ARCH MODELS[a]

TIM BOLLERSLEV

Northwestern University and N.B.E.R.

ROBERT F. ENGLE

University of California, San Diego and N.B.E.R.

DANIEL B. NELSON

University of Chicago and N.B.E.R.

Contents

[a]The authors would like to thank Torben G. Andersen, Patrick Billingsley, William A. Brock, Eric Ghysels, Lars P. Hansen, Andrew Harvey, Blake LeBaron, and Theo Nijman for helpful comments. Financial support from the National Science Foundation under grants SES-9022807 (Bollerslev), SES-9122056 (Engle), and SES-9110131 and SES-9310683 (Nelson), and from the Center for Research in Security Prices (Nelson), is gratefully acknowledged. Inquiries regarding the data for the stock market empirical application should be addressed to Professor G. William Schwert, Graduate School of Management, University of Rochester, Rochester, NY 14627, USA. The GAUSS™ code used in the stock market empirical example is available from Inter-University Consortium for Political and Social Research (ICPSR), P.O. Box 1248, Ann Arbor, MI 48106, USA, telephone (313) 763–5010. Order "Class 5" under this article's name.

Handbook of Econometrics, Volume IV, Edited by R.F. Engle and D.L. McFadden

Abstract

This chapter evaluates the most important theoretical developments in ARCH type modeling of time-varying conditional variances. The coverage include the specification of univariate parametric ARCH models, general inference procedures, conditions for stationarity and ergodicity, continuous time methods, aggregation and forecasting of ARCH models, multivariate conditional covariance formulations, and the use of model selection criteria in an ARCH context. Additionally, the chapter contains a discussion of the empirical regularities pertaining to the temporal variation in financial market volatility. Motivated in part by recent results on optimal filtering, a new conditional variance model for better characterizing stock return volatility is also presented.

1. Introduction

Until a decade ago the focus of most macroeconometric and financial time series modeling centered on the conditional first moments, with any temporal dependencies in the higher order moments treated as a nuisance. The increased importance played by risk and uncertainty considerations in modern economic theory, however, has necessitated the development of new econometric time series techniques that allow for the modeling of time varying variances and covariances. Given the apparent lack of any structural dynamic economic theory explaining the variation in higher order moments, particularly instrumental in this development has been the autoregressive conditional heteroskedastic (ARCH) class of models introduced by Engle (1982). Parallel to the success of standard linear time series models, arising from the use of the conditional versus the unconditional mean, the key insight offered by the ARCH model lies in the distinction between the *conditional* and the *unconditional* second order moments. While the unconditional covariance matrix for the variables of interest may be time invariant, the conditional variances and covariances often depend non-trivially on the past states of the world. Understanding the exact nature of this temporal dependence is crucially important for many issues in macroeconomics and finance, such as irreversible investments, option pricing, the term structure of interest rates, and general dynamic asset pricing relationships. Also, from the perspective of econometric inference, the loss in asymptotic efficiency from neglected heteroskedasticity may be arbitrarily large and, when evaluating economic forecasts, a much more accurate estimate of the forecast error uncertainty is generally available by conditioning on the current information set.

1.1. *Definitions*

Let $\{\varepsilon_t(\theta)\}$ denote a discrete time stochastic process with conditional mean and variance functions parametrized by the finite dimensional vector $\theta \in \Theta \subseteq R^m$, where

θ_0 denotes the true value. For notational simplicity we shall initially assume that $\varepsilon_t(\theta)$ is a scalar, with the obvious extensions to a multivariate framework treated in Section 6. Also, let $E_{t-1}(\cdot)$ denote the mathematical expectation, conditional on the past, of the process, along with any other information available at time $t-1$.

The $\{\varepsilon_t(\theta_0)\}$ process is then defined to follow an ARCH model if the conditional mean equals zero,

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

but the conditional variance,

$$\sigma_t^2(\theta_0) \equiv \mathrm{Var}_{t-1}(\varepsilon_t(\theta_0)) = E_{t-1}(\varepsilon_t^2(\theta_0)) \qquad t = 1, 2, \ldots, \tag{1.2}$$

depends non-trivially on the sigma-field generated by the past observations; i.e. $\{\varepsilon_{t-1}(\theta_0), \varepsilon_{t-2}(\theta_0), \ldots\}$. When obvious from the context, the explicit dependence on the parameters, θ, will be suppressed for notational convenience. Also, in the multivariate case the corresponding time varying conditional covariance matrix will be denoted by $\boldsymbol{\Omega}_t$.

In much of the subsequent discussion we shall focus directly on the $\{\varepsilon_t\}$ process, but the same ideas extend directly to the situation in which $\{\varepsilon_t\}$ corresponds to the innovations from some more elaborate econometric model. In particular, let $\{y_t(\theta_0)\}$ denote the stochastic process of interest with conditional mean

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

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

$$\varepsilon_t(\theta_0) \equiv y_t - \mu_t(\theta_0) \qquad t = 1, 2, \ldots. \tag{1.4}$$

The conditional variance for $\{\varepsilon_t\}$ then equals the conditional variance for the $\{y_t\}$ process. Since very few economic and financial time series have a constant conditional mean of zero, most of the empirical applications of the ARCH methodology actually fall within this framework.

Returning to the definitions in equations (1.1) and (1.2), it follows that the standardized process,

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

will have conditional mean zero, and a time invariant conditional variance of unity. This observation forms the basis for most of the inference procedures that underlie the applications of ARCH type models.

If the conditional distribution for z_t is furthermore assumed to be time invariant with a finite fourth moment, it follows by Jensen's inequality that

$$E(\varepsilon_t^4) = E(z_t^4)E(\sigma_t^4) \geqslant E(z_t^4)E(\sigma_t^2)^2 = E(z_t^4)E(\varepsilon_t^2)^2,$$

where the equality holds true for a constant conditional variance only. Given a normal distribution for the standardized innovations in equation (1.5), the unconditional distribution for ε_t is therefore leptokurtic.

The setup in equations (1.1) through (1.4) is extremely general and does not lend itself directly to empirical implementation without first imposing further restrictions on the temporal dependencies in the conditional mean and variance functions. Below we shall discuss some of the most practical and popular such ARCH formulations for the conditional variance. While the first empirical applications of the ARCH class of models were concerned with modeling inflationary uncertainty, the methodology has subsequently found especially wide use in capturing the temporal dependencies in asset returns. For a recent survey of this extensive empirical literature we refer to Bollerslev et al. (1992).

1.2. *Empirical regularities of asset returns*

Even in the univariate case, the array of functional forms permitted by equation (1.2) is vast, and infinitely larger than can be accommodated by any parametric family of ARCH models. Clearly, to have any hope of selecting an appropriate ARCH model, we must have a good idea of what empirical regularities the model should capture. Thus, a brief discussion of some of the important regularities for asset returns volatility follows.

1.2.1. *Thick tails*

Asset returns tend to be leptokurtic. The documentation of this empirical regularity by Mandelbrot (1963), Fama (1965) and others led to a large literature on modeling stock returns as i.i.d. draws from thick-tailed distributions; see, e.g. Mandelbrot (1963), Fama (1963, 1965), Clark (1973) and Blattberg and Gonedes (1974).

1.2.2. *Volatility clustering*

As Mandelbrot (1963) wrote,

> ...large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes....

This volatility clustering phenomenon is immediately apparent when asset returns are plotted through time. To illustrate, Figure 1 plots the daily capital gains on the Standard 90 composite stock index from 1928–1952 combined with Standard and

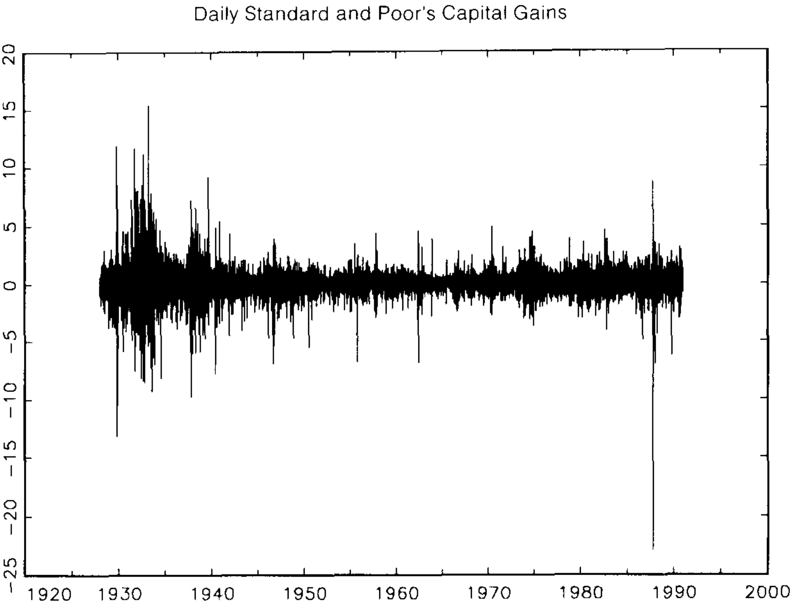


Figure 1

Poor's 500 index from 1953–1990. The returns are expressed in percent, and are continuously compounded. It is clear from visual inspection of the figure, and any reasonable statistical test, that the returns are not i.i.d. through time. For example, volatility was clearly higher during the 1930's than during the 1960's, as confirmed by the estimation results reported in French et al. (1987).

A similar message is contained in Figure 2, which plots the daily percentage Deutschmark/U.S. Dollar exchange rate appreciation. Distinct periods of exchange market turbulence and tranquility are immediately evident. We shall return to a formal analysis of both of these two time series in Section 9 below.

Volatility clustering and thick tailed returns are intimately related. As noted in Section 1.1 above, if the unconditional kurtosis of ε_t is finite, $E(\varepsilon_t^4)/[E(\varepsilon_t^2)]^2 \geqslant E(z_t^4)$, where the last inequality is strict unless σ_t is constant. Excess kurtosis in ε_t can therefore arise from randomness in σ_t, from excess kurtosis in the conditional distribution of ε_t, i.e., in z_t, or from both.

1.2.3. Leverage effects

The so-called "leverage effect," first noted by Black (1976), refers to the tendency for changes in stock prices to be negatively correlated with changes in stock volatility. Fixed costs such as financial and operating leverage provide a partial explanation for this phenomenon. A firm with debt and equity outstanding typically

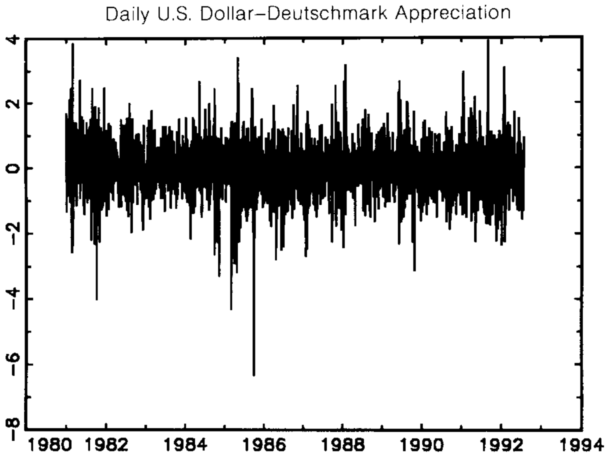


Figure 2

becomes more highly leveraged when the value of the firm falls. This raises equity returns volatility if the returns on the firm as a whole are constant. Black (1976), however, argued that the response of stock volatility to the direction of returns is too large to be explained by leverage alone. This conclusion is also supported by the empirical work of Christie (1982) and Schwert (1989b).

1.2.4. *Non-trading periods*

Information that accumulates when financial markets are closed is reflected in prices after the markets reopen. If, for example, information accumulates at a constant rate over calendar time, then the variance of returns over the period from the Friday close to the Monday close should be three times the variance from the Monday close to the Tuesday close. Fama (1965) and French and Roll (1986) have found, however, that information accumulates more slowly when the markets are closed than when they are open. Variances are higher following weekends and holidays than on other days, but not nearly by as much as would be expected if the news arrival rate were constant. For instance, using data on daily returns across all NYSE and AMEX stocks from 1963–1982, French and Roll (1986) find that volatility is 70 times higher per hour on average when the market is open than when it is closed. Baillie and Bollerslev (1989) report qualitatively similar results for foreign exchange rates.

1.2.5. *Forecastable events*

Not surprisingly, forecastable releases of important information are associated with high ex ante volatility. For example, Cornell (1978), and Patell and Wolfson (1979,

1981) show that individual firms' stock returns volatility is high around earnings announcements. Similarly, Harvey and Huang (1991, 1992) find that fixed income and foreign exchange volatility is higher during periods of heavy trading by central banks or when macroeconomic news is being released.

There are also important predictable changes in volatility across the trading day. For example, volatility is typically much higher at the open and close of stock and foreign exchange trading than during the middle of the day. This pattern has been documented by Harris (1986), Gerity and Mulherin (1992) and Baillie and Bollerslev (1991), among others. The increase in volatility at the open at least partly reflects information accumulated while the market was closed. The volatility surge at the close is less easily interpreted.

1.2.6. *Volatility and serial correlation*

LeBaron (1992) finds a strong inverse relation between volatility and serial correlation for U.S. stock indices. This finding appears remarkably robust to the choice of sample period, market index, measurement interval and volatility measure. Kim (1989) documents a similar relationship in foreign exchange rate data.

1.2.7. *Co-movements in volatilities*

Black (1976) observed that

> ...there is a lot of commonality in volatility changes across stocks: a 1% market volatility change typically implies a 1% volatility change for each stock. Well, perhaps the high volatility stocks are somewhat more sensitive to market volatility changes than the low volatility stocks. In general it seems fair to say that when stock volatilities change, they all tend to change in the same direction.

Diebold and Nerlove (1989) and Harvey et al. (1992) also argue for the existence of a few common factors explaining exchange rate volatility movements. Engle et al. (1990b) show that U.S. bond volatility changes are closely linked across maturities. This commonality of volatility changes holds not only across assets within a market, but also *across* different markets. For example, Schwert (1989a) finds that U.S. stock and bond volatilities move together, while Engle and Susmel (1993) and Hamao et al. (1990) discover close links between volatility changes across international stock markets. The importance of international linkages has been further explored by King et al. (1994), Engle et al. (1990a), and Lin et al. (1994).

That volatilities move together should be encouraging to model builders, since it indicates that a few common factors may explain much of the temporal variation in the conditional variances and covariances of asset returns. This forms the basis for the factor ARCH models discussed in Section 6.2 below.

1.2.8. *Macroeconomic variables and volatility*

Since stock values are closely tied to the health of the economy, it is natural to expect that measures of macroeconomic uncertainty such as the conditional variances of industrial production, interest rates, money growth, etc. should help explain changes in stock market volatility. Schwert (1989a, b) finds that although stock volatility rises sharply during recessions and financial crises and drops during expansions, the relation between macroeconomic uncertainty and stock volatility is surprisingly weak. Glosten et al. (1993), on the other hand, uncover a strong positive relationship between stock return volatility and interest rates.

1.3. *Univariate parametric models*

1.3.1. *GARCH*

Numerous parametric specifications for the time varying conditional variance have been proposed in the literature. In the linear ARCH(q) model originally introduced by Engle (1982), the conditional variance is postulated to be a linear function of the past q squared innovations,

$$\sigma_t^2 = \omega + \sum_{i=1,q} \alpha_i \varepsilon_{t-i}^2 \equiv \omega + \alpha(L)\varepsilon_{t-1}^2, \tag{1.6}$$

where L denotes the lag or backshift operator, $L^i y_t \equiv y_{t-i}$. Of course, for this model to be well defined and the conditional variance to be positive, almost surely the parameters must satisfy $\omega > 0$ and $\alpha_1 \geqslant 0, \ldots, \alpha_q \geqslant 0$.

Defining $v_t \equiv \varepsilon_t^2 - \sigma_t^2$, the ARCH($q$) model in (1.6) may be re-written as

$$\varepsilon_t^2 = \omega + \alpha(L)\varepsilon_{t-1}^2 + v_t. \tag{1.7}$$

Since $E_{t-1}(v_t) = 0$, the model corresponds directly to an AR(q) model for the squared innovations, ε_t^2. The process is covariance stationary if and only if the sum of the positive autoregressive parameters is less than one, in which case the unconditional variance equals $\mathrm{Var}(\varepsilon_t) \equiv \sigma^2 = \omega/(1 - \alpha_1 - \cdots - \alpha_q)$.

Even though the ε_t's are serially uncorrelated, they are clearly not independent through time. In accordance with the stylized facts for asset returns discussed above, there is a tendency for large (small) absolute values of the process to be followed by other large (small) values of unpredictable sign. Also, as noted above, if the distribution for the standardized innovations in equation (1.5) is assumed to be time invariant, the unconditional distribution for ε_t will have fatter tails than the distribution for z_t. For instance, for the ARCH(1) model with conditionally normally distributed errors, $E(\varepsilon_t^4)/E(\varepsilon_t^2)^2 = 3(1-\alpha_1^2)/(1-3\alpha_1^2)$ if $3\alpha_1^2 < 1$, and $E(\varepsilon_t^4)/E(\varepsilon_t^2)^2 = \infty$ otherwise; both of which exceed the normal value of three.

