practice. For higher order ARCH and GARCH models, the condition will restrict models to a small part of the parameter space. Bollerslev and Wooldridge (1992) did not assume finiteness of the fourth moment, but instead, they required $l_t(\theta)$ and its derivatives to satisfy a uniform weak law of large numbers which are not easy to verify. Lumsdaine (1991a) established the consistency and asymptotic normality of the QMLE of the GARCH(1, 1) and IGARCH(1, 1) models under a different set of assumptions. Her basic conditions are

$$E\left[\frac{\partial h_t}{\partial\theta}\, h_t^{-1}\right] < \infty \text{ and } E\left[\left(\frac{\partial h_t}{\partial\theta}\right)\left(\frac{\partial h_t}{\partial\theta}\right)' h_t^{-2}\right] < \infty$$

and these are easy to verify. For simplicity, consider an ARCH(1) model

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2.$$

Then

$$E\left[\left(\frac{\partial h_t}{\partial\alpha_1}\right)^2\right] = E(\varepsilon_{t-1}^4)$$

might not exist, and yet

$$E\left[\left(\frac{\partial h_t}{\partial\alpha_1}\right)^2 h_t^{-2}\right] = E\left[\frac{\varepsilon_{t-1}^4}{\alpha_0^2 + \alpha_1^2\varepsilon_{t-1}^4 + 2\alpha_0\alpha_1\varepsilon_{t-1}^2}\right]$$

may exist because here both the numerator and the denominator grow at the same rate. In terms of the standardized variable $\varepsilon_t^* = \varepsilon_t/\sqrt{h_t}$, Lumsdaine's assumptions are that ε_t^* is IID and drawn from a symmetric and unimodal density with 32 finite moments. Lee and Hansen (1991) obtained similar results under the somewhat weaker condition that ε_t^* is stationary and ergodic with a

bounded fourth conditional moment. Lumsdaine (1991a) and Lee and Hansen (1991) showed that the QMLE for the IGARCH(1, 1) model has the same asymptotic distribution as that of the GARCH(1, 1) model. This result is important because it establishes that the difficulties of the unit root model is not encountered with IGARCH.

Lee (1991) extended all these asymptotic distribution results to the GARCH(1, 1)-M and IGARCH(1, 1)-M models. As discussed in Lee, these models pose additional difficulties because unlike the GARCH model, the conditional variance of a GARCH-M model is a *nonlinear* difference equation. To see this, note from (7.1) that for a GARCH(1, 1)-M model

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \\ &= \alpha_0 + \alpha_1 [y_{t-1} - x_t'\xi - \delta g(h_{t-1})]^2 + \beta_1 h_{t-1}. \end{aligned}$$

To avoid the difficulties associated with nonlinear difference equations, Lee (1991) used the fact that at the true parameter value, h_t is a linear difference equation.

Engle and González-Rivera (1991) pointed out that although the QMLE is consistent and asymptotically normal, it can be inefficient. They demonstrated that the loss of efficiency due to misspecification could be severe when the true distribution is asymmetric and a normal quasi likelihood function is used. They suggested a semiparametric approach in which one maximizes the log likelihood function $l(\theta) = 1/T \Sigma_t l_t(\theta)$, where the nonconstant part of $l_t(\theta)$ in (8.1) is replaced by

$$-\tfrac{1}{2} \log(h_t) + \log(g(\varepsilon_t / h_t^{1/2})). \tag{8.4}$$

Engle and Gonález-Rivera (1991) used a nonparametric method to estimate the function $g(\cdot)$. To do this, they started with an initial estimator of θ, obtained $\hat{\varepsilon}_t / \hat{h}_t^{1/2}$, used these values to estimate $g(\cdot)$ and then maximized $l(\theta)$ to get a revised estimate of θ. The procedure was repeated until it converged. Their Monte Carlo results indicated that there is substantial gain in efficiency from using the semiparametric method over QMLE.

Another attractive way to estimate ARCH models without assuming normality is to apply the generalized method of moments (GMM) approach as advocated by Rich, Raymond and Butler (1991) [see also Sabau (1987, 1988)]. For simplicity, consider an ARCH(1) model and define the following two errors

$$\begin{aligned} \varepsilon_t &= y_t - x_t'\xi \\ \nu_t &= \varepsilon_t^2 - \alpha_0 - \alpha_1 \varepsilon_{t-1}^2 \\ &= (y_t - x_t'\xi)^2 - \alpha_0 - \alpha_1 (y_{t-1} - x_{t-1}'\xi)^2. \end{aligned}$$

Then the GMM estimator is obtained from the following two moment conditions

$$E(\varepsilon_t \mid Z_t) = 0 \text{ and } E(\nu_t \mid Z_t) = 0,$$

where z_t is a set of predetermined variables. The asymptotic distribution of the GMM estimator follows directly from the general formula in Hansen (1982).

Weiss (1986a) and Pantula (1988) studied the asymptotic properties of least squares estimators which also do not require a normality assumption. They proved the consistency and asymptotic normality of such estimators. However, as can be expected, least squares estimators are less efficient than GMM estimators and MLE's with a correct likelihood function. It would be interesting to compare the finite sample properties of all of these estimators.

We previously mentioned the importance of correct specification of the conditional variance function h_t. All the forms of h_t we discussed in Section 4 are fully parametric. Pagan and Hong (1991) argued that the existing parametric forms are not very convincing due to the lack of optimizing theory in their formulation. They advocated nonparametric estimation of h_t, as originally suggested by Pagan and Ullah (1988). They even recommended estimating both the conditional mean, m_t, and the conditional variance, h_t, nonparametrically since misspecification in the conditional mean might exaggerate the variation in h_t. In the statistics literature, many nonparametric techniques are available. For their empirical application, Pagan and Hong (1991) used the kernel method and the Fourier series approximation of Gallant (1982). These procedures estimate the first two conditional moments by relating them to the past values of y_t. If r lags of y_t are chosen, then m_t and h_t can be estimated by using the formulae

$$\hat{m}_t = \sum_{\substack{i=1 \\ i \neq t}}^{T} \omega_{it} y_i, \text{ and } \hat{h}_t = \sum_{\substack{i=1 \\ i \neq t}}^{T} \omega_{it} y_i^2 - \hat{m}_t^2,$$

where ω_{it} are the kernel weights. For the Gaussian kernel

$$\omega_{it} = \frac{\varkappa_{it}}{\sum_{\substack{i=1 \\ i \neq t}}^{T} \varkappa_{it}},$$

where $\varkappa_{it} = \exp\{-\frac{1}{2} \Sigma_{s=1}^{r} h_s^{-2}(y_{i-s} - y_{s-t})^2\}$, h_s being the bandwidth [for details see Pagan and Hong (1991, p. 60)]. The empirical applications of Pagan and Hong (1991) showed the advantages of the nonparametric approach. They plotted the nonparametric $\hat{h}_t$ against y_{t-1} and found a high degree of nonlinearity which would be difficult to capture by simple parametric models. Also, Cox nonnested tests for parametric versus nonparametric models rejected the Engle, Lilien and Robins (1987) specification of the ARCH-M model for excess holding yields on treasury bills. The nonparametric method does require much larger data sets. Fortunately, we do have large data sets for economic and financial variables where ARCH models are generally applied. Of course, results from nonparametric analysis are not as easily interpreted in terms of response coefficients as those obtained from a parametric method. However, at the very least, nonparametric methods can point out deficiencies in the existing parametric models and offer some guidance for modification.