Alternatively the ARCH(q) model may also be represented as a time varying parameter MA(q) model for ε_t,

$$\varepsilon_t = \omega + \alpha(L)\zeta_{t-1}\varepsilon_{t-1}, \tag{1.8}$$

where $\{\zeta_t\}$ denotes a scalar i.i.d. stochastic process with mean zero and variance one. Time varying parameter models have a long history in econometrics and statistics. The appeal of the observational equivalent formulation in equation (1.6) stems from the explicit focus on the time varying conditional variance of the process. For discussion of this interpretation of ARCH models, see, e.g., Tsay (1987), Bera et al. (1993) and Bera and Lee (1993).

In empirical applications of ARCH(q) models a long lag length and a large number of parameters are often called for. To circumvent this problem Bollerslev (1986) proposed the generalized ARCH, or GARCH(p, q), model,

$$\sigma_t^2 = \omega + \sum_{i=1,q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1,p} \beta_j \sigma_{t-j}^2 \equiv \omega + \alpha(L)\varepsilon_{t-1}^2 + \beta(L)\sigma_{t-1}^2. \tag{1.9}$$

For the conditional variance in the GARCH(p, q) model to be well defined all the coefficients in the corresponding infinite order linear ARCH model must be positive. Provided that $\alpha(L)$ and $\beta(L)$ have no common roots and that the roots of the polynomial $\beta(x) = 1$ lie outside the unit circle, this positivity constraint is satisfied if and only if all the coefficients in the infinite power series expansion for $\alpha(x)/(1 - \beta(x))$ are non-negative. Necessary and sufficient conditions for this are given in Nelson and Cao (1992). For the simple GARCH(1, 1) model almost sure positivity of σ_t^2 requires that $\omega \geqslant 0, \alpha_1 \geqslant 0$ and $\beta_1 \geqslant 0$.

Rearranging the GARCH(p, q) model as in equation (1.7), it follows that

$$\varepsilon_t^2 = \omega + [\alpha(L) + \beta(L)]\varepsilon_{t-1}^2 - \beta(L)v_{t-1} + v_t, \tag{1.10}$$

which defines an ARMA[max(p, q), p] model for ε_t^2. By standard arguments, the model is covariance stationary if and only if all the roots of $\alpha(x) + \beta(x) = 1$ lie outside the unit circle; see Bollerslev (1986) for a formal proof. In many applications with high frequency financial data the estimate for $\alpha(1) + \beta(1)$ turns out to be very close to unity. This provides an empirical motivation for the so-called integrated GARCH (p, q), or IGARCH(p, q), model introduced by Engle and Bollerslev (1986). In the IGARCH class of models the autoregressive polynomial in equation (1.10) has a unit root, and consequently a shock to the conditional variance is persistent in the sense that it remains important for future forecasts of all horizons. Further discussion of stationarity conditions and issues of persistence are contained in Section 3 below.

Just as an ARMA model often leads to a more parsimonious representation of the temporal dependencies in the conditional mean than an AR model, the GARCH (p, q) formulation in equation (1.9) provides a similar added flexibility over the linear

ARCH(q) model when parametrizing the conditional variance. This analogy to the ARMA class of models also allows for the use of standard time series techniques in the identification of the orders p and q as discussed in Bollerslev (1988). Because of the higher order dependencies in the v_t process, standard Box and Jenkins (1976) inference procedures will generally be very inefficient, however. Also, as noted above, in most empirical applications with finely sampled data, the simple GARCH (1, 1) model with $\hat{\alpha}_1 + \hat{\beta}_1$ close to one is found to provide a good description of the data. Possible explanations for this phenomenon are discussed in Sections 4 and 5 below.

1.3.2. *EGARCH*

GARCH successfully captures thick tailed returns, and volatility clustering, and can readily be modified to allow for several other stylized facts, such as non-trading periods and predictable information releases. It is not well suited to capture the "leverage effect," however, since the conditional variance in equation (1.9) is a function only of the magnitudes of the lagged residuals and not their signs.

In the exponential GARCH (EGARCH) model of Nelson (1991), σ_t^2 depends on both the size and the sign of lagged residuals. In particular,

$$\ln(\sigma_t^2) = \omega + \left(1 + \sum_{i=1,q} \alpha_i L^i\right)\left(1 - \sum_{j=1,p} \beta_j L^j\right)^{-1} \{\theta z_{t-1} + \gamma[|z_{t-1}| - E|z_{t-1}|]\}. \tag{1.11}$$

Thus, $\{\ln(\sigma_t^2)\}$ follows an ARMA(p, q) process, with the usual ARMA stationarity conditions. Formulas for the higher order moments of ε_t are given in Nelson (1991). As in the GARCH case, ω can easily be made a function of time to accommodate the effect of any non-trading periods or forecastable events.

1.3.3. *Other univariate parametrizations*

Though our list of stylized facts regarding asset volatility narrows the field of candidate ARCH models somewhat, the number of possible formulations is still vast. For example, to capture volatility clustering, GARCH assumes that the conditional variance σ_t^2 equals a distributed lag of squared residuals. An equally natural assumption, employed by Taylor (1986) and Schwert (1989a, b), is that the conditional standard deviation σ_t is a distributed lag of absolute residuals, as in

$$\sigma_t = \omega + \sum_{i=1,q} \alpha_i |\varepsilon_{t-i}| + \sum_{j=1,p} \beta_j \sigma_{t-j}. \tag{1.12}$$

Higgins and Bera (1992) nest the GARCH model and (1.12) in the class of non-linear ARCH (NARCH) models:

$$\sigma_t^\gamma = \omega + \sum_{i=1,q} \alpha_i |\varepsilon_{t-i}|^\gamma + \sum_{j=1,p} \beta_j \sigma_{t-j}^\gamma. \tag{1.13}$$

If (1.13) is modified further by setting

$$\sigma_t^\gamma = \omega + \sum_{i=1,q} \alpha_i |\varepsilon_{t-i} - \kappa|^\gamma + \sum_{j=1,p} \beta_j \sigma_{t-j}^\gamma, \tag{1.14}$$

for some non-zero κ, the innovations in σ_t^γ will depend on the size as well as the sign of lagged residuals, thereby allowing for the leverage effect in stock return volatility. The formulation in equation (1.14) with $\gamma = 2$ is also a special case of Sentana's (1991) quadratic ARCH (QARCH) model, in which σ_t^2 is modeled as a quadratic form in the lagged residuals. A simple version of this model termed asymmetric ARCH, or AARCH, was also proposed by Engle (1990). In the first order case the AARCH model becomes

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \delta\varepsilon_{t-1} + \beta\sigma_{t-1}^2, \tag{1.15}$$

where a negative value of δ means that positive returns increase volatility less than negative returns.

Another route for introducing asymmetric effects is to set

$$\sigma_t^\gamma = \omega + \sum_{i=1,q} [\alpha_i^+ \mathrm{I}(\varepsilon_{t-i} > 0)|\varepsilon_{t-i}|^\gamma + \alpha_i^- \mathrm{I}(\varepsilon_{t-i} \leqslant 0)|\varepsilon_{t-i}|^\gamma] + \sum_{j=1,p} \beta_j \sigma_{t-j}^\gamma, \tag{1.16}$$

where $\mathrm{I}(\cdot)$ denotes the indicator function. For example the threshold ARCH (TARCH) model of Zakoian (1990) corresponds to equation (1.16) with $\gamma = 1$. Glosten, Jagannathan and Runkle (1993) estimate a version of equation (1.16) with $\gamma = 2$. This so-called GJR model allows a quadratic response of volatility to news with different coefficients for good and bad news, but maintains the assertion that the minimum volatility will result when there is no news.[1]

Two additional classes of models have recently been proposed. These models have a somewhat different intellectual heritage but imply particular forms of conditional heteroskedasticity. The first is the unobserved components structural ARCH (STARCH) model of Harvey et al. (1992). These are state space models or factor models in which the innovation is composed of several sources of error where each of the error sources has heteroskedastic specifications of the ARCH form. Since the error components cannot be separately observed given the past observations, the independent variables in the variance equations are not measurable with respect

[1] In a comparison study for daily Japanese TOPIX data, Engle and Ng (1993) found that the EGARCH and the GJR formulation were superior to the AARCH model (1.15) which simply shifted the intercept.

to the available information set, which complicates inference procedures.[2] Following earlier work by Diebold and Nerlove (1989), Harvey et al. (1992) propose an estimation strategy based on the Kalman filter.

To illustrate the issues, consider the factor structure

$$y_t = Bf_t + \varepsilon_t, \tag{1.17}$$

where y_t is an $n \times 1$ vector of asset returns, f_t is a scalar factor with time invariant factor loadings, B, and ε_t is an $n \times 1$ vector of idiosyncratic returns. If the factor follows an ARCH(1) process,

$$\sigma^2_{f,t} = \omega + \alpha f^2_{t-1}, \tag{1.18}$$

then new estimation problems arise since f_{t-1} is not observed, and $\sigma^2_{f,t}$ is not a conditional variance. The Kalman filter gives both $E_{t-1}(f_{t-1})$ and $V_{t-1}(f_{t-1})$, so the proposal by Harvey et al. (1992) is to let the conditional variance of the factor, which is the state variable in the Kalman filter, be given by

$$E_{t-1}(\sigma^2_{f,t}) = \omega + \alpha[V_{t-1}(f_{t-1}) + \{E_{t-1}(f_{t-1})\}^2].$$

Another important class of models is the switching ARCH, or SWARCH, model proposed independently by Cai (1994) and Hamilton and Susmel (1992). This class of models postulates that there are several different ARCH models and that the economy switches from one to another following a Markov chain. In this model there can be an extremely high volatility process which is responsible for events such as the stock market crash in October 1987. Since this could happen at any time but with very low probability, the behavior of risk averse agents will take this into account. The SWARCH model must again be estimated using Kalman filter techniques.

The richness of the family of parametric ARCH models is both a blessing and a curse. It certainly complicates the search for the "true" model, and leaves quite a bit of arbitrariness in the model selection stage. On the other hand, the flexibility of the ARCH class of models means that in the analysis of structural economic models with time varying volatility, there is a good chance that an appropriate parametric ARCH model can be formulated that will make the analysis tractable. For example, Campbell and Hentschell (1992) seek to explain the drop in stock prices associated with an increase in volatility within the context of an economic model. In their model, exogenous rises in stock volatility increase discount rates, lowering stock prices. Using an EGARCH model would have made their formal analysis intractable, but based on a QARCH formulation the derivations are straightforward.

[2] These models sometimes are also called stochastic volatility models; see Andersen (1992a) for a more formal definition.

1.4. ARCH in mean models

Many theories in finance call for an explicit tradeoff between the expected returns and the variance, or the covariance among the returns. For instance, in Merton's (1973) intertemporal CAPM model, the expected excess return on the market portfolio is linear in its conditional variance under the assumption of a representative agent with log utility. In more general settings, the conditional covariance with an appropriately defined benchmark portfolio often serves to price the assets. For example, according to the traditional capital asset pricing model (CAPM) the excess returns on all risky assets are proportional to the non-diversifiable risk as measured by the covariances with the market portfolio. Of course, this implies that the expected excess return on the market portfolio is simply proportional to its own conditional variance as in the univariate Merton (1973) model.

The ARCH in mean, or ARCH–M, model introduced by Engle et al. (1987) was designed to capture such relationships. In the ARCH–M model the conditional mean is an explicit function of the conditional variance,

$$\mu_t(\theta) = g[\sigma_t^2(\theta), \theta], \tag{1.19}$$

where the derivative of the $g(\cdot,\cdot)$ function with respect to the first element is non-zero. The multivariate extension of the ARCH–M model, allowing for the explicit influence of conditional covariance terms in the conditional mean equations, was first considered by Bollerslev et al. (1988) in the context of a multivariate CAPM model. The exact formulation of such multivariate ARCH models is discussed further in Section 6 below.

The most commonly employed univariate specifications of the ARCH–M model postulate a linear relationship in σ_t or σ_t^2; e.g. $g[\sigma_t^2(\theta), \theta] = \mu + \delta\sigma_t^2$. For $\delta \neq 0$ the risk premium will be time-varying, and could change sign if $\mu < 0 < \delta$. Note that any time variation in σ_t will result in serial correlation in the $\{y_t\}$ process.[3]

Because of the explicit dependence of the conditional mean on the conditional variance and/or covariance, several unique problems arise in the estimation and testing of ARCH–M models. We shall return to a discussion of these issues in Section 2.2 below.

1.5. Nonparametric and semiparametric methods

A natural response to the overwhelming variety of parametric univariate ARCH models, is to consider and estimate nonparametric models. One of the first attempts at this problem was by Pagan and Schwert (1990) who used a collection of standard

[3] The exact form of this serial dependence has been formally analyzed for some simple models in Hong (1991).

nonparametric estimation methods, including kernels, Fourier series and least squares regressions, to fit models for the relation between y_t^2 and past y_t's, and then compare the fits with several parametric formulations. Effectively, these models estimate the function $f(\cdot)$ in

$$y_t^2 = f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, \theta) + \eta_t. \tag{1.20}$$

Several problems immediately arise in estimating $f(\cdot)$, however. Because of the problems of high dimensionality, the parameter p must generally be chosen rather small, so that only a little temporal smoothing can actually be achieved directly from (1.20). Secondly, if only squares of the past y_t's are used the asymmetric terms may not be discovered. Thirdly, minimizing the distance between y_t^2 and $f_t \equiv f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, \theta)$ is most effective if η_t is homoskedastic, however, in this case it is highly heteroskedastic. In fact, if f_t were the precise conditional heteroskedasticity, then $y_t^2 f_t^{-1}$ and $\eta_t f_t^{-1}$, would be homoskedastic. Thus, η_t has conditional variance f_t^2, so that the heteroskedasticity is actually more severe than in y_t. Not only does parameter estimation become inefficient, but the use of a simple R^2 measure as a model selection criterion is inappropriate. An R^2 criterion penalizes generalized least squares or maximum likelihood estimators, and corresponds to a loss function which does not even penalize zero or negative predicted variances. This issue will be discussed in more detail in Section 7. Indeed, the conclusion from the empirical analysis for U.S. stock returns conducted in Pagan and Schwert (1990) was that there was in-sample evidence that the nonparametric models could outperform the GARCH and EGARCH models, but that out-of-sample the performance deteriorated. When a proportional loss function was used the superiority of the nonparametric models also disappeared in-sample.

Any nonparametric estimation method must be sensitive to the above mentioned issues. Gourieroux and Monfort (1992) introduce a qualitative threshold ARCH, or QTARCH, model, which has conditional variance that is constant over various multivariate observation intervals. For example, divide the space of y_t into J intervals and let $\mathrm{I}_j(y_t)$ be 1 if y_t is in the jth interval. The QTARCH model is then written as

$$y_t = \sum_{i=1,p\; j=1,J} \alpha_{ij} \mathrm{I}_j(y_{t-i}) + \sum_{i=1,p\; j=1,J} \beta_{ij} \mathrm{I}_j(y_{t-i}) u_{t-i}, \tag{1.21}$$

where u_t is taken to be i.i.d. The α_{ij} parameters govern the mean and the β_{ij} parameters govern the variance of the $\{y_t\}$ process. As the sample size grows, J can be increased and the bins made smaller to approximate any process.

In their most successful application, Gourieroux and Monfort (1992) add a GARCH term resulting in the G–QTARCH(1) model, with a conditional variance given by

$$\sigma_t^2 = \omega + \beta_0 \sigma_{t-1}^2 + \sum_{j=1,J} \beta_j \mathrm{I}_j(y_{t-1}). \tag{1.22}$$

Interestingly, the estimates using four years of daily returns on the French stock index (CAC) showed strong evidence of the leverage effect.

In the same spirit, Engle and Ng (1993) propose and estimate a partially nonparametric, or PNP, model, which uses linear splines to estimate the shape of the response to the most recent news. The name of the model reflects the fact that the long memory component is treated as parametric while the relationship between the news and the volatility is treated nonparametrically.

The semi-nonparametric series expansion developed in a sequence of papers by Gallant and Tauchen (1989) and Gallant et al. (1991, 1992, 1993) has also been employed in characterizing the temporal dependencies in the second order moments of asset returns. A formal description of this innovative nonparametric procedure is beyond the scope of the present chapter, however.

2. Inference procedures

2.1. *Testing for ARCH*

2.1.1. *Serial correlation and Lagrange multiplier tests*

The original Lagrange multiplier (LM) test for ARCH proposed by Engle (1982) is very simple to compute, and relatively easy to derive. Under the null hypothesis it is assumed that the model is a standard dynamic regression model which can be written as

$$y_t = x_t\beta + \varepsilon_t, \tag{2.1}$$

where x_t is a set of weakly exogenous and lagged dependent variables and ε_t is a Gaussian white noise process,

$$\varepsilon_t | \mathrm{I}_{t-1} \sim N(0, \sigma^2), \tag{2.2}$$

where I_t denotes the available information set. Because the null is so easily estimated, the Lagrange multiplier test is a natural choice. The alternative hypothesis is that the errors are ARCH(q), as in equation (1.6). A straightforward derivation of the Lagrange multiplier test as in Engle (1984) leads to the TR^2 test statistic, where the R^2 is computed from the regression of ε_t^2 on a constant and $\varepsilon_{t-1}^2, \ldots, \varepsilon_{t-q}^2$. Under the null hypothesis that there is no ARCH, the test statistic is asymptotically distributed as a chi-square distribution with q degrees of freedom.

The intuition behind this test is very clear. If the data are homoskedastic, then the variance cannot be predicted and variations in ε_t^2 will be purely random. However, if ARCH effects are present, large values of ε_t^2 will be predicted by large values of the past squared residuals.

While this is a simple and widely used statistic, there are several points which should be made. First and most obvious, if the model in (2.1) is misspecified by omission of a relevant regressor or failure to account for some non-linearity or serial correlation, it is quite likely that the ARCH test will reject as these errors may induce serial correlation in the squared errors. Thus, one cannot simply assume that ARCH effects are necessarily present when the ARCH test rejects. Second, there are several other asymptotically equivalent forms of the test, including the standard F-test from the above regression. Another version of the test simply omits the constant but subtracts the estimate of the unconditional variance, σ^2, from the dependent variable, and then uses one half the explained sum of squares as a test statistic. It is also quite common to use asymptotically equivalent portmanteau tests, such as the Ljung and Box (1978) statistic, for ε_t^2.

As described above, the parameters of the ARCH(q) model must be positive. Hence, the ARCH test could be formulated as a one tailed test. When $q = 1$ this is simple to do, but for higher values of q, the procedures are not as clear. Demos and Sentana (1991) has suggested a one sided ARCH test which is presumably more powerful than the simple TR^2 test described above. Similarly, since we find that the GARCH(1, 1) is often a superior model and is surely more parsimoniously parametrized, one would like a test which is more powerful for this alternative. The Lagrange multiplier principle unfortunately does not deliver such a test because, for models close to the null, α_1 and β_1 cannot be separately identified. In fact, the LM test for GARCH(1, 1) is just the same as the LM test for ARCH(1); see Lee and King (1993) which proposes a locally most powerful test for ARCH and GARCH.

Of course, Wald type tests for GARCH may also be computed. These too are non-standard, however. The t-statistic on α_1 in the GARCH(1, 1) model will not have a t-distribution under the null hypothesis since there is no time-varying input and β_1 will be unidentified. Finally, likelihood ratio test statistics may be examined, although again they have an uncertain distribution under the null. Practical experience, however, suggests that the latter is a very powerful approach to testing for GARCH effects. We shall return to a more detailed discussion of these tests in Section 2.2.2 below.

2.1.2. BDS test for ARCH

The tests for ARCH discussed above are tests for volatility clustering rather than general conditional heteroskedasticity, or general non-linear dependence. One widely used test for general departures from i.i.d. observations is the BDS test introduced by Brock, Dechert and Scheinkman (1987). We will consider only the univariate version of the test; the multivariate extension is made in Baek and Brock (1992). The BDS test has inspired quite a large literature and several applications have appeared in the finance area; see, e.g. Scheinkman and LeBaron (1989), Hsieh (1991) and Brock et al. (1991).

To set up the test, let $\{x_t\}_{t=1,T}$ denote a scalar sequence which under the null hypothesis is assumed to be i.i.d. through time. Define the *m-histories* of the x_t process as the vectors $(x_1,\ldots,x_m)$, $(x_2,\ldots,x_{m+1})$, $(x_3,\ldots,x_{m+2}),\ldots,(x_{T-m},\ldots,X_{T-1})$, $(x_{T-m+1},\ldots,x_T)$. Clearly, there are $T-m+1$ such m-histories, and therefore $(T-m+1)(T-m)/2$ distinct pairs of m-histories. Next, define the *correlation integral* as the fraction of the distinct pairs of m-histories lying within a distance c in the sup norm; i.e.

$$C_{m,T}(c) \equiv [(T-m+1)(T-m)/2]^{-1} \sum_{t=m,s} \sum_{s=m,T} \mathrm{I}(\max_{j=0,m-1}|x_{t-j}-x_{s-j}|<c). \tag{2.3}$$

Under weak dependence conditions, $C_{m,T}(c)$ converges almost surely to a limit $C_m(c)$. By the basic properties of order-statistics, $C_m(c)=C_1(c)^m$ when $\{x_t\}$ is i.i.d. The BDS test is based on the difference, $[C_{m,T}(c)-C_{1,T}(c)^m]$. Intuitively, $C_{m,T}(c)>C_{1,T}(c)^m$ means that when x_{t-j} and x_{s-j} are "close" for $j=1$ to $m-1$, i.e. $\max_{j=1,m-1}|x_{t-j}-x_{s-j}|<c$, then x_t and x_s are more likely than average to be close, also. In other words, nearest-neighbor methods work in predicting the $\{x_t\}$ series, which is inconsistent with the i.i.d. assumption.[4]

Brock et al. (1987) show that for fixed m and c, $T^{1/2}[C_{m,T}(c)-C_{1,T}(c)^m]$ is asymptotically normal with mean zero and variance $V(m,c)$ given by

$$V(m,c) \equiv 4[K(c)^m + 2\sum_{j=1,m-1} K(c)^{m-j}C_1(c)^{2j} + (m-1)^2C_1(c)^{2m} - m^2K(c)C_1(c)^{2m-2}], \tag{2.4}$$

where $K(c)=E\{[F(x_t+c)-F(x_t-c)]^2\}$, and $F(\cdot)$ is the cumulative distribution function of x_t. The BDS test is then computed as

$$T^{1/2}[C_{m,T}(c)-C_{1,T}(c)^m]/\hat{V}(T,m,c), \tag{2.5}$$

where $\hat{V}(T,m,c)$ denotes a consistent estimator of $V(m,c)$, details of which are given by Brock et al. (1987, 1991). For fixed $m \geqslant 2$ and $c>0$, the BDS statistic in equation (2.5) is asymptotically standard normal.

The BDS test has power against many, though not all, departures from i.i.d. In particular, as documented by Brock et al. (1991) and Hsieh (1991), the power against ARCH alternatives is close to Engle's (1982) test. For other conditionally heteroskedastic alternatives, the power of the BDS test may be superior. To illustrate, consider the following example from Brock et al. (1991), where σ_t^2 is deterministically

[4] $C_{m,T}(c)<C_{1,T}(c)^m$ indicates the reverse of nearest-neighbors predictability. It is important not to push the nearest-neighbors analogy too far, however. For example, suppose $\{x_t\}$ is an ARCH process with a constant conditional mean of 0. In this case, the conditional mean of x_t is always 0, and the nearest-neighbors analogy breaks down for minimum mean-squared-error forecasting of x_t. It still holds for forecasting, say, the probability that x_t lies in some interval.

determined by the tent map,

$$\sigma_{t+1}^2 = 1 - 2|\sigma_t^2 - 0.5|, \tag{2.6}$$

with $\sigma_0^2 \in (0,1)$. The model is clearly heteroskedastic, but does not exhibit volatility clustering, since the empirical serial correlations of $\{\sigma_t^2\}$ approach zero in large samples for almost all values of σ_0^2.

In order to actually implement the BDS test a choice has to be made regarding the values of m and c. The Monte Carlo experiments of Brock et al. (1991) suggest that c should be between $\frac{1}{2}$ and 2 standard deviations of the data, and that T/m should be greater than 200 with m no greater than 5. For the asymptotic distribution to be a good approximation to the finite-sample behavior of the BDS test a sample size of at least 500 observations is required.

Since the BDS test is a test for i.i.d., it requires some adaptation in testing for ARCH errors in the presence of time-varying conditional means. One of the most convenient properties of the BDS test is that unlike many other diagnostic tests, including the portmanteau statistic, its distribution is unchanged when applied to residuals from a linear model. If, for example, the null hypothesis is a stationary, invertible, ARMA model with i.i.d. errors and the alternative hypothesis is the same ARMA model but with ARCH errors, the standard BDS test remains valid when applied to the fitted residuals from the homoskedastic ARMA model. A similar invariance property holds for residuals from a wide variety of non-linear regression models, but as discussed in Section 2.4.2 below, this does not carry over to the standardized residuals from a fitted ARCH model. Of course, the BDS test may reject due to misspecification of the conditional mean rather than ARCH effects in the errors. The same is true, however, of the simple TR^2 Lagrange multiplier test for ARCH, which has power against a wide variety of non-linear alternatives.

2.2. *Maximum likelihood methods*

2.2.1. *Estimation*

The procedure most often used in estimating θ_0 in ARCH models involves the maximization of a likelihood function constructed under the auxiliary assumption of an i.i.d. distribution for the standardized innovations in equation (1.5). In particular, let $f(z_t;\eta)$ denote the density function for $z_t(\theta) \equiv \varepsilon_t(\theta)/\sigma_t(\theta)$ with mean zero, variance one, and nuisance parameters $\eta \in H \subseteq R^k$. Also, let $\{y_T, y_{T-1}, \ldots, y_1\}$ refer to the sample realizations from an ARCH model as defined by equations (1.1) through (1.4), and $\psi' \equiv (\theta', \eta')$ the combined $(m+k)\times 1$ parameter vector to be estimated for the conditional mean, variance and density functions.

The log likelihood function for the tth observation is then given by

$$l_t(y_t;\psi) = \ln\{f[z_t(\theta);\eta]\} - 0.5\ln[\sigma_t^2(\theta)] \qquad t = 1,2,\ldots. \tag{2.7}$$

The second term on the right hand side is a Jacobian that arises in the transformation from the standardized innovations, $z_t(\theta)$, to the observables, $y_t(\theta)$.[5] By a standard prediction error decomposition argument, the log likelihood function for the full sample equals the sum of the conditional log likelihoods in equation (2.7),[6]

$$L_T(y_T, y_{T-1}, \ldots, y_1; \psi) = \sum_{t=1,T} l_t(y_t; \psi). \tag{2.8}$$

The maximum likelihood estimator (MLE) for the true parameters $\psi'_0 \equiv (\theta'_0, \eta'_0)$, say $\hat{\psi}_T$, is found by the maximization of equation (2.8).

Assuming the conditional density and the mean and variance functions to be differentiable for all $\psi \in \Theta \times H \equiv \Psi$, $\hat{\psi}_T$ therefore solves

$$S_T(y_T, y_{T-1}, \ldots, y_1; \psi) \equiv \sum_{t=1,T} S_t(y_t; \psi) = 0, \tag{2.9}$$

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

$$\nabla_\theta l_t(y_t; \psi) = f[z_t(\theta); \eta]^{-1} f'[z_t(\theta); \eta] \nabla_\theta z_t(\theta) - 0.5\sigma_t^2(\theta)^{-1} \nabla_\theta \sigma_t^2(\theta), \tag{2.10}$$

where $f'(z_t(\theta); \eta)$ denotes the derivative of the density function with respect to the first element, and

$$\nabla_\theta z_t(\theta) = -\nabla_\theta \mu_t(\theta) \sigma_t^2(\theta)^{-1/2} - 0.5\varepsilon_t(\theta) \sigma_t^2(\theta)^{-3/2} \nabla_\theta \sigma_t^2(\theta). \tag{2.11}$$

In practice, the actual solution to the set of $m + k$ non-linear equations in (2.9) will have to proceed by numerical techniques. Engle (1982) and Bollerslev (1986) provide a discussion of some of the alternative iterative procedures that have been successfully employed in the estimation of ARCH models.

Of course, the actual implementation of the maximum likelihood procedure requires an explicit assumption regarding the conditional density in equation (2.7). By far the most commonly employed distribution in the literature is the normal,

$$f[z_t(\theta)] = (2\pi)^{-1/2} \exp[-0.5 z_t(\theta)^2]. \tag{2.12}$$

Since the normal distribution is uniquely determined by its first two moments, only the conditional mean and variance parameters enter the likelihood function in

[5] In the multivariate context, $l_t(y_t; \psi) \equiv \ln\{f[\varepsilon_t(\theta) \Omega_t(\theta)^{-1/2}; \eta]\} - 0.5 \ln(|\Omega_t(\theta)|)$, where $|\cdot|$ denotes the determinant.

[6] In most empirical applications the likelihood function is conditioned on a number of initial observations and nuisance parameters in order to start up the recursions for the conditional mean and variance functions. Subject to proper stationarity conditions this practice does not alter the asymptotic distribution of the resulting MLE.

equation (2.8); i.e. $\psi = \theta$. If the conditional mean and variance functions are both differentiable for all $\theta \in \Theta$, it follows that the score vector in equation (2.10) takes the simple form,

$$s_t(\dot{y}_t; \theta) = \nabla_\theta \mu_t(\theta) \varepsilon_t(\theta) \sigma_t^2(\theta)^{-1/2} + 0.5 \nabla_\theta \sigma_t^2(\theta) \sigma_t^2(\theta)^{-1/2} [\varepsilon_t(\theta)^2 \sigma_t^2(\theta)^{-1} - 1]. \quad (2.13)$$

From the discussion in Section 2.1 the ARCH model with conditionally normal errors results in a leptokurtic unconditional distribution. However, the degree of leptokurtosis induced by the time-varying conditional variance often does not capture all of the leptokurtosis present in high frequency speculative prices. To circumvent this problem Bollerslev (1987) suggested using a standardized t-distribution with $\eta > 2$ degrees of freedom,

$$f[z_t(\theta); \eta] = [\pi(\eta - 2)]^{-1/2} \Gamma[0.5(\eta + 1)] \Gamma(0.5\eta)^{-1} [1 + z_t(\theta)(\eta - 2)^{-1}]^{-(\eta+1)/2}, \quad (2.14)$$

where $\Gamma(\cdot)$ denotes the gamma function. The t-distribution is symmetric around zero, and converges to the normal distribution for $\eta \to \infty$. However, for $4 < \eta < \infty$ the conditional kurtosis equals $3(\eta - 2)/(\eta - 4)$, which exceeds the normal value of three.

Several other conditional distributions have been employed in the literature to fully capture the degree of tail fatness in speculative prices. The density function for the generalized error distribution (GED) used in Nelson (1991) is given by

$$f[z_t(\theta); \eta] = \eta \lambda^{-1} 2^{-(1+1/\eta)} \Gamma(\eta^{-1})^{-1} \exp[-0.5 | z_t(\theta) \lambda^{-1} |^\eta], \quad (2.15)$$

where

$$\lambda = [2^{(-2/\eta)} \Gamma(\eta^{-1}) \Gamma(3\eta^{-1})^{-1}]^{1/2}. \quad (2.16)$$

For the tail-thickness parameter $\eta = 2$ the density equals the standard normal density in equation (2.10). For $\eta < 2$ the distribution has thicker tails than the normal, while $\eta > 2$ results in a distribution with thinner tails than the normal.

Both of these candidates for the conditional density impose the restriction of symmetry. From an economic point of view the hypothesis of symmetry is of interest since risk averse agents will induce correlation between shocks to the mean and shocks to the variance as developed more fully by Campbell and Hentschel (1992).

Engle and Gonzalez-Rivera (1991) propose to estimate the conditional density nonparametrically. The procedure they develop first estimates the parameters of the model using the Gaussian likelihood. The density of the residuals standardized by their estimated conditional standard deviations is then estimated using a linear spline with smoothness priors. The estimated density is then taken to be the true density and the new likelihood function is maximized. The use of the linear spline simplifies the estimation in that the derivatives with respect to the conditional density are easy to compute and store, which would not be the case for kernels or many other methods. In a Monte Carlo study, this approach improved the efficiency beyond the quasi MLE, particularly when the density was highly non-normal and skewed.

2.2.2. Testing

The primary appeal of the maximum likelihood technique stems from the well-known optimality conditions of the resulting estimators under ideal conditions. Crowder (1976) gives one set of sufficient regularity conditions for the MLE in models with dependent observations to be consistent and asymptotically normally distributed. Verification of these regularity conditions has proven extremely difficult for the general ARCH class of models, and a formal proof is only available for a few special cases, including the GARCH (1, 1) model in Lumsdaine (1992a) and Lee and Hansen (1993).[7] The common practice in empirical studies has been to proceed under the assumption that the necessary regularity conditions are satisfied.

In particular, if the conditional density is correctly specified and the true parameter vector $\psi_0 \in \mathrm{int}(\Psi)$, then a central limit theorem argument yields that

$$T^{1/2}(\hat{\psi}_T - \psi_0) \rightarrow N(0, A_0^{-1}), \tag{2.17}$$

where $\rightarrow$ denotes convergence in distribution. Again, the technical difficulties in verifying (2.17) are formidable. The asymptotic covariance matrix for the MLE is equal to the inverse of the information matrix evaluated at the true parameter vector ψ_0,

$$A_0 = -T^{-1} \sum_{t=1,T} E[\nabla_\psi s_t(y_t;\psi_0)]. \tag{2.18}$$

The inverse of this matrix is less than the asymptotic covariance matrix for all other estimators by a positive definite matrix. In practice, a consistent estimate for A_0 is available by evaluating the corresponding sample analogue at $\hat{\psi}_T$; i.e. replace $E[\nabla_\psi s_t(y_t;\psi_0)]$ in equation (2.18) with $\nabla_\psi s_t(y_t;\hat{\psi}_T)$. Furthermore, as shown below, the terms with second derivatives typically have expected value equal to zero and therefore do not need to be calculated.