Not much is known about the finite sample distribution of the different estimators discussed above. Engle, Hendry and Trumble (1985), Bollerslev and Wooldridge (1992) and Lumsdaine (1991b) reported some Monte Carlo results

on the QMLE. For the GARCH(1, 1) model, Bollerslev and Wooldridge (1992) found the QMLE of α_1 to be biased upward, the QMLE of β_1 to be biased downward, and the overall estimate of $\alpha_1 + \beta_1$ to be slightly biased downward. This was consistent with the ARCH(1) results of Engle, Hendry and Trumble (1985). Lumsdaine (1991b) reported that in small samples, QMLE's are not normally distributed and rather skewed. For example, she found $\hat{\beta}_1$ to be skewed to the right. This is similar to the downward bias observed by Bollerslev and Wooldridge (1992). Lumsdaine (1991b) also observed some pile-up for the estimator of β_1. Surprisingly, the pile-up was at the zero boundary. In most applications β_1 seems to take values above 0.5, so this may not be taken as a small sample effect.

Geweke (1988a,b, 1989) argued that a Bayesian approach rather than the classical one might be more suitable for estimating ARCH models due to two distinct features of these models. First, as we noted earlier, some inequality restrictions must be imposed on the parameters to ensure positivity of h. In the classical estimation framework, these restrictions are somewhat impractical to impose. However, under the Bayesian paradigm, diffuse priors can incorporate these inequalities. Second, most of the time the main interest is not in the individual parameters, rather in h_t, which is a function of the parameters. Exact posterior distributions and means of h_t can be obtained quite easily using Monte Carlo integration with importance sampling. The recent introduction of Gibbs sampling to the Bayesian econometrics literature might make the task even easier. Geweke's successful application to inflation and stock price data demonstrated the viability of the Bayesian approach for estimating ARCH models. Unfortunately, this approach has not been pursued by other researchers.

9. Testing

The introduction of ARCH to econometrics has led to many interesting testing problems. The basic test for the ARCH model is testing for the presence of ARCH, i.e., a test for the null hypothesis H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_q = 0$ in (2.3). Engle (1982) derived the LM statistic for testing H_0, which is computed as TR^2, where T is the number of observations and R^2 is the coefficient of multiple determination from the regression of $\hat{\varepsilon}_t^2$ on a constant and $\hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-q}^2$, $\hat{\varepsilon}_t$'s being the OLS residuals from the model (2.1). Under H_0, the LM statistic asymptotically follows a χ_q^2 distribution. The structure of the test is the same as that of the Breusch and Pagan (1979) and Godfrey (1978) 'static' heteroskedasticity test in the regression model. As noted in Bera and Lee (1992), this test is also a special case of the IM test applied to the regression model (2.1) with an $AR(q)$ error structure and can be viewed as a test for randomness of the AR parameters $\phi_1, \ldots, \phi_q$.

A convenient way of looking at a general test for ARCH is to give it a moment test interpretation, the moment condition being

$$E\left(\frac{\varepsilon_t^2}{\alpha_0} - 1 \mid z_t\right) = 0, \tag{9.1}$$

where z_t is some vector of variables. For Engle's test, $z_t = (1, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots, \varepsilon_{t-q}^2)$. If the alternative model is GARCH, as given in (2.7), then z_t would be specified as $z_t = (1, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots, \varepsilon_{t-q}^2, h_{t-1}, \ldots, h_{t-p})$. When estimated under H_0, z_t becomes $z_t = (1, \hat{\varepsilon}_{t-1}^2, \hat{\varepsilon}_{t-2}^2, \ldots, \hat{\varepsilon}_{t-q}^2, \hat{\alpha}_0, \ldots, \hat{\alpha}_0)$. Therefore, the last p elements of $\hat{z}_t$ are redundant and a test for no conditional heteroskedasticity against an ARCH(q) or a GARCH(p, q) will be identical [see Bollerslev (1986) and J. H. Lee (1991)]. For the AARCH model (3.3), $z_t = (1, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots, \varepsilon_{t-q}^2, \varepsilon_{t-1}\varepsilon_{t-2}, \ldots, \varepsilon_{t-1}\varepsilon_{t-q})$ and the test is carried out by running a regression of $\hat{\varepsilon}_t^2$ on a constant and the squares and cross products of $\hat{\varepsilon}_{t-i}$, $i = 1, 2, \ldots, q$ [see Bera and Lee (1992), and Bera, Higgins and Lee (1992)].

A complication arises when H_0 is tested against an ARCH-M model given in (7.1). The conditioning set z_t is the same as in the ARCH case, namely $z_t = (1, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots, \varepsilon_{t-q}^2)$. However, we note that when the null hypothesis of no ARCH is imposed on the model, the nuisance parameter δ is not identified. This renders the information matrix to be singular under H_0, and thereby invalidates the standard distribution of the LM test. However, note that for a given value of δ, say δ^*, the LM statistic is perfectly well behaved and has the form [see Domowitz and Hakkio (1985)]

$$LM(\delta^*) = \frac{1}{2+\delta^{*2}}\, \gamma' Z \left[Z'Z - \frac{\delta^{*2}}{2+\delta^*}\, Z'X(X'X)^{-1}X'Z \right]^{-1} Z'\gamma, \qquad (9.2)$$

where γ is a $T \times 1$ vector with t-th element $\gamma_t = (\hat{\varepsilon}_t^2/\hat{\alpha}_0 - 1) + (\delta^* \hat{\varepsilon}_t/\hat{\alpha}_0)$. The second component of this LM test is due to the non-block diagonality of the information matrix between the conditional mean and variance parameters. It is clear that when $\delta^* = 0$, LM(δ^*) reduces to Engle's test for ARCH. Any arbitrary choice of δ will lead to a suboptimal test.

The same problem is faced for the NARCH model (4.2). When $\alpha_1 = \alpha_2 = \cdots = \alpha_q = 0$ in (4.2), the parameter δ becomes unidentified. For a fixed value of δ, again say δ^*, the LM statistic $S(\delta^*)$ can be computed as TR^2, where the R^2 is obtained by a regression of $\hat{\varepsilon}_t^2$ on an intercept and

$$\frac{(\hat{\varepsilon}_{t-i}^2)^{\delta^*} - 1}{\delta^*}, \quad i = 1, 2, \ldots, q.$$

Therefore, in our conditional moment test framework

$$z_t = \left[1, \frac{(\varepsilon_{t-1}^2)^{\delta^*} - 1}{\delta^*}, \ldots, \frac{(\varepsilon_{t-q}^2)^{\delta^*} - 1}{\delta^*} \right],$$

which is the Box-Cox transformation of the lagged squared residuals. It is obvious that when $\delta^* = 1$, z_t reduces to the conditioning set of Engle's test. To overcome the nonidentification of δ, Bera and Higgins (1992) followed the procedure of Davies (1977, 1987) and suggested basing the test on a critical region of the form

$$\{S = \sup_{\delta^*} S(\delta^*) > \omega\}, \qquad (9.3)$$

where ω is a suitably chosen constant. However, unlike $S(\delta^*)$, S does not have

an asymptotic χ_q^2 distribution under the null hypothesis. It is clear that if χ_q^2 critical values are used, the type-*I* error probability of the test will be too high. Davies (1987) provided an approximation to the *p*-value of the test as

$$Pr[\chi_q^2 > S] + V\frac{e^{-S/2}S^{(p-1)/2}}{2^{p/2}\Gamma p/2}, \tag{9.4}$$

where V measures the variation in $\sqrt{S(\delta)}$ over values of δ corresponding to different alternative hypotheses. This V can be estimated by

$$V = \sum_{j=1}^{R} |\sqrt{S(\delta_j)} - \sqrt{S(\delta_{j-1})}|,$$

where δ_0 and δ_R are the lower and upper bounds for δ, and $\delta_1, \delta_2, \ldots, \delta_{R-1}$ are the turning points of $\sqrt{S(\delta)}$. The second component in (9.4) can be viewed as the correction factor to the standard χ^2 *p*-value due to the scanning across a range of values of δ. Monte Carlo results and an empirical illustration presented in Bera and Higgins (1992) suggest that the above procedure is more powerful than the standard LM test for ARCH when the true process has $\delta \neq 1$. Bera and Ra (1991) applied the same technique to the ARCH-M model and obtained similar results. Hansen (1991) developed a simulation approach which approximates the asymptotic null distribution of statistics which have the structure of S. Andrews and Ploberger (1992) also considered the general problem of testing when a nuisance parameter exists only under the alternative hypothesis and derived asymptotically optimal tests in terms of weighted average power in the class of all tests with a given significance level. Andrews (1993) used this latter approach for testing the presence of conditional heteroskedasticity with GARCH(1, 1) as the alternative model.

One drawback of using the LM test principle in testing H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_q = 0$ is that it does not take account of the one-sided nature of the alternative hypothesis, i.e., that the α_i's cannot take negative values. We can expect some loss of power due to this omission, although the two-sided LM test will have the correct size asymptotically. Demos and Sentana (1991) and Lee and King (1991) suggested some one-sided versions of the LM test. Demos and Sentana's version of the one-sided LM test can be obtained as the sum of the squared *t*-ratios associated with the positive coefficients of the OLS regression of $\hat{\varepsilon}_t^2$ on 1, $\hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-q}^2$, while Lee and King's version is based on the sum of the scores $\partial l(\theta)/\partial\alpha_i$, $i = 1, \ldots, q$. Lee and King (1992) carried out a Monte Carlo study of the finite sample power properties of the two-sided and their one-sided LM statistics, and found that the one-sided version of the test has better power.

In the moment condition (9.1), the term $\varepsilon_t^2/\alpha_0 - 1$ is essentially a result of the normality assumption. If we consider a general log-density function of the form (8.4), then $\varepsilon_t^2/\alpha_0 - 1$ could be replaced by $\phi(\varepsilon_t/\sqrt{\alpha_0})\cdot\varepsilon_t/\sqrt{\alpha_0} - 1$, where $\phi = -g'/g$ is the score function. For the normal distribution $\phi(\varepsilon_t/\sqrt{\alpha_0}) = \varepsilon_t/\sqrt{\alpha_0}$. As in Engle and González-Rivera (1991), the score function can be estimated nonparametrically. As a general test for ARCH under nonnormality, we can think of running a regression of $\phi_t\varepsilon_t$ on z_t. In our discussion above of the TR^2

type test statistics, we noted various tests by changing the independent variable set z_t with the same dependent variable $\hat{\varepsilon}_t^2$. Now we can think of different dependent variables corresponding to various nonnormal distributions. For example, if we assume a double exponential distribution, we need to run the regression of $|\hat{\varepsilon}_t|$ on z_t. This is known as the Glejser (1969) test for heteroskedasticity. In the context of testing static heteroskedasticity and autocorrelation in the regression model, Bera and Ng (1991) successfully used such nonparametric tests and these could easily be adapted to ARCH models [also see Pagan and Pak (1993)].

Any general test for nonlinear dependence may also detect conditional heteroskedasticity. The BDS test of Brock, Dechert and Scheinkman (1987) is frequently used in empirical work with ARCH models [see for example Hsieh (1989), Gallant, Hsieh and Tauchen (1991) and Higgins and Bera (1992)]. The BDS test measures nonlinearity by the proportion of 'm-histories', $y_t^m = \{y_t, y_{t+1}, \ldots, y_{t+m-1}\}$, which lie in within a specified distance of one another. Hsieh (1989) and Brock, Hsieh and LeBaron (1991) demonstrated by Monte Carlo experiments that the BDS test has good power against ARCH alternatives.