Under the assumption of a correctly specified conditional density, the information matrix equality implies that $A_0 = B_0$, where B_0 denotes the expected value of the

[7]As discussed in Section 3 below, the condition that $E(\ln(\alpha_1 z_t^2 + \beta_1)] < 0$ in Lunsdaine (1992a) ensures that the GARCH(1,1) model is strictly stationary and ergodic. Note also, that by Jensen's inequality $E(\ln(\alpha_1 z_t^2 + \beta_1)] < \ln E(\alpha_1 z_t^2 + \beta_1) = \ln(\alpha_1 + \beta_1)$, so the parameter region covers the interesting IGARCH(1,1) case in which $\alpha_1 + \beta_1 = 1$.

outer product of the gradients evaluated at the true parameters,

$$B_0 = T^{-1} \sum_{t=1,T} E[s_t(y_t;\psi_0)s_t(y_t;\psi_0)']. \tag{2.19}$$

The outer product of the sample gradients evaluated at $\hat{\psi}_T$ therefore provides an alternative covariance matrix estimator; that is, replace the summand in equation (2.19) by the sample analogues $s_t(y_t;\hat{\psi}_T)s_t(y_t;\hat{\psi}_T)'$. Since analytical derivatives in ARCH models often involve very complicated recursive expressions, it is common in empirical applications to make use of numerical derivatives to approximate their analytical counterparts. The estimator defined from equation (2.19) has the computational advantage that only first order derivatives are needed, as numerical second order derivatives are likely to be unstable.[8]

In many applications of ARCH models the parameter vector may be partitioned as $\theta' = (\theta_1', \theta_2')$ where θ_1 and θ_2 operate a sequential cut on $\Theta_1 \times \Theta_2 = \Theta$, such that θ_1 parametrizes the conditional mean and θ_2 parametrizes the conditional variance function for y_t. Thus, $\nabla_{\theta_2}\mu_t(\theta) = 0$, and although $\nabla_{\theta_1}\sigma_t^2(\theta) \neq 0$ for all $\theta \in \Theta$, it is possible to show that, under fairly general symmetrical distributional assumptions regarding z_t and for particular functional forms of the ARCH conditional variance, the information matrix for $\theta' = (\theta_1', \theta_2')$ becomes block diagonal. Engle (1982) gives conditions and provides a formal proof for the linear ARCH(q) model in equation (1.6) under the assumption of conditional normality. As a result, asymptotically efficient estimates for θ_{02} may be calculated on the basis of a consistent estimate for θ_{01}, and vice versa. In particular, for the linear regression model with covariance stationary ARCH disturbances, the regression coefficients may be consistently estimated by OLS, and asymptotically efficient estimates for the ARCH parameters in the conditional variance calculated on the basis of the OLS regression residuals. The loss in asymptotic efficiency for the OLS coefficient estimates may be arbitrarily large, however. Also, the conventional OLS standard errors are generally inappropriate, and should be modified to take account of the heteroskedasticity as in White (1980). In particular, as noted by Milhøj (1985), Diebold (1987), Bollerslev (1988) and Stambaugh (1993) when testing for serial correlation in the mean in the presence of ARCH effects, the conventional Bartlett standard error for the estimated autocorrelations, given by the inverse of the square root of the sample size, may severely underestimate the true standard error.

There are several important cases in which block-diagonality does not hold. For example, block-diagonality typically fails for functional forms, such as EGARCH, in which σ_t^2 is an asymmetric function of lagged residuals. Another important exception is the ARCH–M class of models discussed in Section 1.4. Consistent

[8] In the Berndt, Hall, Hall and Hausman (1974) (BHHH) algorithm, often used in the maximization of the likelihood function, the covariance matrix from the auxiliary OLS regression in the last iteration provides an estimate of B_0. In a small scale Monte Carlo experiment Bollerslev and Wooldridge (1992) found that this estimator performed reasonably well under ideal conditions.

estimation of the parameters in ARCH–M models generally requires that both the conditional mean and variance functions be correctly specified and estimated simultaneously. A formal analysis of these issues is contained in Engle et al. (1987), Pagan and Hong (1991), Pagan and Sabau (1987a, 1987b) and Pagan and Ullah (1988).

Standard hypothesis testing procedures concerning the true parameter vector are directly available from equation (2.17). To illustrate, let the null hypothesis of interest be stated as $r(\psi_0) = 0$, where $r: \Theta \times H \to R^l$ is differentiable on $\mathrm{int}(\Psi)$ and $l < m + k$. If $\psi_0 \in \mathrm{int}(\Psi)$ and $\mathrm{rank}[\nabla_\psi r(\psi_0)] = l$, the Wald statistic takes the familiar form

$$W_T = T \cdot r(\hat{\psi}_T)' [\nabla_\psi r(\hat{\psi}_T) C_T^{-1} \nabla_\psi r(\hat{\psi}_T)']^{-1} r(\hat{\psi}_T),$$

where C_T denotes a consistent estimator of the covariance matrix for the parameter estimates under the alternative. If the null hypothesis is true and the regularity conditions are satisfied, the Wald statistic is asymptotically chi-square distributed with $(m + k) - l$ degrees of freedom.

Similarly, let $\hat{\psi}_{0T}$ denote the MLE under the null hypothesis. The conventional likelihood ratio (LR) statistic,

$$\mathrm{LR}_T = 2[L_T(y_T, y_{T-1}, \ldots, y_1; \hat{\psi}_T) - L_T(y_T, y_{T-1}, \ldots, y_1; \hat{\psi}_{0T})],$$

should then be the realization of a chi-square distribution with $(m + k) - l$ degrees of freedom if the null hypothesis is true and $\psi_0 \in \mathrm{int}(\Psi)$.

As discussed already in Section 2.1 above, when testing hypotheses about the parameters in the conditional variance of estimated ARCH models, non-negativity constraints must often be imposed, so that ψ_0 is on the boundary of the admissible parameter space. As a result the two-sided critical value from the standard asymptotic chi-square distribution will lead to a conservative test; recent discussions of general issues related to testing inequality constraints are given in Gourieroux et al. (1982), Kodde and Palm (1986) and Wolak (1991).

Another complication that often arises when testing in ARCH models, also alluded to in Section 2.1 above, concerns the lack of identification of certain parameters under the null hypothesis. This in turn leads to a singularity of the information matrix under the null and a breakdown of standard testing procedures. For instance, as previously noted in the GARCH(1, 1) model, β_1 and ω are not jointly identified under the null hypothesis that $\alpha_1 = 0$. Similarly, in the ARCH–M model, $\mu_t(\theta) = \mu + \delta\sigma_t^2$ with $\mu \neq 0$, the parameter δ is only identified if the conditional variance is time-varying. Thus, a standard joint test for ARCH effects and $\delta = 0$ is not feasible. Of course, such identification problems are not unique to the ARCH class of models, and a general discussion is beyond the scope of the present chapter; for a more detailed analysis along these lines we refer the reader to Davies (1977), Watson and Engle (1985) and Andrews and Ploberger (1992, 1993).

The finite sample evidence on the performance of ARCH MLE estimators and test statistics is still fairly limited: examples include Engle et al. (1985), Bollerslev and Wooldridge (1992), Lumsdaine (1992b) and Baillie et al. (1993). For the GARCH(1, 1) model with conditional normal errors, the available Monte Carlo evidence suggests that the estimate for $\alpha_1 + \beta_1$ is downward biased and skewed to the right in small samples. This bias in $\hat{\alpha}_1 + \hat{\beta}_1$ comes from a downward bias in $\hat{\beta}_1$, while $\hat{\alpha}_1$ is upward biased. Consistent with the theoretical results in Lumsdaine (1992a) there appears to be no discontinuity in the finite sample distribution of the estimators at the IGARCH(1, 1) boundary; i.e. $\alpha_1 + \beta_1 = 1$. Reliable inference from the LM, Wald and LR test statistics generally does require moderately large sample sizes of at least two hundred or more observations, however.

2.3. *Quasi-maximum likelihood methods*

The assumption of conditional normality for the standardized innovations are difficult to justify in many empirical applications. This has motivated the use of alternative parametric distributional assumptions such as the densities in equation (2.14) or (2.15). Alternatively, the MLE based on the normal density in equation (2.12) may be given a quasi-maximum likelihood interpretation.

If the conditional mean and variance functions are correctly specified, the normal quasi-score in equation (2.13) evaluated at the true parameters θ_0 will have the martingale difference property,

$$E_t\{\nabla_\theta \mu_t(\theta_0)\varepsilon_t(\theta_0)\sigma_t^{-2}(\theta_0) + 0.5\nabla_\theta\sigma_t^2(\theta_0)\sigma_t^2(\theta_0)^{-1}[\varepsilon_t(\theta_0)^2\sigma_t^2(\theta_0)^{-1} - 1]\} = 0. \tag{2.20}$$

Since equation (2.20) holds for any value of the true parameters, the QMLE obtained by maximizing the conditional normal likelihood function defined by equations (2.7), (2.8) and (2.12), say $\hat{\theta}_{T,\mathrm{QMLE}}$, is Fisher-consistent; that is, $E[S_T(y_T, y_{T-1}, \ldots, y_1; \theta)] = 0$ for any $\theta \in \Theta$. Under appropriate regularity conditions this is sufficient to establish consistency and asymptotic normality of $\hat{\theta}_{T,\mathrm{QMLE}}$. Wooldridge (1994) provides a formal discussion. Furthermore, following Weiss (1984, 1986), the asymptotic distribution for the QMLE takes the form

$$T^{1/2}(\hat{\theta}_{T,\mathrm{QMLE}} - \theta_0) \to N(0, A_0^{-1}B_0A_0^{-1}). \tag{2.21}$$

Under appropriate, and difficult to verify, regularity conditions, the A_0 and B_0 matrices are consistently estimated by the sample counterparts from equations (2.18) and (2.19), respectively.

Provided that the first two conditional moments are correctly specified, it follows from equation (2.13) that

$$E_t[\nabla_\theta s_t(y_t; \theta_0)] = -\nabla_\theta \mu_t(\theta)\nabla_\theta \mu_t(\theta)'\sigma_t^2(\theta)^{-1} - \tfrac{1}{2}\nabla_\theta\sigma_t^2(\theta)\nabla_\theta\sigma_t^2(\theta)'\sigma_t^2(\theta)^{-2}. \tag{2.22}$$

As pointed out by Bollerslev and Wooldridge (1992), a convenient estimate of the information matrix, A_0, involving only first derivatives is therefore available by replacing the right hand side of equation (2.18) with the sample realizations from equation (2.22).

The finite sample distribution of the QMLE and the Wald statistics based on the robust covariance matrix estimator constructed from equations (2.18), (2.19) and (2.22) has been investigated by Bollerslev and Wooldridge (1992). For symmetric departures from conditional normality, the QMLE is generally close to the exact MLE. However, as noted by Engle and Gonzales-Rivera (1991), for non-symmetric conditional distributions both the asymptotic and the finite sample loss in efficiency may be quite large, and semiparametric density estimation, as discussed in Section 1.5, may be preferred.

2.4. *Specification checks*

2.4.1. *Lagrange multiplier diagnostic tests*

After a model is selected and estimated, it is generally desirable to test whether it adequately represents the data. A useful array of tests can readily be constructed from calculating Lagrange multiplier tests against particular parametric alternatives. Since almost any moment condition can be formulated as the score against some alternative, these tests may also be interpreted as conditional moment tests; see Newey (1985) and Tauchen (1985). Whenever one computes a collection of test statistics, the question of the appropriate size of the full procedure arises. It is generally impossible to control precisely the size of a procedure when there are many correlated test statistics and conventional econometric practice does not require this. When these tests are viewed as diagnostic tests, they are simply aids in the model building process and may well be part of a sequential testing procedure anyway. In this section, we will show how to develop tests against a variety of interesting alternatives to any particular model. We focus on the simplest and most useful case.

Suppose we have estimated a parametric model with the assumption that each observation is conditionally normal with mean zero and variance $\sigma_t^2 = \sigma_t^2(\theta)$. Then the score can be written as a special case of (2.13),

$$s_t(y_t, \theta) = \nabla_\theta \ln \sigma_t^2(\theta)[\varepsilon_t^2(\theta)\sigma_t^2(\theta)^{-1} - 1]. \tag{2.23}$$

In order to conserve space, equation (2.23) may be written more compactly as

$$s_{\theta t} \equiv x_{\theta t} u_t, \tag{2.24}$$

where $x_{\theta t}$ denotes the $k \times 1$ vector of derivatives of the logarithm of the conditional variance equation with respect to the parameters θ, and $u_t \equiv \varepsilon_t^2(\theta)\sigma_t^2(\theta)^{-1} - 1$ defines the generalized residuals. From the first order conditions in equation (2.9), the MLE for θ, $\hat{\theta}_T$, solves

$$\sum_{t=1,T} \hat{s}_{\theta t} = \sum_{t=1,T} \hat{x}_{\theta t}\hat{u}_t = 0. \tag{2.25}$$

Suppose that the additional set of r parameters, represented by the $r \times 1$ vector γ, have been implicitly set to zero during estimation. We wish to test whether this restriction is supported by the data. That is, the null hypothesis may be expressed as $\gamma_0 = 0$, where $\sigma_t^2 = \sigma_t^2(\theta, \gamma)$. Also, suppose that the score with respect to γ has the same form as in equation (2.24),

$$s_{\gamma t} = x_{\gamma t}u_t. \tag{2.26}$$

Under fairly general regularity conditions, the scores themselves when evaluated at the true parameter under the null hypothesis, θ_0, will satisfy a martingale central limit theorem. Therefore,

$$T^{1/2}s_{\psi}(\theta_0) \to N(0, \mathrm{V}), \tag{2.27}$$

where $V = A_0$ denotes the covariance matrix of the scores. The conventional form of the Lagrange multiplier test, as in Breusch and Pagan (1979) or Engle (1984), is then given by

$$\xi_T = T^{-1} \sum_{t=1,T} \hat{s}'_{\psi t}\hat{V}^{-1} \sum_{t=1,T} \hat{s}_{\psi t}, \tag{2.28}$$

where $\psi = (\theta, \gamma)$, represent estimates evaluated under the null hypothesis, and $\hat{V}$ denotes a consistent estimate of V. As discussed in Section 2.2, a convenient estimate of the information matrix is given by the outer product of the scores,

$$\hat{B}_T = T^{-1} \sum_{t=1,T} \hat{s}_{\psi t}\hat{s}'_{\psi t}, \tag{2.29}$$

so that the test statistic can be computed in terms of a regression. Specifically, let the $T \times 1$ vector of ones be denoted $\imath$, and the $T \times (k + r)$ matrix of scores evaluated under the null hypothesis be denoted by $\hat{S}' = \{\hat{s}_{\psi 1}, \hat{s}_{\psi 2}, \ldots, \hat{s}_{\psi T}\}$. Then a simple form of the LM test is obtained from

$$\xi_{1T} = \imath'\hat{S}(\hat{S}'\hat{S})^{-1}\hat{S}'\imath = TR^2, \tag{2.30}$$

where the R^2 is the uncentered fraction of variance explained by the regression of a vector of ones on all the scores. The test statistic in equation (2.30) is often referred to as the outer product of the gradient, or OPG, version of the test. It is very easy to compute. In particular, using the BHHH estimation algorithm, the test statistic is simply obtained by one step of the BHHH algorithm from the maximum achieved under the null hypothesis.

Studies of this version of the LM test, such as MacKinnon and White (1985) and Bollerslev and Wooldridge (1992), often find that it has size distortions and is not very powerful as it does not utilize the structure of the problem under the null hypothesis to obtain the best estimate of the information matrix. Of course the R^2 in (2.30) will be overstated if the likelihood function has not been fully maximized under the null so that (2.25) is not satisfied. One might recommend a first step correction by BHHH to be certain that this is achieved.