All of the above tests are only for detection of the possible presence of conditional heteroskedasticity and do not provide any information regarding the form of the conditional variance function h_t. As we mentioned earlier, correct specification of h_t is very important. The accuracy of forecast intervals depends on selecting an h_t which correctly relates the future variances to the current information set. Also Pagan and Sabau (1987a) showed that an incorrect functional form for h_t can result in inconsistent maximum likelihood estimates of the conditional mean parameters. This is more likely to happen when h_t is asymmetric or for the ARCH-M models. Most of the empirical papers indirectly test for the correct specification of h_t and other accompanying assumptions by studying the properties of the standardized residuals $\hat{\varepsilon}_t^* = \hat{\varepsilon}_t / \hat{h}_t^{1/2}$. The basis of considering ε_t^* is that under our setup

$$\varepsilon_t^* = \frac{\varepsilon_t}{h_t^{1/2}} \,|\, \Psi_{t-1} \sim N(0, 1).$$

Therefore, if the model is correctly specified, $\hat{\varepsilon}_t^*$ should behave as white noise. The various diagnostic checks that are commonly used include testing the normality of $\hat{\varepsilon}_t^*$ and considering the sample autocorrelations of $\hat{\varepsilon}_t^*$. These diagnostics are helpful in detecting certain misspecifications, but we cannot expect them to be very powerful tests.

Using the Newey (1985) and Tauchen (1985) principle of moment tests, Pagan and Sabau (1987b) suggested a consistency test for ARCH models. The test is based on the moment condition $E[h_t(\varepsilon_t^2 - h_t)] = 0$. Therefore, the test could be carried out by regressing $\hat{\varepsilon}_t^2 - \hat{h}_t$ on a constant and $\hat{h}_t$, and testing whether the coefficient of $\hat{h}_t$ is zero. However, if h_t is symmetric and the model is not of the ARCH-M type, then misspecificaton of h_t will not lead to inconsistency and consequently the test will not have any power.

In most ARCH models, misspecification may not lead to inconsistency, but it might make likelihood based inference invalid. In that case, misspecification can be tested through the IM equality, i.e., by testing $A = B$ which are defined in equation (8.3). Bera and Zuo (1991) suggested such a test. One component of the IM test can be calculated by running a regression of $\hat{\varepsilon}_t^{*4} - 6\hat{\varepsilon}_t^{*2} + 3$ on the cross products of $\hat{\varepsilon}_{t-i}^{*2}$, where $\hat{\varepsilon}_t^* = \hat{\varepsilon}_t / \hat{h}_t^{1/2}$ is as above. This is essentially a test for heterokurtosis, and it can also be viewed as a test for randomness of the parameters in the specified h_t.

Another simple test for an estimated ARCH model like (2.3) is derived in Higgins and Bera (1992). The relevant null hypothesis for this is H_0: $\delta = 1$ in the NARCH model (4.2). The LM statistic for testing H_0 can be calculated by running a regression of $\hat{\varepsilon}_t^2$ on z_t, where

$$z_t = (1,\ \hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-q}^2,\ \pi_t - \hat{h}_t \log(\hat{h}_t))$$

with

$$\pi_t = \sum_{i=1}^{q} \hat{\alpha}_i \hat{\varepsilon}_{t-i}^2 \log(\hat{\varepsilon}_{t-i}^2).$$

The test can be viewed as a diagnostic check of the adequacy of the ARCH model (2.3) after it has been estimated. Starting from a different alternative model, Hall (1990) derived a simple LM test for an estimated ARCH model. The alternative distribution for his heteroskedastic normal model is that the distribution is a member of the family with semiparametric probability density functions considered by Gallant and Tauchen (1989). His test is based on the possible correlations of $\hat{\varepsilon}_t / \hat{h}_t$ and $\hat{\varepsilon}_t^2 / \hat{h}_t^2$ with the information set. Simulation results reported in his paper indicate that the LM test which uses all the information under the null hypothesis has good finite sample properties in moderate to large samples.

Engle and Ng (1991) proposed a battery of tests designed to detect misspecification of a maintained conditional variance function. Let S_t^- be a dummy variable that takes the value 1 when ε_{t-1} is negative, and zero otherwise. Similarly, let S_t^+ be a dummy variable that takes the value 1 when ε_{t-1} positive, and zero otherwise. Engle and Ng suggested standardizing the residual with the null h_t, regressing $\hat{\varepsilon}_t^{*2}$ on an intercept, S_t^-, $S_t^- \hat{\varepsilon}_{t-1}$ and $S_t^+ \hat{\varepsilon}_{t-1}$, and testing that the coefficients on the three constructed regressors are zero using an F or TR^2 statistic. The first regressor, S_t^-, represents the *sign bias* test which is intended to detect an asymmetric influence by the lagged negative and positive errors on the conditional variance which may not be incorporated in the conditional variance function specified under the null hypothesis. The second regressor, $S_t^- \hat{\varepsilon}_{t-1}$, should be significant if the impact of large negative errors versus small negative errors on the conditional variance is different from the impact implied by the null h_t. This component of the regression is called the *negative size bias* test. The third regressor, $S_t^+ \hat{\varepsilon}_{t-1}$, represents the *positive size bias test* and should detect different impacts of large positive errors versus small positive errors on the

conditional variance. Engle and Ng point out that the components of the test can be conducted individually if a particular form of misspecification is suspected.

The introduction of conditional heteroskedasticity in econometrics also lead to another interesting problem. Diebold (1986b) demonstrated that the presence of ARCH invalidates the standard asymptotic distribution theory of the sample autocorrelations, and hence of the Box-Pierce and Box-Ljung test statistics for serial correlation. For simplicity, consider the test for $\phi_1 = 0$ in

$$\begin{aligned} y_t &= x_t\xi + \varepsilon_t \\ \varepsilon_t &= \phi_1\varepsilon_{t-1} + u_t, \end{aligned} \tag{9.5}$$

where $u_t \sim (0, \sigma_u^2)$ and $\mathrm{Var}(\varepsilon_t \mid \Psi_{t-1}) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2$. As we noted, having ARCH of the above form is equivalent to ϕ_1 being random or u_t being heteroskedastic. Then the problem is equivalent to testing for the significance of a regression coefficient under heteroskedasticity. We know that the use of White's (1980) consistent estimator for the variance-covariance matrix provides asymptotically valid inference in the presence of an unknown form of heteroskedasticity. Therefore, a robust way to test $\phi_1 = 0$ is to run a regression of the OLS residuals $\hat{\varepsilon}_t$ on x_t and $\hat{\varepsilon}_{t-1}$, and test the significance of the coefficient of $\hat{\varepsilon}_{t-1}$ using White's standard error. Wooldridge (1990) suggested exactly this procedure for testing autocorrelation in the presence of ARCH [see also Davidson and MacKinnon (1985), Bollerslev and Wooldridge (1992) and MacKinnon (1992)]. Note that the standard LM approach for testing first order autocorrelation is to regress $\hat{\varepsilon}_t$ on x_t and $\hat{\varepsilon}_{t-1}$ and use a TR^2 statistic. The robust procedure involves two regressions:

(1) Run $\hat{\varepsilon}_{t-1}$ on x_t and save the residuals as $\tilde{\varepsilon}_{t-1}$.
(2) Compute TR^2 from running 1 on $\hat{\varepsilon}_t\tilde{\varepsilon}_{t-1}$.

The statistic TR^2 asymptotically follows a χ_1^2 distribution under the null hypothesis of no serial correlation. Steps (1) and (2) are equivalent to using White's consistent variance-covariance matrix estimator as mentioned above. Monte Carlo results reported in Bollerslev and Wooldridge (1992) indicate that the size of the robust version of the LM test is much closer to the nominal size than the size of the standard LM test. Bera, Higgins and Lee (1992) derived LM tests for autocorrelation which take account of specific forms of ARCH disturbances. Of course, validity of such tests depend on the correct specification of the ARCH process. In practice, the tests could be very useful by specifying different forms of conditional heteroskedasticity and then testing for serial correlation. Being fully parametric, this test could be expected to have higher power when h_t is specified correctly.

10. Epilogue

Research on modeling conditional first moments started many decades ago, and that field is still very active. The problems currently being investigated, just

to name a few, are structural change, different kinds of nonlinearities, cointegration and finite sample properties of estimators and test statistics. It is safe to say that most of the problems encountered in modeling the first moment also transmits to ARCH, i.e., conditional second moment modeling. In this survey paper, we have provided a brief account of these problems. For years to come, researchers will be occupied with topics like structural change in ARCH, co-persistence, asymptotic and finite sample statistical inference for ARCH, and procedures robust in the presence of ARCH. We have also noted that ARCH models have their own unique problems which are not present in modeling the conditional mean. Gradually, we will also see more rigorous economic foundations for ARCH models than those currently available. Therefore, the frontiers of ARCH will keep on moving further, though possibly not at the spectacular rate as we observed in its first decade of existence. The success of ARCH might even tempt researchers to model higher order moments — the third and fourth — in a systematic way. From that we might learn more about the behavior of speculative prices, and economic variables in general, a tradition started by Louis Bachelier almost a century ago.

Acknowledgements

We are grateful to three anonymous referees and editor Les Oxley for many helpful suggestions and detailed comments on the earlier version of the paper. Thanks are also due to Ken Kroner for his comments. However, we retain the responsibility for any remaining errors. Financial support from the Research Board of the University of Illinois is gratefully acknowledged.

Glossary

AARCH:	Augmented autoregressive conditional heteroskedasticity.
ARCD:	Autoregressive conditional density.
ARCH:	Autoregressive conditional heteroskedasticity.
ARCH-M:	Autoregressive conditional heteroskedasticity in the mean.
EGARCH:	Exponential autoregressive conditional heteroskedasticity.
GARCH:	Generalized autoregressive conditional heteroskedasticity.
IGARCH:	Integrated generalized autoregressive conditional heteroskedasticity.
NARCH:	Nonlinear autoregressive conditional heteroskedasticity
PNP ARCH:	Partially nonparametric autoregressive conditional heteroskedasticity.
QARCH:	Quadratic autoregressive conditional heteroskedasticity.
QTARCH:	Qualitative threshold autoregressive conditional heteroskedasticity.
TARCH:	Threshold autoregressive conditional heteroskedasticity.

References

Andel, J. (1976) Autoregressive series with random parameters, *Mathematische Operationsforschung und Statistik, Series Statistics 7*, 736–741.

Andrews, D. W. K. (1993) An introduction to econometric applications of empirical process theory for dependent random variables, *Econometric Reviews*, forthcoming.

Andrews, D. W. K. and Ploberger, W. (1992) Optimal tests when a nuisance parameter is present only under the alternative, *Cowles Foundation Discussion Paper*, No. 1015.

Baba, Y., Engle, R. F., Kraft, D. F. and Kroner, K. F. (1990) Multivariate simultaneous generalized ARCH, *Mimeo*, Department of Economics, University of California, San Diego.

Bachelier, L. (1900) Théorie de la spéculation, *Annáles de l'Ecole Normale Supérieure*, 17, 21–86.

Baillie, R. T. and Bollerslev, T. (1989) The message in daily exchange rates: a conditional variance tale, *Journal of Business and Economic Statistics*, 7, 297–305.

——(1990) A multivariate generalized ARCH approach to modeling risk premia in forward foreign exchange rate markets, *Journal of International Money and Finance*, 16, 109–124.

——(1992) Prediction in dynamic models with time-dependent conditional variances, *Journal of Econometrics*, 52, 91–113.

Baillie, R. T. and Myers, R. J. (1991) Bivariate GARCH estimation of optimal commodity futures hedge, *Journal of Applied Econometrics*, 16, 109–124.

Bekaert, G. (1992) The time-variation of expected returns and volatility in foreign exchange markets, *Mimeo*, Northwestern University.

Bera, A. K. and Higgins, M. L. (1992) A test for conditional heteroskedasticity in time series models, *Journal of Time Series Analysis*, 13, 501–519.

Bera, A. K. and Lee, S. (1993) Information matrix test, parameter heterogeneity and ARCH: a synthesis, *Review of Economic Studies*, 60, 229–240.

Bera, A. K. and Ng, P. T. (1991) Robust tests for heteroskedasticity and autocorrelation using score functions, *Mimeo*, Department of Economics, University of Illinois at Urbana-Champaign.

Bera, A. K. and Ra, S. (1991) A test for conditional heteroskedasticity in the ARCH-M model, *Mimeo*, Department of Economics, University of Illinois at Urbana-Champaign.

Bera, A. K. and Roh, J-S. (1991) A moment test of the constancy of the correlation coefficient in the bivariate GARCH model, *Mimeo*, Department of Economics, University of Illinois at Urbana-Champaign.

Bera, A. K. and Zuo, X. (1991) Specification test for a linear regression model with ARCH process, *Mimeo*, Department of Economics, University of Illinois at Urbana-Champaign.

Bera, A. K., Garcia, P. and Roh, J-S. (1991) Estimation of time-varying hedge ratios for agricultural commodities: BGARCH and random coefficient approaches, *Mimeo*, Department of Economics, University of Illinois at Urbana-Champaign.