An alternative estimate of V corresponding to equation (2.19) is available from taking expectations of $S'S$. In the simplified notation of this section,

$$E(S'S) = \sum_{t=1,T} E(u_t^2 x_t x_t') = E(u_t^2) \sum_{t=1,T} E(x_t x_t'), \tag{2.31}$$

where it is assumed that the conditional expectation $E_{t-1}(u_t^2)$ is time invariant. Of course, this will be true if the standardized innovations $\varepsilon_t(\theta)\sigma_t^2(\theta)^{-1/2}$ has a distribution which does not depend upon time or past information, as typically assumed in estimation. Consequently, an alternative consistent estimator of V is given by

$$\hat{V}_T = (T^{-1}\hat{u}'\hat{u})(T^{-1}\hat{X}'\hat{X}), \tag{2.32}$$

where $u' = \{u_1, \ldots, u_T\}$, $X' = \{x_1, \ldots, x_T\}$, and $x_t' = \{x_{\theta t}', X_{\gamma t}'\}$. Since $\iota' S = u'X$, the Lagrange multiplier test based on the estimator in equation (2.32) may also be computed from an auxiliary regression,

$$\xi_{2T} = \hat{u}'\hat{X}(\hat{X}'\hat{X})^{-1}\hat{X}'\hat{u} = TR^2. \tag{2.33}$$

Here the regression is of the percentage difference between the squared residuals and the estimated conditional variance regressed on the gradient of the logarithm of the conditional variance with respect to all the parameters including those set to zero under the null hypothesis. This test statistic is similar to one step of a Gauss–Newton iteration from an estimate under the null. It is called the Hessian estimate by Bollerslev and Wooldridge (1992) because it can also be derived by setting components of the Hessian equal to their expected value, assuming only that the first two moments are correctly specified, as discussed in Section 2.3. This version of the test has considerable intuitive appeal as it checks for remaining conditional heteroskedasticity in u_t as a function of x_t. It also performed better than the OPG test in the simulations reported by Bollerslev and Wooldridge (1992). This is also the version of the test used by Engle and Ng (1993) to compare various model specifications. As noted by Engle and Ng (1993), the likelihood must be fully maximized under the null if the test is to have the correct size. An approach to dealing with this issue would be to first regress $\hat{u}_t$ on $\hat{x}_{\theta t}$ and then form the test on the basis of the residuals from this regression. The R^2 of this regression should be zero if the likelihood is maximized, so this is merely a numerical procedure to purge the test statistic of contributions from loose convergence criteria.

Both of these procedures develop the asymptotic distribution under the null hypothesis that the model is correctly specified including the normality assumption. Recently, Wooldridge (1990) and Bollerslev and Wooldridge (1992) have developed robust LM tests which have the same limiting distribution under any null specifying that the first two conditional moments are correct. This follows in the line of conditional moment tests for GMM or QMLE as in Newey (1985), Tauchen (1985) and White (1987, 1994).

To derive these tests, consider the Taylor series expansions of the scores around the true parameter values, $s_\gamma(\theta)$ and $s_\theta(\theta_0)$,

$$T^{1/2}s_\gamma(\theta_0) = T^{1/2}s_\gamma(\hat{\theta}_T) + T^{1/2}\frac{\partial s_\gamma}{\partial \theta'}(\hat{\theta}_T - \theta_0), \tag{2.34}$$

$$T^{1/2}s_\theta(\theta_0) = T^{1/2}s_\theta(\hat{\theta}_T) + T^{1/2}\frac{\partial s_\theta}{\partial \theta'}(\hat{\theta}_T - \theta_0), \tag{2.35}$$

where the derivatives of the scores are evaluated at $\hat{\theta}_T$. The derivatives in equations (2.34) and (2.35) are simply the $H_{\gamma\theta}$ and $H_{\theta\theta}$ elements of the Hessian, respectively. The distribution of the score with respect to γ evaluated at $\hat{\theta}_T$ is readily obtained from the left hand side of equation (2.34). In particular substituting in (2.35), and using (2.26) to give the limiting distribution of the scores,

$$T^{1/2}s_\gamma(\hat{\theta}_T) \to N(0, W), \tag{2.36}$$

where

$$W \equiv V_{\gamma\gamma} - H_{\gamma\theta}H_{\theta\theta}^{-1}V_{\theta\gamma} - V_{\gamma\theta}H_{\theta\theta}^{-1}H_{\theta\gamma} + H_{\gamma\theta}H_{\theta\theta}^{-1}V_{\theta\theta}H_{\theta\theta}^{-1}H_{\theta\gamma}. \tag{2.37}$$

Notice first, that if the scores are the derivatives of the true likelihood, then the information matrix equality will hold, and therefore $H = V$ asymptotically. In this case we get the conventional LM test described in (2.28) and computed generally either as (2.30) or (2.33). If the normality assumption underlying the likelihood is false so that the estimates are viewed as quasi-maximum likelihood estimators, then the expressions in equations (2.36) and (2.37) are needed.

As pointed out by Wooldridge (1990), any score which has the additional property that $H_{\theta\gamma}$ converges in probability to zero can be tested simply as a limiting normal with covariance matrix $V_{\gamma\gamma}$, or as a TR^2 type test from a regression of a vector of ones on $\hat{s}_{\gamma t}$. By proper redefinition of the score, such a test can always be constructed. To illustrate, suppose that $s_{\gamma t} = x_{\gamma t}u_t$, $s_{\theta t} = x_{\theta t}u_t$ and $\partial u_t/\partial\theta = -x_{\theta t}$. Also define

$$s^*_{\gamma t} \equiv (x_{\gamma t} - x^p_{\gamma t})u_t, \tag{2.38}$$

where

$$x^p_{\gamma t} \equiv x_{\theta t}\left(\sum_{t=1,T} x_{\theta t}x'_{\theta t}\right)^{-1}\left(\sum_{t=1,T} x_{\theta t}x'_{\gamma t}\right). \tag{2.39}$$

The statistic based on $s^*_{\gamma t}$ in equation (2.38) then tests only the part of $x_{\gamma t}$ which is orthogonal to the scores used to estimate the model under the null hypothesis. This strategy generalizes to more complicated settings as discussed by Bollerslev and Wooldridge (1992).

2.4.2. *BDS specification tests*

As discussed in Section 2.1.2, the asymptotic distribution of the BDS test is unaffected by passing the data through a linear, e.g. ARMA, filter. Since an ARCH model typically assumes that the standardized residuals $z_t \equiv \varepsilon_t\sigma_t^{-1}$ are i.i.d., it seems reasonable to use the BDS test as a specification test by applying it to the fitted standardized residuals from an ARCH model. Fortunately, the BDS test applied to the standardized residuals has considerable power to detect misspecification in ARCH models. Unfortunately, the asymptotic distribution of the test is strongly affected by the fitting of the ARCH model. As documented by Brock et al. (1991) and Hsieh (1991), BDS tests on the standardized residuals from fitted ARCH models reject much too infrequently. In light of the filtering properties of misspecified ARCH models, discussed in Section 4 below, this may not be too surprising.

The asymptotic distribution of the BDS test for ARCH residuals has not yet been derived. One commonly employed procedure to get around this problem is to simply simulate the critical values of the test statistic; i.e. in each replication generate data by Monte Carlo methods from the specific ARCH model, then estimate the ARCH model and compute the BDS test for the standardized residuals. This approach is obviously very demanding computationally.

Brock and Potter (1992) suggest another possibility for the case in which the conditional mean of the observed data is known. Applying the BDS test to the logarithm of the squared known residuals, i.e. $\ln(\varepsilon_t^2) = \ln(z_t^2) + \ln(\sigma_t^2)$, separates $\ln(\varepsilon_t^2)$ into an i.i.d. component, $\ln(z_t^2)$, and a component which can be estimated by non-linear regression methods. Under the null of a correctly specified ARCH model, $\ln(z_t^2) = \ln(\varepsilon_t^2) - \ln(\sigma_t^2)$ is i.i.d. and, subject to the regularity conditions of Brock and Potter (1992) or Brock et al. (1991), the asymptotic distribution of the BDS test is the same whether applied to $\ln(z_t^2)$ or to the fitted values $\ln(\hat{z}_t^2) \equiv \ln(\varepsilon_t^2) - \ln(\hat{\sigma}_t^2)$. While the assumption of a known conditional mean is obviously unrealistic in some applications, it may be a reasonable approximation for high frequency financial time series, where the noise component tends to swamp the conditional mean component.

3. Stationary and ergodic properties

3.1. Strict stationarity

In evaluating the stationarity of ARCH models, it is convenient to recursively substitute for the lagged ε_t's and σ_t^2's. For completeness, consider the multivariate case where

$$\varepsilon_t = \Omega_t^{1/2} Z_t, \quad Z_t \sim \text{i.i.d.}, \quad E(Z_t) = 0_{n \times 1}, \quad E(Z_t Z_t') = \mathrm{I}_{n \times n} \tag{3.1}$$

and

$$\Omega_t = \Omega(t, Z_{t-1}, Z_{t-2}, \ldots). \tag{3.2}$$

Using the ergodicity criterion from Corollary 1.4.2 in Krengel (1985), it follows that *strict stationarity* of $\{\varepsilon_t\}_{t=-\infty,\infty}$ is equivalent to the condition

$$\Omega_t = \Omega(Z_{t-1}, Z_{t-2}, \ldots), \tag{3.3}$$

with $\Omega(\cdot, \cdot, \ldots)$ measurable, and

$$\text{Trace}(\Omega_t \Omega_t') < \infty \quad \text{a.s.} \tag{3.4}$$

Equation (3.3) eliminates direct dependence of $\{\Omega_t\}$ on t, while (3.4) ensures that random shocks to $\{\Omega_t\}$ die out rapidly enough to keep $(\Omega_t\}$ from exploding asymptotically.

In the univariate EGARCH(p, q) model, for example, equation (3.2) is obtained by exponentiating both sides of the definition in equation (1.11). Since $\ln(\sigma_t^2)$ is written in ARMA(p, q) form, it is easy to see that if $(1 + \sum_{j=1,q} \alpha_j x^j)$ and $(1 - \sum_{i=1,p} \beta_i x^i)$ have no common roots, equations (3.3) and (3.4) are equivalent to all the roots of $(1 - \sum_{i=1,p} \beta_i x^i)$ lying outside the unit circle. Similarly, in the bivariate EGARCH model defined in Section 6.4 below, $\ln(\sigma_{m,t}^2)$, $\ln(\sigma_{p,t}^2)$ and $\beta_{p,t}$ all follow ARMA processes giving rise to ARMA stationarity conditions.

One sufficient condition for (3.4) is moment boundedness; i.e. clearly $E[\text{Trace}(\Omega_t \Omega_t')^p]$ finite for some $p > 0$ implies $\text{Trace}(\Omega_t \Omega_t') < \infty$ a.s. For example, Bollerslev (1986) shows that in the univariate GARCH(p, q) model defined by equation (1.9), $E(\sigma_t^2)$ is finite and $\{\varepsilon_t\}$ is covariance stationary, when $\sum_{i=1,p} \beta_i + \sum_{j=1,q} \alpha_j < 1$. This is a sufficient, but not a necessary condition for strict stationarity, however. Because ARCH processes are thick tailed, the conditions for "weak" or covariance stationarity are often more stringent than the conditions for "strict" stationarity.

For instance, in the univariate GARCH(1, 1) model, (3.2) takes the form

$$\sigma_t^2 = \omega\left[1 + \sum_{k=1,\infty} \prod_{i=1,k} (\beta_1 + \alpha_1 z_{t-i}^2)\right]. \tag{3.5}$$

Nelson (1990b) shows that when $\omega > 0$, $\sigma_t^2 < \infty$ a.s., and $\{\varepsilon_t, \sigma_t^2\}$ is strictly stationary if and only if $E[\ln(\beta_1 + \alpha_1 z_t^2)] < 0$. An easy application of Jensen's inequality shows that this is a much weaker requirement than $\alpha_1 + \beta_1 < 1$, the necessary and sufficient condition for $\{\varepsilon_t\}$ to be covariance stationary. For example, the simple ARCH(1) model with $z_t \sim N(0, 1)$ and $\alpha_1 = 3$ and $\beta_1 = 0$, is strictly but not weakly stationary.

To grasp the intuition behind this seemingly paradoxical result, consider the terms in the summation in (3.5); i.e. $\prod_{i=1,k}(\beta_1 + \alpha_1 z_{t-i}^2)$. Taking logarithms, it follows directly that $\sum_{i=1,k}\ln(\beta_1 + \alpha_1 z_{t-i}^2)$ is a random walk with drift. If $E[\ln(\beta_1 + \alpha_1 z_{t-i}^2)] \geqslant 0$, the drift is positive and the random walk diverges to ∞ a.s. as $k \to \infty$. If, on the other hand, $E[\ln(\beta_1 + \alpha_1 z_{t-i}^2)] < 0$, the drift is negative and the random walk diverges to $-\infty$ a.s. as $k \to \infty$, in which case $\prod_{i=1,k}(\beta_1 + \alpha_1 z_{t-i}^2)$ tends to zero at an exponential rate in k a.s. as $k \to \infty$. This, in turn, implies that the sum in equation (3.5) converges a.s., establishing (3.4). Measurability in (3.3) follows easily using Theorems 3.19 and 3.20 in Royden (1968).

This result for the univariate GARCH(1, 1) model generalizes fairly easily to other closely related ARCH models. For example, in the multivariate diagonal GARCH(1, 1) model, discussed in Section 6.1 below, the diagonal elements of Ω_t follow univariate GARCH(1, 1) processes. If each of these processes is stationary, the Cauchy–Schwartz inequality ensures that all of the elements in Ω_t are bounded a.s. The case of the constant conditional correlation multivariate GARCH(1, 1) model in Section 6.3 is similar. The same method can also be used in a number of other univariate cases as well. For instance, when $p = q = 1$, the stationarity condition for the model in equation (1.16) is $E[\ln(\alpha_1^+ \mathrm{I}(z_t > 0)|z_t|^\gamma + \alpha_1^- \mathrm{I}(z_t \leqslant 0)|z_t|^\gamma)] < 0$.

Establishing stationarity becomes much more difficult when we complicate the models even slightly. The extension to the higher order univariate GARCH(p, q) model has recently been carried out by Bougerol and Picard (1992) with methods which may be more generally applicable. There exists a large mathematics literature on conditions for stationarity and ergodicity for Markov chains; see, e.g. Nummelin and Tuominen (1982) and Tweedie (1983a). These conditions can sometimes be verified for ARCH models, although much work remains establishing useful stationarity criteria even for many commonly-used models.

3.2. *Persistence*

The notion of "persistence" of a shock to volatility within the ARCH class of models is considerably more complicated than the corresponding concept of persistence in the mean for linear models. This is because even strictly stationary ARCH models frequently do not possess finite moments.

Suppose that $\{\sigma_t^2\}$ is strictly stationary and ergodic. Let $F(\sigma_t^2)$ denote the unconditional cumulative distribution function (cdf) for σ_t^2, and let $F_s(\sigma_t^2)$ denote the cdf of σ_t^2 given information at time $s < t$. Then for any s, $F_s(\sigma_t^2) - F(\sigma_t^2)$ converges to 0 at all continuity points as $t \to \infty$; i.e. time s information drops out of the forecast distribution as $t \to \infty$. Therefore, one perfectly reasonable definition of "persistence" would be to say that shocks fail to persist when $\{\sigma_t^2\}$ is stationary and ergodic.

It is equally natural, however, to define persistence of shocks in terms of forecast moments; i.e. to choose some $\eta > 0$ and to say that shocks to σ_t^2 fail to persist if and only if for every s, $E_s(\sigma_t^{2\eta})$ converges, as $t \to \infty$, to a finite limit independent of time s information. Such a definition of persistence may be particularly appropriate when an economic theory makes a forecast moment, as opposed to a forecast distribution, the object of interest.

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

Consider, for example, the GARCH(1, 1) model, in which

$$\sigma_{t+1}^2 = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2 = \omega + \sigma_t^2(\alpha_1 z_t^2 + \beta_1). \tag{3.6}$$

The expectation of σ_t^2 as of time s, is given by

$$E_s(\sigma_t^2) = \omega\left[\sum_{k=0,t-s-1} (\alpha_1 + \beta_1)^k\right] + \sigma_s^2(\alpha_1 + \beta_1)^{t-s}. \tag{3.7}$$

It is easy to see that $E_s(\sigma_t^2)$ converges to the unconditional variance of $\omega/(1 - \alpha_1 - \beta_1)$ as $t \to \infty$ if and only if $\alpha_1 + \beta_1 < 1$. In the IGARCH model with $\omega > 0$ and $\alpha_1 + \beta_1 = 1$, it follows that $E_s(\sigma_t^2) \to \infty$ a.s. as $t \to \infty$. Nevertheless, as discussed in the previous section, IGARCH models are strictly stationary and ergodic. In fact, as shown by Nelson (1990b) in the IGARCH(1, 1) model $E_s(\sigma_t^{2\eta})$ converges to a finite limit independent of time s information as $t \to \infty$ whenever $\eta < 1$. This ambiguity of "persistence" holds more generally. When the support of z_t is unbounded it follows from Nelson (1990b) that in any stationary and ergodic GARCH(1, 1) model, $E_s(\sigma_t^{2\eta})$ diverges for all sufficiently large η, and converges for all sufficiently small η. For many other ARCH models, moment convergence may be most easily established with the methods used in Tweedie (1983b).

While the relevant criterion for persistence may be dictated by economic theory, in practice tractability may also play an important role. For example, $E_s(\sigma_t^2)$, and its multivariate extension discussed in Section 6.5 below, can often be evaluated even when strict stationarity is difficult to establish, or when $E_s(\sigma_t^{2\eta})$ for $\eta \neq 1$ is intractable.

Even so, in many applications, simple moment convergence criterion have not been successfully developed. This includes quite simple cases, such as the univariate GARCH(p, q) model when $p > 1$ or $q > 1$. The same is true for multivariate models, in which co-persistence is an issue. In such cases, the choice of $\eta = 1$ may be impossible to avoid. Nevertheless, it is important to recognize that apparent persistence of shocks may be driven by thick-tailed distributions rather than by inherent non-stationarity.