Bera, A. K., Higgins, M. L. and Lee, S. (1990) On the formulation of a general structure for conditional heteroskedasticity, *Mimeo*, Department of Economics, University of Illinois at Urbana-Champaign.

——(1992) Interaction between autocorrelation and conditional heteroskedasticity: a random coefficient approach, *Journal of Business and Economic Statistics*, 10, 133–142.

Berndt, E. K., Hall, B. H., Hall, R. E. and Hausman, J. (1974) Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement*, 4, 653–665.

Bollerslev, T. (1986) A generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307–327.
——(1987) A conditionally heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics*, 69, 542–547.
——(1988) On the correlation structure of the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis*, 9, 121–131.
——(1990) Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach, *Review of Economics and Statistics*, 72, 498–505.
Bollerslev, T. and Engle, R. F. (1989) Common persistence in conditional variances, *Econometrica*, 61, 167–186.
Bollerslev, T. and Hodrick, R. J. (1992) Financial market efficiency tests, *Kellogg Graduate School of Management Working Paper*, No. 132.
Bollerslev, T. and Wooldridge, J. M. (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews*, 11, 143–179.
Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992) ARCH modelling in finance: a review of the theory and empirical evidence, *Journal of Econometrics*, 52, 5–59.
Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988) A capital asset pricing model with time-varying covariances, *Journal of Political Economy*, 96, 116–131.
Bougerol, P. and Picard, N. (1992) Stationarity of GARCH processes and of some nonnegative time series, *Journal of Econometrics*, 52, 115–127.
Box, G. and Jenkins, G. (1976) *Time Series Analysis, Forecasting and Control*, revised edition, Holden-Day, San Francisco.
Breusch, T. S. and Pagan, A. R. (1979) A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, 47, 239–253.
Brock, W., Dechert, W. D. and Scheinkman, J. (1987) A test for independence based on the correlation dimension, *Mimeo*, University of Wisconsin-Madison.
Brock, W., Hsieh, D. A. and LeBaron, B. (1991) *Nonlinear Dynamics, Chaos and Instability*, MIT Press, Cambridge.
Chesher, A. D. (1984) Testing for neglected heterogeneity, *Econometrica*, 52, 865–871.
Cheung, Y-W. and Pauly, P. (1990) Random coefficient modeling of conditionally heteroskedastic processes: short run exchange rate dynamics. Paper presented at the International Conference on ARCH Models, June 1990, Paris.
Clark, P. K. (1973) A subordinated stochastic process model with finite variance for speculative prices, *Econometrica*, 41, 135–156.
Davidson, R. and MacKinnon, J. G. (1985) Heteroskedasticity-robust tests in regression directions, *Annales de l'INSEE*, 59/60, 183–218.
Davies, R. B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, 64, 247–254.
——(1987) Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, 74, 33–43.
Demos, A. and Sentana, E. (1991) Testing for GARCH effects: a one-sided approach, *Mimeo*, Financial Market Group, London School of Economics.
Diebold, F. X. (1986a) Modelling the persistence of conditional variances: a comment, *Econometric Reviews*, 5, 51–56.
——(1986b) Testing for serial correlation in the presence of heteroskedasticity, *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 323–328.
——(1988) *Empirical Modeling of Exchange Rate Dynamics*, Springer Verlag, New York.
Diebold, F. X. and Nason, J. A. (1990) Nonparametric exchange rate prediction, *Journal of International Economics*, 28, 315–332.
Diebold, F. X. and Nerlove, M. (1989) The dynamics of exchange rate volatility: a multivariate latent factor ARCH model, *Journal of Applied Econometrics*, 4, 1–21.

Domowitz, I. and Hakkio, C. S. (1985) Conditional variance and the risk premium in the foreign exchange market, *Journal of International Economics*, 19, 47–66.

Drost, F. C. and Nijman, T. E. (1992) Temporal aggregation of GARCH processes, *Econometrica*, forthcoming.

Duffee, G. R. (1992) Reexamining the relationship between stock returns and stock return volatility, *Federal Reserve Board Finance and Economics Discussion Series*, No 191.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, *Econometrica*, 50, 987–1008.

——(1983) Estimates of the variance of U.S. inflation based upon the ARCH model, *Journal of Money, Credit, and Banking*, 15, 286–301.

——(1987) Multivariate ARCH with factor structures — cointegration in variance, *Discussion Paper*, No. 87-27, University of California, San Diego.

Engle, R. F. and Bollerslev, T. (1986) Modelling the persistence of conditional variances, *Econometric Reviews*, 5, 1–87.

Engle, R. F. and González-Rivera, G. (1991) Semiparametric ARCH models, *Journal of Business and Economic Statistics*, 9, 345–359.

Engle, R. F. and Kraft, D. (1983) Multiperiod forecast error variances of inflation estimated from ARCH models, in A. Zellner (ed.), *Applied Time Series Analysis of Economic Data*, Bureau of the Census, Washington D.C.

Engle, R. F. and Ng, V. K. (1991) Measuring and testing the impact of news on volatility, *Mimeo*, Department of Economics, University of California, San Diego.

Engle, R. F., Granger, C. W. J. and Kraft, D. (1986). Combining competing forecasts of inflation using a bivariate ARCH model, *Journal of Economic Dynamics and Control*, 8, 151–165.

Engle, R. F., Hendry, D. and Trumble, D. (1985) Small-sample properties of ARCH estimates and tests, *Canadian Journal of Economics*, 18, 66–93.

Engle, R. F., Ito, T. and Lin W-L. (1990) Meteor showers or heat waves? heteroskedastic intra-daily volatility in the foreign exchange market, *Econometrica*, 58, 525–542.

Engle, R. F., Lilien, D. M. and Robins, R. P. (1987) Estimating time varying risk premia in the term structure: the ARCH-M model, *Econometrica*, 55, 391–407.

Engle, R. F., Ng, V. and Rothschild, M. (1990) Asset pricing with a FACTOR-ARCH covariance structure: empirical estimates for treasury bills, *Journal of Econometrics*, 45, 213–237.

French, K. R., Schwert, G. W. and Stambaugh, R. F. (1987) Expected stock returns and volatility, *Journal of Financial Economics*, 19, 3–30.

Gallant, R. (1982) Unbiased determination of production technologies, *Journal of Econometrics*, 20, 285–323.

Gallant, A. R. and Tauchen G. (1989) Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications, *Econometrica*, 57, 1091–1129.

Gallant, A. R., Hsieh, D. A. and Tauchen, G. (1991) On fitting a recalcitrant series: the pound/dollar exchange rate, 1974–1983, in Barnett, W. A., Powell, J. and Tauchen, G. (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press, Cambridge.