4. Continuous time methods

ARCH models are systems of non-linear stochastic difference equations. This makes their probabilistic and statistical properties, such as stationarity, moment finiteness, consistency and asymptotic normality of MLE, more difficult than is the case with linear models. One way to simplify the analysis of ARCH models is to approximate the stochastic difference equations with more tractable stochastic differential equations. On the other hand, for certain purposes, notably in the computation of point forecasts and maximum likelihood estimates, ARCH models are more convenient than the stochastic differential equation models of time-varying volatility common in the finance literature; see, e.g. Wiggins (1987), Hull and White (1987), Gennotte and Marsh (1991), Heston (1991) and Andersen (1992a).

Suppose that the process $\{X_t\}$ is governed by the stochastic integral equation

$$X_t = X_0 + \int_0^t \mu(X_s)\,\mathrm{d}s + \int_0^t \Omega^{1/2}(X_s)\,\mathrm{d}W_s, \tag{4.1}$$

where $\{W_t\}$ is an $N \times 1$ standard Brownian motion, and $\mu(\cdot)$ and $\Omega^{1/2}(\cdot)$ are continuous functions from R^N into R^N and the space of $N \times N$ real matrices respectively. The starting value, X_0, may be fixed or random. Following Karatzas and Shreve (1988) and Ethier and Kurtz (1986), if equation (4.1) has a unique weak-sense solution, the distribution of the $\{X_t\}$ process is then completely determined by the following four characteristics:[9]

(i) the cumulative distribution function, $F(x_0)$, of the starting point X_0;
(ii) the drift $\mu(x)$;
(iii) the conditional covariance matrix $\Omega(x) \equiv \Omega(x)^{1/2}[\Omega(x)^{1/2}]'$;[10]
(iv) the continuity, with probability one, of $\{X_t\}$ as a function of time.

Our interest here is either in approximating (4.1) by an ARCH model or visa versa. To that end, consider a sequence of first-order Markov processes $\{_hX_t\}$, whose

[9] Formally, we consider $\{X_t\}$ and the approximating discrete time processes $\{_hX_t\}$ as random variables in $D_{R^n}[0, \infty)$, the space of right continuous functions with finite left limits, equipped with the Skorohod topology. $D_{R^n}[0, \infty)$ is a complete, separable metric space [see, e.g. Chapter 3 in Ethier and Kurtz (1986)].

[10] $\Omega(x)^{1/2}$ is a matrix square root of $\Omega(x)$, though it need not be the symmetric square root since we require only that $\Omega(x)^{1/2}[\Omega(x)^{1/2}]' = \Omega(x)$, not $\Omega(x)^{1/2}\Omega(x)^{1/2} = \Omega(x)$.

sample paths are random step functions with jumps at times $h, 2h, 3h, \ldots$. For each $h > 0$, and each non-negative integer k, define the drift and covariance functions by $\mu_h(x) \equiv h^{-1}E[({}_hX_{k+1} - {}_hX_k)|{}_hX_k = x]$, and $\Omega_h(x) \equiv h^{-1}\,\mathrm{Cov}[({}_hX_{k+1} - {}_hX_k)|{}_hX_k = x]$, respectively. Also, let $F_h({}_hx_0)$ denote the cumulative distribution function for ${}_hX_0$. Since (i)–(iv) completely characterize the distribution of the $\{X_t\}$ process, it seems intuitive that weak convergence of $\{{}_hX_t\}$ to $\{X_t\}$ can be achieved by "matching" these properties in the limit as $h \downarrow 0$. Stroock and Varadhan (1979) showed that this is indeed the case.

Theorem 4.1. [Stroock and Varadhan (1979)]

Let the stochastic integral equation (4.1) have a unique weak-sense solution. Then $\{{}_hX_t\}$ converges weakly to $\{X_t\}$ for $h \downarrow 0$ if

(i′) $F_h(\cdot) \to F(\cdot)$ as $h \downarrow 0$ at all continuity points of $F(\cdot)$,
(ii′) $\mu_h(x) \to \mu(x)$ uniformly on every bounded x set as $h \downarrow 0$,
(iii′) $\Omega_h(x) \to \Omega(x)$ uniformly on every bounded x set as $h \downarrow 0$,
(iv′) for some $\delta > 0$, $h^{-1}E[\|{}_hX_{k+1} - {}_hX_k\|^{2+\delta}|{}_hX_k = x] \to 0$ uniformly on every bounded x set as $h \downarrow 0$.[11]

This result, along with various extensions, is fundamental in all of the continuous record asymptotics discussed below.

Deriving the theory of continuous time approximation for ARCH models in its full generality is well beyond the scope of this chapter. Instead, we shall simply illustrate the use of these methods by explicit reference to a diffusion model frequently applied in the options pricing literature; see e.g. Wiggins (1987). The model considers an asset price, Y_t, and its instantaneous returns volatility, σ_t. The continuous time process for the joint evolution of $\{Y_t, \sigma_t\}$ with fixed starting values, (Y_0, σ_0), is given by

$$\mathrm{d}Y_t = \mu Y_t\,\mathrm{d}t + Y_t\sigma_t\,\mathrm{d}W_{1,t} \tag{4.2}$$

and

$$\mathrm{d}[\ln(\sigma_t^2)] = -\beta[\ln(\sigma_t^2) - \alpha]\,\mathrm{d}t + \psi\,\mathrm{d}W_{2,t}, \tag{4.3}$$

where μ, ψ, β and α denote the parameters of the process, and $W_{1,t}$ and $W_{2,t}$ are driftless Brownian motions independent of (Y_0, σ_0^2) that satisfy

$$\begin{bmatrix} \mathrm{d}W_{1,t} \\ \mathrm{d}W_{2,t} \end{bmatrix} [\mathrm{d}W_{1,t} \quad \mathrm{d}W_{2,t}] = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \mathrm{d}t. \tag{4.4}$$

[11] We define the matrix norm, $\|\cdot\|$, by $\|A\| \equiv [\mathrm{Trace}(AA')]^{1/2}$. It is easy to see why (i′)–(iii′) match (i)–(iii) in the limit as $h \downarrow 0$. That (iv′) leads to (iv) follows from Hölder's inequality; see Theorem 2.2 in Nelson (1990a) for a formal proof.

Of course in practice, the price process is only observable at discrete time intervals. However, the continuous time model in equations (4.2)–(4.4) provides a very convenient framework for analyzing issues related to theoretical asset pricing, in general, and option pricing, in particular. Also, by Ito's lemma equation (4.2) may be equivalently written as

$$\mathrm{d}y_t = \left(\mu - \frac{\sigma_1^2}{2}\right)\mathrm{d}t + \sigma_t\,\mathrm{d}W_{1,t}, \tag{4.2$'$}$$

where $y_t \equiv \ln(Y_t)$. For many purposes this is a more tractable differential equation.

4.1. ARCH models as approximations to diffusions

Suppose that an economic model specifies a diffusion model such as equation (4.1), where some of the state variables, including $\Omega(x_t)$, are unobservable. Is it possible to formulate an ARCH data generation process that is similar to the true process, in the sense that the distribution of the sample paths generated by the ARCH model and the diffusion model in equation (4.1) becomes "close" for increasingly finer discretizations?

Specifically, consider the diffusion model given by equations (4.2)–(4.4). Strategies for approximating diffusions such as this are well known. For example, Melino and Turnbull (1990) use a standard Euler approximation in defining (y_t, σ_t),[12]

$$y_{t+h} = y_t + \left(\mu - \frac{\sigma_t^2}{2}\right)h + h^{1/2}\sigma_t Z_{1,t+h}, \tag{4.5}$$

$$\ln(\sigma_{t+h}^2) = \ln(\sigma_t^2) - h\beta[\ln(\sigma_t^2) - \alpha] + h^{1/2}\psi Z_{2,t+h}, \tag{4.6}$$

for $t = h, 2h, 3h, \ldots$. Here (y_0, σ_0) is taken to be fixed, and $(Z_{1,t}, Z_{2,t})$ is assumed to be i.i.d. bivariate normal with mean vector $(0, 0)$ and

$$\mathrm{Var}\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{4.7}$$

Convergence of this set of stochastic difference equations to the diffusion in equations (4.2)–(4.4) as $h \downarrow 0$ may be verified using Theorem 4.1. In particular, (i$'$) holds trivially, since (y_0, σ_0) are constants. To check conditions (ii$'$) and (iii$'$), note that

$$h^{-1}E_t\begin{bmatrix} y_{t+h} - y_t \\ \ln(\sigma_{t+h}^2) - \ln(\sigma_t^2) \end{bmatrix} = \begin{bmatrix} (\mu - \sigma_t^2/2) \\ -\beta[\ln(\sigma_t^2) - \alpha] \end{bmatrix} \tag{4.8}$$

[12] See Pardoux and Talay (1985) for a general discussion of the Euler approximation technique.

and

$$h^{-1}\operatorname{Var}_t\begin{bmatrix} y_{t+h}-y_t \\ \ln(\sigma_{t+h}^2)-\ln(\sigma_t^2)\end{bmatrix}=\begin{bmatrix}\sigma_t^2 & \sigma_t\rho\psi \\ \sigma_t\rho\psi & \psi^2\end{bmatrix}, \tag{4.9}$$

which matches the drift and diffusion matrix of (4.2)–(4.4). Condition (iv′) is nearly trivially satisfied, since $Z_{1,t}$ and $Z_{2,t}$ are normally distributed with arbitrary finite moments. The final step of verifying that the limit diffusion has a unique weak-sense solution is often the most difficult and least intuitive part of the proof for convergence. Nelson (1990a) summarizes several sets of sufficient conditions, however, and formally shows that the process defined by (4.5)–(4.7) satisfies these conditions.

While conditionally heteroskedastic, the model defined by the stochastic difference equations (4.5)–(4.7) is *not* an ARCH model. In particular, for $\rho \neq 1$ σ_t^2 is not simply a function of the discretely observed sample path of $\{y_t\}$ combined with a startup value σ_0^2. More technically, while the conditional variance $(y_{t+h}-y_t)$ given the σ-algebra generated by $\{y_\tau, \sigma_\tau^2\}_{0\leqslant\tau\leqslant t}$ equals $h\sigma_t^2$, it is not, in general, the conditional variance of $(y_{t+h}-y_t)$ given the smaller σ-algebra generated by $\{y_\tau\}_{0,h,2h\ldots h[t/h]}$ and σ_0^2. Unfortunately, this latter conditional variance is not available in closed form.[13]

To create an ARCH approximation to the diffusion in (4.2)–(4.4), simply replace (4.6) by

$$\ln(\sigma_{t+h}^2)=\ln(\sigma_t^2)-h\beta[\ln(\sigma_t^2)-\alpha]+h^{1/2}g(Z_{1,t+h}), \tag{4.10}$$

where $g(\cdot)$ is measurable with $E[|g(Z_{1,t+h})|^{2+\delta}]<\infty$ for some $\delta>0$, and

$$\operatorname{Var}\begin{bmatrix} Z_{1,t} \\ g(Z_{1,t})\end{bmatrix}=\begin{bmatrix}1 & \rho\psi \\ \rho\psi & \psi^2\end{bmatrix}. \tag{4.11}$$

As an ARCH model, the discretization defined by (4.5), (4.10) and (4.11) inherits the convenient properties usually associated with ARCH models, such as the easily computed likelihoods and inference procedures discussed in Section 2 above. As such, it is a far more tractable approximation to (4.2)–(4.4) than the discretization defined by equations (4.5)–(4.7).

To complete the formulation of the ARCH approximation, an explicit $g(\cdot)$ function is needed. Since $E(|Z_{1,t}|)=(2/\pi)^{1/2}$, $E(Z_{1,t}|Z_{1,t}|)=0$ and $\operatorname{Var}(|Z_{1,t}|)=1-(2/\pi)$, one possible formulation would be

$$g(Z_{1,t})=\rho\psi Z_{1,t}+\psi\left(\frac{1-\rho^2}{1-2/\pi}\right)^{1/2}\left[|Z_{1,t}|-\left(\frac{2}{\pi}\right)^{1/2}\right]. \tag{4.12}$$

[13] Jacquier et al. (1994) have recently proposed a computationally tractable algorithm for computing this conditional variance.

This corresponds directly to the EGARCH model in equation (1.11). Alternatively,

$$g(Z_{1,t}) = \rho\psi Z_{1,t} + \psi\left(\frac{1-\rho^2}{2}\right)^{1/2}(Z_{1,t}^2 - 1) \tag{4.13}$$

also satisfies equation (4.11). This latter specification turns out to be the asymptotically optimal filter for $h\downarrow 0$, as discussed in Nelson and Foster (1991, 1994) and Section 4.3 below.

4.2. *Diffusions as approximations to ARCH models*

Now consider the question of how to best approximate a discrete time ARCH model with a continuous time diffusion. This can yield important insights into the workings of a particular ARCH model. For example, the stationary distribution of σ_t^2 in the AR(1) version of the EGARCH model given by equaton (1.11) is intractable. However, the sequence of EGARCH models defined by equations (4.5) and (4.10)–(4.12) converges weakly to the diffusion process in (4.2)–(4.4). When $\beta > 0$, the stationary distribution of $\ln(\sigma_t^2)$ is $N(\alpha, \psi^2/2\beta)$. Nelson (1990a) shows that this is also the limit of the stationary distribution of $\ln(\sigma_t^2)$ in the sequence of EGARCH models (4.5) and (4.10)–(4.12) as $h\downarrow 0$. Similarly, the continuous limit may result in convenient approximations for forecast moments of the $\{y_t, \sigma_t^2\}$ process.

Different ARCH models will generally result in different limit diffusions. To illustrate, suppose that the data are generated by a simple martingale model with a GARCH(1, 1) error structure as in equation (1.9). In the present notation, the process takes the form,

$$y_{t+h} = y_t + \sigma_t h z_{t+h} = y_t + \varepsilon_{t+h} \tag{4.14}$$

and

$$\sigma_{t+h}^2 = \omega h + (1 - \theta h - \alpha h^{1/2})\sigma_t^2 + \mathrm{h}^{1/2}\alpha\varepsilon_{t+h}^2, \tag{4.15}$$

where given time t information, ε_{t+h} is $N(0, \sigma_t^2)$, and (x_0, σ_0) is assumed to be fixed. Note that using the notation for the GARCH(p, q) model in equation (1.9) $\alpha_1 + \beta_1 = 1 - \theta h$, so for increasing sampling frequencies, i.e., as $h\downarrow 0$, the parameters of the process approach the IGARCH(1, 1) boundary as discussed in Section 3. Following Nelson (1990a)

$$h^{-1}E_t\begin{bmatrix} y_{t+h} - y_t \\ \sigma_{t+h}^2 - \sigma_t^2 \end{bmatrix} = \begin{bmatrix} 0 \\ \omega - \theta\sigma_t^2 \end{bmatrix} \tag{4.16}$$

and

$$h^{-1}\,\mathrm{Var}_t\begin{bmatrix} y_{t+h}-y_t \\ \sigma^2_{t+h}-\sigma^2_t \end{bmatrix} = \begin{bmatrix} \sigma^2_t & 0 \\ 0 & 2\alpha^2\sigma^4_t \end{bmatrix}. \tag{4.17}$$

Thus, from Theorem 4.1 the limit diffusion is given by

$$\mathrm{d}x_t = \sigma_t\,\mathrm{d}W_{1,t} \tag{4.18}$$

and

$$\mathrm{d}\sigma^2_t = (\omega - \theta\sigma^2_t)\,\mathrm{d}t + \sqrt{2}\alpha\sigma^2_t\,\mathrm{d}W_{2,t}, \tag{4.19}$$

where $W_{1,t}$ and $W_{2,t}$ are independent standard Brownian motions.

The diffusion defined by equations (4.18) and (4.19) is quite different from the EGARCH limit in equations (4.2)–(4.4). For example, if $\theta/2\alpha^2 > -1$, the stationary distribution of σ^2_t in (4.19) is an inverted gamma, so as $h\downarrow 0$ and $t\to\infty$, the normalized increments $h^{-1/2}(y_{t+h}-y_t)$ are conditionally normally distributed but unconditionally Student's t. In particular, in the IGARCH case corresponding to $\theta = 0$, as $h\downarrow 0$ and $t\to\infty$, $h^{-1/2}(y_{t+h}-y_t)$ approaches a Student's t distribution with two degrees of freedom. In the EGARCH case, however, $h^{-1/2}(y_{t+h}-y_t)$ is conditionally normal but is unconditionally a normal–lognormal mixture. When σ^2_t is stationary, the GARCH formulation in (1.9) therefore gives rise to unconditionally thicker-tailed residuals than the EGARCH model in equation (1.11).