Geweke, J. (1986) Modelling the persistence of conditional variances: comment, *Econometric Reviews*, 5, 57–61.

——(1988a) Comments on Poirier: operational Bayesian methods in econometrics, *Journal of Economic Perspectives*, 2, 159–166.

——(1988b) Exact inference in models with autoregressive conditional heteroskedasticity, in Berndt, E., White, H. and Barnett, W. (eds.), *Dynamic Econometric Modeling*, Cambridge University Press, Cambridge.

——(1989) Exact predictive densities in linear models with ARCH disturbances, *Journal of Econometrics*, 44, 307–325.

Glejser, H. (1969) A new test for heteroskedasticity, *Journal of the American Statistical Association*, 64, 316–323.

Glosten, L. R., Jagannathan, R. and Runkle, D. (1991) Relationship between the expected value and the volatility of the nominal excess return on stocks, *Mimeo*, Northwestern University.

Godfrey, L. G. (1978) Testing for multiplicative heteroskedasticity, *Journal of Econometrics*, 8, 227–236.

Gourieroux, C. and Monfort, A. (1992) Qualitative threshold ARCH models, *Journal of Econometrics*, 52, 159–199.

Granger, C. W. J. and Andersen, A. P. (1978) *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen.

Granger, C. W. J., Robins R. P. and Engle R. F. (1984) Wholesale and retail prices: bivariate time series modeling with forecastable error variances, in D. A. Belsley and E. Kuh (eds.), *Model Reliability*, MIT Press, Cambridge.

Hall, A. (1990) Lagrange multiplier tests for normality against semiparametric alternatives, *Journal of Business & Economic Statistics*, 8, 417–426.

Hansen, B. E. (1991) Inference when a nuisance parameter is not identified under the null hypothesis, *Rochester Center for Economic Research Working Paper*, No 296.

——(1992) Autoregressive conditional density estimation, *Rochester Center for Economic Research Working Paper*, No 332.

Hansen, L. P. (1982) Large sample properties of the method of moment estimators, *Econometrica*, 50, 1029–1054.

Harvey, A. Ruiz, E. and Sentana, E. (1992) Unobserved component time series models with ARCH disturbances, *Journal of Econometrics*, 52, 129–157.

Higgins, M. L. and Bera, A. K. (1989) A joint test for ARCH and bilinearity in the regression model, *Econometric Reviews*, 7, 171–181.

——(1991) ARCH and bilinearity as competing models for nonlinear dependence, *Mimeo*, Department of Economics, University of Wisconsin-Milwaukee.

——(1992) A class of nonlinear ARCH models, *International Economic Review*, 33, 137–158.

Higgins, M. L. and Majin, S. (1992) Measuring the volatility of the ex ante real interest rate: a structural ARCH approach, *Mimeo*, University of Wisconsin-Milwaukee.

Hong, P. Y. (1991) The autocorrelation structure for the GARCH-M process, *Economic Letters*, 37, 129–132.

Hsieh, D. A. (1988) The statistical properties of daily foreign exchange rates: 1974–1983, *Journal of International Economics*, 24, 129–145.

——(1989) Testing for nonlinear dependence in daily foreign exchange rate changes, *Journal of Business*, 62, 339–368.

Kim, C. M. (1989) Volatility effect on time series behavior of exchange rate changes, *Working Paper*, Korea Institute for International Economic Policy.

Klien, B. (1977) The demand for quality-adjusted cash balances: price uncertainty in the U.S. demand for money function, *Journal of Political Economy*, 85, 692–715.

Koedijk, K. G., Stork, P. A. and De Vries, C. G. (1992) Conditional heteroskedasticity, realignments and the European monetary system, *Mimeo*, Katholieke Universiteit Leuven.

Kroner, K. F. (1988) Estimating and testing for factor ARCH, *Mimeo*, University of Arizona.

Kroner, K. F. and Claessens, S. (1991) Optimal currency composition of external debt: applications to Indonesia and Turkey, *Journal of International Money and Finance*, 10, 131–148.

Kroner, K. F. and Sultan, J. (1991) Exchange rate volatility and time varying hedge ratios, in Rhee, S. G. and Change, R. P. (eds.), *Pacific-Basin Capital Markets Research*, Vol. II, North-Holland, Amsterdam.

Lamoureux, G. C. and Lastrapes, W. D. (1990a) Heteroskedasticity in stock return data: volume versus GARCH effects, *The Journal of Finance*, 45, 221–229.

——(1990b) Persistence in variance, structural change, and the GARCH model, *Journal of Business & Economic Statistics*, 8, 225–234.

LeBaron, B. (1992) Some relations between volatility and serial correlations in stock market returns, *Journal of Business*, 65, 199–219.

Lee, J. H. (1991) A Lagrange multiplier test for GARCH models, *Economic Letters*, 37, 265–271.

Lee, J. H. and M. L. King (1991) A locally most mean powerful based score test for ARCH and GARCH regression disturbances, *Journal of Business and Economic Statistics*, 11, 17–27.

Lee, S-W. (1991) Asymptotic properties of the maximum likelihood estimator of the GARCH-M and IGARCH-M models, *Mimeo*, Department of Economics, University of Rochester.

Lee, S-W. and Hansen, B. E. (1991) Asymptotic properties of the maximum likelihood estimator and test of the stability of parameters of the GARCH and IGARCH models, *Mimeo*, Department of Economics, University of Rochester.

Lee, T. K. Y. and Tse, Y. K. (1991) Term structure of interest rates in the Singapore Asian dollar market, *Journal of Applied Econometrics*, 6, 143–152.

Lin, W-L. (1992) Alternative estimators for factor GARCH models — a Monte Carlo comparison, *Journal of Applied Econometrics*, 7, 259–279.

Lumsdaine, R. L. (1991a) Asymptotic properties of the quasi-maximum likelihood estimator in GARCH(1, 1) and IGARCH(1, 1) models, *Mimeo*, Department of Economics, Princeton University.

——(1991b) Finite sample properties of the maximum likelihood estimator in GARCH(1, 1) and IGARCH(1, 1) models: a Monte Carlo investigation. *Mimeo*, Department of Economics, Princeton University.

MacKinnon, J. G. (1992) Model specification tests and artificial regression, *Journal of Economic Literature*, 30, 102–146.

Mandelbrot, B. (1963a) The variation of certain speculative prices, *Journal of Business*, 36, 394–419.

——(1963b) New methods in statistical economics, *Journal of Political Economy*, 71, 421–440.