4.3. *ARCH models as filters and forecasters*

Suppose that discretely sampled observations are only available for a subset of the state variables in (4.1), and that interest centers on estimating the unobservable state variable(s), $\Omega(x_t)$. Doing this optimally via a non-linear Kalman filter is computationally burdensome; see, e.g. Kitagawa (1987).[14] Alternatively, the data might be passed through a discrete time ARCH model, and the resulting conditional variances from the ARCH model viewed as estimates for $\Omega(x_t)$. Nelson (1992) shows that under fairly mild regularity conditions, a wide variety of misspecified ARCH models consistently extract conditional variances from high frequency time series. The regularity conditions require that the conditional distribution of the observable series is not too thick tailed, and that the conditional covariance matrix moves smoothly over time. Intuitively, the GARCH filter defined by equation (1.9)

[14] An approximate linear Kalman filter for a discretized version of (4.1) has been employed by Harvey et al. (1994). The exact non-linear filter for a discretized version of (4.1) has been developed by Jacquier et al. (1994). Danielson and Richard (1993) and Shephard (1993) also calculate the exact likelihood by computer intensive methods.

estimates the conditional variance by averaging squared residuals over some time window, resulting in a nonparametric estimate for the conditional variance at each point in time. Many other ARCH models can be similarly interpreted.

While many different ARCH models may serve as consistent filters for the same diffusion process, efficiency issues may also be relevant in the design of an ARCH model. To illustrate, suppose that the y_t process is observable at time intervals of length h, but that σ_t^2 is not observed. Let $\hat{\sigma}_0^2$ denote some initial estimate of the conditional variance at time 0, with subsequent estimates generated by the recursion

$$\ln(\hat{\sigma}_{t+h}^2) = \ln(\hat{\sigma}_t^2) + h\kappa(\hat{\sigma}_t^2) + h^{1/2}g[\hat{\sigma}_t^2, h^{-1/2}(y_{t+h} - y_t)]. \tag{4.20}$$

The set of admissible $g(\cdot,\cdot)$ functions is restricted by the requirement that $E_t\{g[\sigma_t^2, h^{-1/2}(y_{t+h} - y_t)]\}$ be close to zero for small values of h.[15] Define the normalized estimation error from this filter extraction as $q_t \equiv h^{-1/4}[\ln(\hat{\sigma}_t^2) - \ln(\sigma_t^2)]$.

Nelson and Foster (1994) derive a diffusion approximation for q_t when the data have been generated by the diffusion in equations (4.2)–(4.4) and the time interval shrinks to zero. In particular, they show that q_t is approximately normally distributed, and that by choosing the $g(\cdot,\cdot)$ function to minimize the asymptotic variance of q_t, the drift term for $\ln(\sigma_t^2)$ in the ARCH model, $\kappa(\cdot)$, does not appear in the resulting minimized asymptotic variance for the measurement error. The effect is second order in comparison to that of the $g(\cdot,\cdot)$ term, and creates only an asymptotically negligible bias in q_t. However, for $\kappa(\sigma_t^2) \equiv -\beta[\ln(\sigma_t^2) - \alpha]$, the leading term of this asymptotic bias also disappears. It is easy to verify that the conditions of Theorem 4.1 are satisfied for the ARCH model defined by equation (4.20) with $\kappa(\sigma^2) \equiv -\beta[\ln(\sigma^2) - \alpha]$ and the variance minimizing $g(\cdot,\cdot)$. Thus, as a data generation process this ARCH model converges weakly to the diffusion in (4.2)–(4.4). In the diffusion limit the first two conditional moments completely characterize the process, and the optimal ARCH filter matches these moments.

The above result on the optimal choice of an ARCH filter may easily be extended to other diffusions and more general data generating processes. For example, suppose that the true data generation process is given by the stochastic difference equation analogue of (4.2)–(4.4),

$$y_{t+h} = y_t + h\left(\mu - \frac{\sigma_t^2}{2}\right) + \xi_{1,t}, \tag{4.21}$$

$$\ln(\sigma_{t+h}^2) = \ln(\sigma_t^2) - h\beta[\ln(\sigma_t^2) - \alpha] + h^{1/2}\xi_{2,t}, \tag{4.22}$$

where $(\xi_{1,t}\sigma_t^{-1}, \xi_{2,t})$ is i.i.d. and independent of t, h and y_t, with conditional density $f(\xi_{1,t}, \xi_{2,t}|\sigma_t)$ with mean $(0,0)$, bounded $2+\delta$ absolute moments, $\mathrm{Var}_t(\xi_{1,t}) = \sigma_t^2$,

[15] Formally, the function must satisfy that $h^{-1/4}E_t\{g[\sigma_t^2, h^{-1/2}(y_{t+h} - y_t)]\} \to 0$ uniformly on bounded (y_t, σ_t) sets as $h \downarrow 0$.

and $\mathrm{Var}_t(\xi_{2,t}) = \psi^2$. This model can be shown to converge weakly to (4.2)–(4.4) as $h \downarrow 0$. The asymptotically optimal filter for the model given by equations (4.21) and (4.22) has been derived in Nelson and Foster (1994). This optimal ARCH filter when (4.21) and (4.22) are the data generation process is not necessarily the same as the optimal filter for (4.2)–(4.4). The increments in a diffusion such as (4.2)–(4.4) are approximately conditionally normal over very short time intervals, whereas the innovations $(\xi_{1,t}, \xi_{2,t})$ in (4.21) and (4.22) may be non-normal. This affects the properties of the ARCH filter. Consider estimating a variance based on i.i.d. draws from some distribution with mean zero. If the distribution is normal, averaging squared residuals is an asymptotically efficient method of estimating the variance. Least squares, however, can be very inefficient if the distribution is thicker tailed than the normal. This theory of robust scale estimation, discussed in Davidian and Carroll (1987) and Huber (1977), carries over to the ARCH case. For example, estimating σ_t^2 by squaring a distributed lag of absolute residuals, as proposed by Taylor (1986) and Schwert (1989a, b), will be more efficient than estimating σ_t^2 with a distributed lag of squared residuals if the conditional distribution of the innovations is sufficiently thicker tailed than the normal.

One property of optimally designed ARCH filters concerns their resemblance to the true data generating process. In particular, if the data were generated by the asymptotically optimal ARCH filter, the functional form for the second conditional moment of the state variables would be the same as in the true data generating process. If the conditional first moments also match, the second order bias is similarly eliminated. Nelson and Foster (1991) show that ARCH models which match these first two conditional moments also have the desirable property that the forecasts generated by the possibly misspecified ARCH model approach the forecasts from the true model as $h \downarrow 0$. Thus, even when ARCH models are misspecified, they may consistently estimate the conditional variances. Unfortunately, the behavior of ARCH filters with estimated as opposed to known parameters, and the properties of the parameter estimates themselves, are not yet well understood.

5. Aggregation and forecasting

5.1. Temporal aggregation

The continuous record asymptotics discussed in the preceding section summarizes the approximate relationships between continuous time stochastic differential equations and discrete time ARCH models defined at increasingly higher sampling frequencies. While the approximating stochastic differential equations may result in more manageable theoretical considerations, the relationship between high frequency ARCH stochastic difference equations and the implied stochastic process for less frequently sampled, or temporally aggregated, data is often of direct importance for empirical work. For instance, when deciding on the most appropriate sampling interval for inference purposes more efficient parameter estimates for the low frequency process may be available from the model estimates obtained with high frequency data. Conversely, in some instances the high frequency process may be of primary interest, while only low frequency data is available. The non-linearities in ARCH models severely complicate a formal analysis of temporal aggregation. In contrast to the linear ARIMA class of models for conditional means, most parametric ARCH models are only closed under temporal aggregation subject to specific qualifications.

Following Drost and Nijman (1993) we say that $\{\varepsilon_t\}$ is a weak GARCH(p,q) process if ε_t is serially uncorrelated with unconditional mean zero, and σ_t^2, as defined in equation (1.9), corresponds to the best linear projection of ε_t^2 on the space spanned by $\{1, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots\}$. More specifically,

$$E(\varepsilon_t^2 - \sigma_t^2) = E[(\varepsilon_t^2 - \sigma_t^2)\varepsilon_{t-i}] = E[(\varepsilon_t^2 - \sigma_t^2)\varepsilon_{t-i}^2] = 0 \qquad i = 1, 2, \ldots. \tag{5.1}$$

This definition of a weak GARCH(p,q) model obviously encompasses the conventional GARCH(p,q) model in which σ_t^2 is equal to the conditional expectation of ε_t^2 based on the full information set at time $t-1$ as a special case. Whereas the conventional GARCH(p,q) class of models is not closed under temporal aggregation, Drost and Nijman (1993) show that temporal aggregation of ARIMA models with weak GARCH(p,q) errors lead to another ARIMA model with weak GARCH(p',q') errors. The orders of this temporally aggregated model and the model parameters depend on the original model characteristics.

To illustrate, suppose that $\{\varepsilon_t\}$ follows a weak GARCH(1, 1) model with parameters ω, α_1 and β_1. Let $\{\varepsilon_t^{(m)}\}$ denote the discrete time temporally aggregated process defined at $t, t+m, t+2m, \ldots$. For a stock variable $\varepsilon_t^{(m)}$ is obtained by sampling ε_t every mth period. For a flow variable $\varepsilon_t^{(m)} \equiv \varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_{t-m+1}$. In both cases, it is possible to show that the temporally aggregated process, $\{\varepsilon_t^{(m)}\}$, is also weak GARCH(1, 1) with parameters $\omega^{(m)} = \omega[1 - (\alpha_1 + \beta_1)^m]/(1 - \alpha_1 - \beta_1)$ and $\alpha_1^{(m)} = (\alpha_1 + \beta_1)^m - \beta_1^{(m)}$, where $\beta_1^{(m)}$ is a complicated function of the parameters for the original process. Thus, $\alpha_1^{(m)} + \beta_1^{(m)} = (\alpha_1 + \beta_1)^m$, and conditional heteroskedasticity disappears as the sampling frequency decreases, provided that $\alpha_1 + \beta_1 < 1$. Moreover, for flow variables the conditional kurtosis of the standardized residuals, $\varepsilon_t^{(m)}[\sigma_t^{(m)}]^{-1}$, converges to the normal value of three for less frequently sampled observations. This convergence to asymptotic normality for decreasing sampling frequencies of temporally aggregated covariance stationary GARCH(p,q) flow variables has been shown previously by Diebold (1988), using a standard central limit theorem type argument.

These results highlight the fact that the assumption of i.i.d. innovations invoked in maximum likelihood estimation of GARCH models is necessarily specific to the particular sampling frequency employed in the estimation. If $\varepsilon_t\sigma_t^{-1}$ is assumed i.i.d., the distribution of $\varepsilon_t^{(m)}[\sigma_t^{(m)}]^{-1}$ will generally not be time invariant. Following the discussion in Section 2.3, the estimation by maximum likelihood methods could be given a quasi-maximum likelihood type interpretation, however. Issues pertaining to the efficiency of the resulting estimators remain unresolved.

The extension of the aggregation results for the GARCH(p, q) model to other parametric specifications is in principle straightforward. The cross sectional aggregation of multivariate GARCH processes, which may be particularly relevant in the formation of portfolios, have been addressed in Nijman and Sentana (1993).

5.2. Forecast error distributions

One of the primary objectives of econometric time series model building is often the construction of out-of-sample predictions. In conventional econometric models with time invariant innovation variances, the prediction error uncertainty is an increasing function of the prediction horizon, and does not depend on the origin of the forecast. In the presence of ARCH errors, however, the forecast accuracy will depend non-trivially on the current information set. The proper construction of forecast error intervals and post-sample structural stability tests, therefore, both require the evaluation of future conditional error variances.[16]

A detailed analysis of the forecast moments for various GARCH models is available in Engle and Bollerslev (1986) and Baillie and Bollerslev (1992). Although both of these studies develop expressions for the second and higher moments of the forecast error distributions, this is generally not enough for the proper construction of confidence intervals, since the forecast error distributions will be leptokurtic and time-varying.

A possible solution to this problem is suggested by Baillie and Bollerslev (1992), who argue for the use of the Cornish–Fisher asymptotic expansion to take account of the higher order dependencies in the construction of the prediction error intervals. The implementation of this expansion requires the evaluation of higher order conditional moments of ε_{t+s}, which can be quite complicated. Interestingly, in a small scale Monte Carlo experiment, Baillie and Bollerslev (1992) find that under the assumption of conditional normality for $\varepsilon_t\sigma_t^{-1}$, the ninety-five percent confidence interval for multi-step predictions from the GARCH(1, 1) model, constructed under the erroneous assumption of conditional normality of $\varepsilon_{t+s}[E(\sigma_{t+s}^2)]^{-1/2}$ for $s > 1$, has a coverage probability quite close to ninety-five percent. The one percent fractile is typically underestimated by falsely assuming conditional normality of the multi-step leptokurtic prediction errors, however.

Most of the above mentioned results are specialized to the GARCH class of models, although extensions to allow for asymmetric or leverage terms and multivariate formulations in principle would be straightforward. Analogous results on forecasting $\ln(\sigma_t^2)$ for EGARCH models are easily obtained. Closed form expressions

[16] Also, as discussed earlier, the forecasts of the future conditional variances are often of direct interest in applications with financial data.

for the moments of the forecast error distribution for the EGARCH model are not available, however.

As discussed in Section 4.3, an alternative approximation to the forecast error distribution may be based upon the diffusion limit of the ARCH model. If the sampling frequency is "high" so that the discrete time ARCH model is a "close" approximation to the continuous time diffusion limit, the distribution of the forecasts should be "good" approximations too; see Nelson and Foster (1991). In particular, if the unconditional distribution of the diffusion limit can be derived, this would provide an approximation to the distribution of the long horizon forecasts from a strictly stationary model.

Of course, the characteristics of the prediction error distribution may also be analyzed through the use of numerical methods. In particular, let $f_{t,s}(\varepsilon_{t+s})$ denote the density function for ε_{t+s} conditional on information up through time t. Under the assumption of a time invariant conditional density function for the standardized innovations, $f(\varepsilon_t\sigma_t^{-1})$, the prediction error density for ε_{t+s} is then given by the convolution

$$f_{t,s}(\varepsilon_{t+s}) = \int\cdots\int f(\varepsilon_{t+s}\sigma_{t+s}^{-1})f(\varepsilon_{t+s-1}\sigma_{t+s-1}^{-1})\cdots f(\varepsilon_{t+1}\sigma_{t+1}^{-1})\,\mathrm{d}\varepsilon_{t+s-1}\,\mathrm{d}\varepsilon_{t+s-2}\cdots\mathrm{d}\varepsilon_{t+1}.$$

Evaluation of this multi-step prediction error density may proceed directly by numerical integration. This is illustrated within a Bayesian context by Geweke (1989a, b), who shows how the use of importance sampling and antithetic variables can be employed in accelerating the convergence of the Monte Carlo integration. In accordance with the results in Baillie and Bollerslev (1992), Geweke (1989a) finds that for conditional normally distributed one-step-ahead prediction errors, the shorter the forecast horizon s, and the more tranquil the periods before the origin of the forecast, the closer to normality is the prediction error distribution for ε_{t+s}.

6. Multivariate specifications

Financial market volatility moves together over time across assets and markets. Recognizing this commonality through a multivariate modeling framework leads to obvious gains in efficiency. Several interesting issues in the structural analysis of asset pricing theories and the linkage of different financial markets also call for an explicit multivariate ARCH approach in order to capture the temporal dependencies in the conditional variances and covariances.

In keeping with the notation of the previous sections, the $N \times 1$ vector stochastic process $\{\varepsilon_t\}$ is defined to follow a multivariate ARCH process if $E_{t-1}(\varepsilon_t) = 0$, but the $N \times N$ conditional covariance matrix,

$$E_{t-1}(\varepsilon_t\varepsilon_t') = \Omega_t, \tag{6.1}$$

depends non-trivially on the past of the process. From a theoretical perspective, inference in multivariate ARCH models poses no added conceptual difficulties in comparison to the procedures for the univariate case outlined in Section 2 above.

To illustrate, consider the log likelihood function for $\{\varepsilon_T, \varepsilon_{T-1}, \ldots, \varepsilon_1\}$ obtained under the assumption of conditional multivariate normality,

$$L_T(\varepsilon_T, \varepsilon_{T-1}, \ldots, \varepsilon_1; \psi) = -0.5[TN\ln(2\pi) + \sum_{t=1,T} (\ln|\Omega_t| + \varepsilon_t'\Omega_t^{-1}\varepsilon_t)]. \tag{6.2}$$

This function corresponds directly to the conditional likelihood function for the univariate ARCH model defined by equations (2.7), (2.8) and (2.12), and maximum likelihood, or quasi-maximum likelihood, procedures may proceed as discussed in Section 2. Of course, the actual implementation of a multivariate ARCH model necessarily requires some assumptions regarding the format of the temporal dependencies in the conditional covariance matrix sequence, $\{\Omega_t\}$.

Several key issues must be faced in choosing a parametrization for Ω_t. Firstly, the sheer number of potential parameters in a geneal formulation is overwhelming. All useful specifications must necessarily restrict the dimensionality of the parameter space, and it is critical to determine whether they impose important untested characteristics on the conditional variance process. A second consideration is whether such restrictions impose the required positive semi-definiteness of the conditional covariance matrix estimators. Thirdly, it is important to recognize whether Granger causality in variance as in Granger et al. (1986) is allowed by the chosen parametrization; that is, does the past information on one variable predict the conditional variance of another. A fourth issue is whether the correlations or regression coefficients are time-varying and, if so, do they have the same persistence properties as the variances? A fifth issue worth considering is whether there are linear combinations of the variables, or portfolios, with less persistence than individual series, or assets. Closely related is the question of whether there exist simple statistics which are sufficient to forecast the entire covariance matrix. Finally, it is natural to ask whether there are multivariate asymmetric effects, and if so how these may influence both the variances and covariances. Below we shall briefly review some of the parametrizations that have been applied in the literature, and comment on their appropriateness for answering each of the questions posed above.