——(1967) The variation of some other speculative prices, *Journal of Business*, 40, 393–413.

McCurdy, T. H. and Morgan, I. (1988) Testing the martingale hypothesis in Deutsche mark futures with models specifying the form of the heteroskedasticity, *Journal of Applied Econometrics*, 3, 187–202.

McLeod, A. I. and Li, W. K. (1983) Diagnostic checking ARMA time series models using squared-residual autocorrelations, *Journal of Time Series Analysis*, 4, 269–273.

Milhoj, A. (1985) The moment structure of ARCH processes, *Scandinavian Journal of Statistics*, 12, 281–292.

——(1987a) A multiplicative parameterization of ARCH models. *Research Report*, No. 101, Institute of Statistics, University of Copenhagen.

——(1987b) A conditional variance model for daily observations of an exchange rate, *Journal of Business and Economic Statistics*, 5, 99–103.

Mizrach, B. (1990) Learning and conditional heteroskedasticity in asset returns, *Mimeo*, Department of Finance, The Wharton School, University of Pennsylvania.

Nelson, D. B. (1990a) Stationarity and persistence in the GARCH(1, 1) model, *Econometric Theory*, 6, 318–334.

——(1990b) ARCH models as diffusion approximations, *Journal of Econometrics*, 45, 7–38.

——(1991) Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, 59, 347–370.

Nelson, D. B. and C. Q. Cao (1992) Inequality constraints in the univariate GARCH model, *Journal of Business & Economic Statistics*, 10, 229–235.

Newey, W. (1985) Maximum likelihood specification testing and conditional moment tests, *Econometrica*, 53, 1047–1070.

Ng, V. K. and Pirrong, S. C. (1992) Disequilibrium adjustment, volatility, and price discovery in spot and futures markets, *Mimeo*, University of Michigan.

Ng, V. K., Engle, R. F. and Rothschild, M. (1992) A multi-dynamic-factor model for stock returns, *Journal of Econometrics*, 52, 245–266.

Nicholls, D. F. and Quinn, B. G. (1982) *Random Coefficient Autoregressive Models: An Introduction*, Springer-Verlag, New York.

Oedegaard, B. A. (1991) Empirical tests of changes in autocorrelation of stock index returns, *Mimeo*, Graduate School of Industrial Administration, Carnegie Mellon University.

Pagan, A. R. and Hong, Y. S. (1991) Nonparametric estimation and the risk premium, in Barnett, W. A., Powell, J. and Tauchen, G. (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press, Cambridge.

Pagan, A. R. and Pak, Y. (1993) Tests for heteroskedasticity, in Maddala, G. S., Rao, C. R. and Vinod, H. D. (eds.), *Handbook of Statistics*, vol. 11, North Holland Pub. Co.

Pagan, A. R. and Sabau, H. (1987a) On the inconsistency of the MLE in certain heteroskedastic regression models, *Mimeo*, Department of Economics, University of Rochester.

——(1987b) Consistency tests for heteroskedastic and risk models, *Mimeo*, Department of Economics, University of Rochester.

Pagan, A. R. and Schwert, G. W. (1990) Alternative models for conditional stock volatility, *Journal of Econometrics*, 45, 267–290.

Pagan, A. R. and Ullah, A. (1988) The econometric analysis of models with risk terms, *Journal of Applied Econometrics*, 3, 87–105.

Pantula, S. G. (1986) Modelling the persistence of conditional variances: comment, *Econometric Reviews*, 5, 79–97.

——(1988) Estimation of autoregressive models with ARCH errors, *Sankhyā: The Indian Journal of Statistics*, 50, Series B, 119–138.

Rabemananjara, R. and Zakoian, J. M. (1993) Threshold ARCH models and asymmetries in volatility, *Journal of Applied Econometrics*, 8, 31–49.

Ray, D. (1983) On the autoregressive model with random coefficients, *Calcutta Statistical Association Bulletin*, 32, 135–142.

Rich, R. W., Raymond, J. and Butler, J. S. (1991) Generalized instrumental variables estimation of autoregressive conditional heteroskedastic models, *Economics Letters*, 35, 179–185.

Sabau, H. C. L. (1987) The structure of GMM and ML estimators in conditionally heteroskedastic models, *Working Paper*, No. 153, Faculty of Economics, Australian National University.

——(1988) *Some Theoretical Aspects of Econometric Inference with Heteroskedastic Models*, Unpublished Ph.D. dissertation, Australian National University.

Sentana, E. (1991) Quadratic ARCH models: a potential re-interpretation of ARCH models, *Mimeo*, Department of Economics and Financial Markets Group, London School of Economics.

Sentana, E. and Wadhwani, S. (1991) Feedback traders and stock returns autocorrelations: evidence from a century of daily data, *Review of Economic Studies*, 58, 547–563.

Spanos, A. (1991) A parametric approach to dynamic heteroskedasticity: the Student's t and related models, *Mimeo*, Department of Economics, Virginia Polytechnic Institute and State University.

Stock, J. H. (1988) Estimating continuous-time processes subject to time deformation, *Journal of the American Statistical Association*, 83, 77–85.

Tauchen, G. (1985) Diagnostic testing and evaluation of maximum likelihood models, *Journal of Econometrics*, 30, 415–443.

Tauchen, G. and Pitts, M. (1983) The price variability-volume relationship on speculative markets, *Econometrica*, 51, 485–505.

Taylor, S. J. (1986) *Modelling Financial Time Series*, John Wiley, Chichester, UK.

——(1990) Modelling stochastic volatility, *Mimeo*, Department of Accounting and Finance, University of Lancaster, UK.

Tong, H. (1990) *Non-linear Time Series, A Dynamical System Approach*, Clarendon Press, Oxford.

Tsay, R. S. (1987) Conditional heteroscedastic time series models, *Journal of the American Statistical Association*, 82, 590–604.

Weiss, A. A. (1986a) Asymptotic theory for ARCH models: estimation and testing, *Econometric Theory*, 2, 107–131.

——(1986b) ARCH and bilinear time series models: comparison and combination, *Journal of Business and Economic Statistics*, 4, 59–70.

White, H. (1980) A heteroscedastic-consistent covariance matrix and a direct test for heteroscedasticity, *Econometrica*, 48, 421–448.

——(1982) Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 1–25.

Wooldridge, J. M. (1990) A unified approach to robust regression-based specification tests, *Econometric Theory*, 6, 17–43.

Zakoian, J-M. (1990) Threshold heteroskedastic model, *Mimeo*, INSEE, Paris.