6.1. *Vector ARCH and diagonal ARCH*

Let vech$(\cdot)$ denote the vector-half operator, which stacks the lower triangular elements of an $N \times N$ matrix as an $[N(N+1)/2] \times 1$ vector. Since the conditional covariance matrix is symmetric, vech(Ω_t) contains all the unique elements in Ω_t. Following Kraft and Engle (1982) and Bollerslev et al. (1988), a natural multivariate extension of the univariate GARCH(p,q) model defined in equation (1.9) is then given by

$$\begin{aligned}\mathrm{vech}(\Omega_t) &= W + \sum_{i=1,q} A_i \,\mathrm{vech}(\varepsilon_{t-i}\varepsilon'_{t-i}) + \sum_{j=1,p} B_j \,\mathrm{vech}(\Omega_{t-j}) \\ &\equiv W + A(L)\mathrm{vech}(\varepsilon_{t-1}\varepsilon'_{t-1}) + B(L)\,\mathrm{vech}(\Omega_{t-1}),\end{aligned} \tag{6.3}$$

where W is an $[N(N+1)/2] \times 1$ vector, and the A_i and B_j matrices are of dimension $[N(N+1)/2] \times [N(N+1)/2]$. This general formulation is termed the vec representation by Engle and Kroner (1993). It allows each of the elements in $\{\Omega_t\}$ to depend on all of the most recent q past cross products of the ε_t's and all of the most recent p lagged conditional variances and covariances, resulting in a total of $[N(N+1)/2] \cdot [1 + (p+q)N(N+1)/2]$ parameters. Even for low dimensions of N and small values of p and q the number of parameters is very large; e.g. for $N = 5$ and $p = q = 1$ the unrestricted version of (6.3) contains 465 parameters. This allows plenty of flexibility to answer most, but not all, of the questions above.[17] However, this large number of parameters is clearly unmanageable, and conditions to ensure that the conditional covariance matrices are positive definite a.s. for all t are difficult to impose and verify; Engle and Kroner (1993) provides one set of sufficient conditions discussed below.

In practice, some simplifying assumptions will therefore have to be imposed. In the diagonal GARCH(p,q) model, originally suggested by Bollerslev et al. (1988), the A_i and B_j matrices are all taken to be diagonal. Thus, the (i,j)th element in $\{\Omega_t\}$ only depends on the corresponding past (i,j)th elements in $\{\varepsilon_t\varepsilon'_t\}$ and $\{\Omega_t\}$. This restriction reduces the number of parameters to $[N(N+1)/2](1+p+q)$. These restrictions are intuitively reasonable, and can be interpreted in terms of a filtering estimate of each variance and covariance. However, this model clearly does not allow for causality in variance, co-persistence in variance, as discussed in Section 6.5 below, or asymmetries.

Necessary and sufficient conditions on the parameters to ensure that the conditional covariance matrices in the diagonal GARCH(p,q) model are positive definite a.s. are most easily derived by expressing the model in terms of Hadamard products. In particular, define the symmetric $N \times N$ matrices A_i^* and B_j^* implicitly by $A_i = \mathrm{diag}[\mathrm{vech}(A_i^*)]$ $i = 1,\ldots,q$, $B_j = \mathrm{diag}[\mathrm{vech}(B_j^*)]$ $j = 1,\ldots,p$, and $W \equiv \mathrm{vech}(W^*)$. The diagonal model may then be written as

$$\Omega_t = W^* + \sum_{i=1,q} A_i^* \odot (\varepsilon_{t-i}\varepsilon'_{t-i}) + \sum_{j=1,p} B_j^* \odot \Omega_{t-j}, \tag{6.4}$$

where $\odot$ denotes the Hadamard product.[18] It follows now by the algebra of

[17]Note, that even with this number of parameters, asymmetric terms are excluded by the focus on squared residuals.

[18]The Hadamard product of two $N \times N$ matrices A and B is defined by $\{A \odot B\}_{ij} \equiv \{A\}_{ij}\{B\}_{ij}$; see, e.g. Amemiya (1985).

Hadamard products, that Ω_t is positive definite a.s. for all t provided that W^* is positive definite, and the A_i^* and B_j^* matrices are positive semi-definite for all $i = 1, \ldots, q$ and $j = 1, \ldots, p$; see Attanasio (1991) and Marcus and Minc (1964) for a formal proof. These conditions are easy to impose and verify through a Cholesky decomposition for the parameter matrices in equation (6.4). Even simpler versions of this model which let either A_i^* or B_j^* be rank one matrices, or even simply a scalar times a matrix of ones, may be useful in some applications.

In the alternative representation of the multivariate GARCH(p,q) model termed by Engle and Kroner (1993) the Baba, Engle, Kraft and Kroner, or BEKK, representation, the conditional covariance matrix is parametrized as

$$\Omega_t = V'V + \sum_{k=1,K} \sum_{i=1,q} A'_{ki}\varepsilon_{t-i}\varepsilon'_{t-i}A_{ki} + \sum_{k=1,K} \sum_{j=1,p} B'_{kj}\Omega_{t-j}B_{kj}, \tag{6.5}$$

where V, A_{ik} $i = 1, \ldots, q$, $k = 1, \ldots, K$, and B_{jk} $j = 1, \ldots, p$, $k = 1, \ldots, K$ are all $N \times N$ matrices. This formulation has the advantage over the general specification in equation (6.3) that Ω_t is guaranteed to be positive definite a.s. for all t. The model in equation (6.5) still involves a total of $[1 + (p+q)K]N^2$ parameters. By taking $\mathrm{vech}(\Omega_t)$ we can express any model of the form (6.5) in terms of (6.3). Thus any vec model in (6.3) whose parameters can be expressed as (6.5) must be positive definite. However, in empirical applications, the structure of the A_{ik} and B_{jk} matrices must be further simplified as this model is also overparametrized. A choice made by McCurdy and Stengos (1992) is to set $K = p = q = 1$ and make A_1 and B_1 diagonal. This leads to the simple positive definite version of the diagonal vec model

$$\Omega_t = W^* + \alpha_1\alpha'_1 \odot (\varepsilon_{t-1}\varepsilon'_{t-1}) + \beta_1\beta'_1 \odot \Omega_{t-1}, \tag{6.6}$$

where $A_1 = \mathrm{diag}[\alpha_1]$ and $B_1 = \mathrm{diag}[\beta_1]$.

6.2. *Factor ARCH*

The Factor ARCH model can be thought of as an alternative simple parametrization of (6.5). Part of the appeal of this parametrization in applications with asset returns stems from its derivation in terms of a factor type model. Specifically, suppose that the $N \times 1$ vector of returns y_t has a factor structure with K factors given by the $K \times 1$ vector ξ_t, and time invariant factor loadings given by the $N \times K$ matrix B:

$$y_t = B\xi_t + \varepsilon_t. \tag{6.7}$$

Assume that the idiosyncratic shocks, ε_t, have constant conditional covariances Ψ, and that the factors, ξ_t, have conditional covariance matrix Λ_t. Also, suppose that ε_t and ξ_t are uncorrelated, or that they have constant correlations. The conditional covariance matrix of y_t then equals

$$V_{t-1}(y_t) = \Omega_t = \Psi + B\Lambda_t B'. \tag{6.8}$$

If Λ_t is diagonal with elements λ_{kt}, or if the off-diagonal elements are constant and combined into Ψ, the model may therefore be written as

$$\Omega_t = \Psi + \sum_{k=1,K} \beta_k \beta_k' \lambda_{kt}, \tag{6.9}$$

where β_k denotes the kth column in B. Thus, there are K statistics which determine the full covariance matrix. Forecasts of the variances and covariances or of any portfolio of assets, will be based only on the forecasts of these K statistics. This model was first proposed in Engle (1987), and implemented empirically by Engle et al. (1990b) and Ng et al. (1992) for treasury bills and stocks, respectively.

Diebold and Nerlove (1989) suggested a closely related latent factor model,

$$\Omega_t = \Psi + \sum_{k=1,K} \beta_k \beta_k' \delta_{kt}^2, \tag{6.9'}$$

in which the factor variances, δ_{kt}^2, were not functions of the past information set. An estimation approach based upon an approximate Kalman filter was used by Diebold and Nerlove (1989). More recently King et al. (1994) have estimated a similar latent factor model using theoretical developments in Harvey et al. (1994).

An immediate implication of (6.8) and (6.9) is that, if $K < N$, there are some portfolios with constant variance. Indeed a useful way to determine K is to find how many assets are required to form such portfolios. Engle and Kozicki (1993) present this as an application of a test for common features. This test is applied by Engle and Susmel (1993) to determine whether there is evidence that international equity markets have common volatility components. Only for a limited number of pairs of the countries analyzed can a one factor model not be rejected.

A second implication of the formulation in (6.8) is that there exist factor-representing portfolios with portfolio weights that are orthogonal to all but one set of factor loadings. In particular, consider the portfolio $r_{kt} = \phi_k' y_t$, where $\phi_k' \beta_j$ equals 1 if $j = k$ and zero otherwise. The conditional variance of r_{kt} is then given by

$$\mathrm{Var}_{t-1}(r_{kt}) = \phi_k' \Omega_t \phi_k = \psi_k + \lambda_{kt}, \tag{6.10}$$

where $\psi_k = \phi_k' \Psi \phi_k$. Thus, the portfolios r_{kt} have exactly the same time variation as the factors, which is why they are called factor-representing portfolios.

In order to estimate this model, the dependence of the λ_{kt}'s upon the past information set must also be parametrized. The simplest assumption is that there is a set of factor-representing portfolios with univariate GARCH(1,1) representations. Thus,

$$\mathrm{Var}_{t-1}(r_{kt}) = \psi_k + \alpha_k(\phi_k'\varepsilon_{t-1})^2 + \gamma_k V_{t-2}(r_{kt-1})' \tag{6.11}$$

and, therefore,

$$\Omega_t = \Psi^* + \sum_{k=1,K} \alpha_k[\beta_k\phi_k'\varepsilon_{t-1}\varepsilon_{t-1}'\phi_k\beta_k'] + \sum_{k=1,K} \gamma_k[\beta_k\phi_k'\Omega_{t-1}\phi_k\beta_k'], \tag{6.12}$$

so that the factor ARCH model is a special case of the BEKK parametrization. Clearly, more general factor ARCH models would allow the factor representing portfolios to depend upon a broader information set than the simple univariate assumption underlying (6.11).

Estimation of the factor ARCH model by full maximum likelihood together with several variations has been considered by Lin (1992). However, it is often convenient to assume that the factor-representing portfolios are known a priori. For example, Engle et al. (1990b) assumed the existence of two such portfolios: one an equally weighted treasury bill portfolio and one the Standard and Poor's 500 composite stock portfolio. A simple two step estimation procedure is then available, by first estimating the univariate models for each of the factor-representing portfolios. Taking the variance estimates from this first stage as given, the factor loadings may then be consistently estimated up to a sign, by noticing that each of the individual assets has a variance process which is linear in the factor variances, where the coefficients equal the squares of the factor loadings. While this is surely an inefficient estimator, it has the advantage that it allows estimation for arbitrarily large matrices using simple univariate procedures.

6.3. *Constant conditional correlations*

In the constant conditional correlations model of Bollerslev (1990), the time-varying conditional covariances are parametrized to be proportional to the product of the corresponding conditional standard deviations. This assumption greatly simplifies the computational burden in estimation, and conditions for Ω_t to be positive definite a.s. for all t are also easy to impose.

More explicitly, let D_t denote the $N \times N$ diagonal matrix with the conditional variances along the diagonal; i.e. $\{D_t\}_{ii} = \{\Omega_t\}_{ii}$ and $\{D_t\}_{ij} \equiv 0$ for $i \neq j$, $i, j = 1, \ldots, N$. Also, let Γ_t denote the matrix of conditional correlations; i.e. $\{\Gamma_t\}_{ij} \equiv \{\Omega_t\}_{ij}[\{\Omega_t\}_{ii} \cdot \{\Omega_t\}_{jj}]^{-1/2}$, $i, j = 1, \ldots, N$. The constant conditional correlation model then assumes that $\Gamma_t = \Gamma$ is time-invariant, so that the temporal variation in $\{\Omega_t\}$ is determined solely by the time-varying conditional variances,

$$\Omega_t = D_t^{1/2}\Gamma D_t^{1/2}. \tag{6.13}$$

If the conditional variances along the diagonal in the D_t matrices are all positive, and the conditional correlation matrix Γ is positive definite, the sequence of conditional covariance matrices $\{\Omega_t\}$ is guaranteed to be positive definite a.s. for all t. Furthermore, the inverse of Ω_t is simply given by $\Omega_t^{-1} = D_t^{-1/2}\Gamma^{-1}D_t^{-1/2}$. Thus, when calculating the likelihood function in equation (6.2), or some other multivariate objective function involving Ω_t^{-1} $t = 1,\ldots,T$, only one matrix inversion is required for each evaluation. This is especially relevant from a computational point of view when numerical derivatives are being used. Also, by a standard multivariate SURE analogy, Γ may be concentrated out of the normal likelihood function by $(D_t^{-1/2}\varepsilon_t)(D_t^{-1/2}\varepsilon_t)'$, simplifying estimation even further.

Of course, the validity of the assumption of constant conditional correlations remains an empirical question.[19] However, this particular formulation has already been successfully applied by a number of authors, including Baillie and Bollerslev (1990), Bekaert and Hodrick (1993), Bollerslev (1990), Kroner and Sultan (1991), Kroner and Claessens (1991) and Schwert and Seguin (1990).

6.4. *Bivariate EGARCH*

A bivariate version of the EGARCH model in equation (1.11) has been introduced by Braun et al. (1992) in order to model any "leverage effects," as discussed in Section 1.2.3, in conditional betas. Specifically, let $\varepsilon_{m,t}$ and $\varepsilon_{p,t}$ denote the residuals for a market index and a second portfolio or asset. The model is then given by

$$\varepsilon_{m,t} = \sigma_{m,t} z_{m,t} \tag{6.14}$$

and

$$\varepsilon_{p,t} = \beta_{p,t}\varepsilon_{m,t} + \sigma_{p,t} z_{p,t}, \tag{6.15}$$

where $\{z_{m,t}, z_{p,t}\}$ is assumed to be i.i.d with mean (0,0) and identity covariance matrix. The conditional variance of the market index, $\sigma^2_{m,t}$, is modeled by a univariate EGARCH model,

$$\ln(\sigma^2_{m,t}) = \alpha_m + \delta_m[\ln(\sigma^2_{m,t}) - \alpha_m] + \theta_m z_{m,t-1} + \gamma_m(|z_{m,t-1}| - E|z_{m,t-1}|). \tag{6.16}$$

The conditional beta of $\varepsilon_{p,t}$ with respect to $\varepsilon_{m,t}$, $\beta_{p,t}$, is modeled as

$$\beta_{p,t} = \lambda_0 + \lambda_4(\beta_{p,t-1} - \lambda_0) + \lambda_1 z_{m,t-1} z_{p,t-1} + \lambda_2 z_{m,t-1} + \lambda_3 z_{p,t-1}. \tag{6.17}$$

The coefficients λ_2 and λ_3 allow for "leverage effects" in $\beta_{p,t}$. The non-market, or

[19] A formal moment based test for the assumption of constant conditional correlations has been developed by Bera and Roh (1991).

idiosyncratic, variance of the second portfolio, $\sigma^2_{p,t}$, is parametrized as a modified univariate EGARCH model, to allow for both market and idiosyncratic news effects,

$$\ln(\sigma^2_{p,t}) = \alpha_p + \delta_p[\ln(\sigma^2_{p,t}) - \alpha_p] + \theta_p z_{p,t-1} + \gamma_p(|z_{p,t-1}| - E|z_{p,t-1}|) + \theta_{p,m} z_{m,t-1} + \gamma_{p,m}(|z_{m,t-1}| - E|z_{m,t-1}|). \tag{6.18}$$

Braun et al. (1992) find that this model provides a good description of the returns for a number of industry and size sorted portfolios.

6.5. Stationarity and co-persistence

Stationarity and moment convergence criteria for various univariate specifications were discussed in Section 3 above. Corresponding convergence criteria for multivariate ARCH models are generally complex, and explicit results are only available for a few special cases.

Specifically, consider the multivariate vec GARCH(1, 1) model defined in equation (6.3). Analogous to the expression for the univariate GARCH(1, 1) model in equation (3.10), the minimum mean square error forecast for $\mathrm{vech}(\Omega_t)$ as of time $s < t$ takes the form

$$E_s(\mathrm{vech}(\Omega_t)) = W\left[\sum_{k=0,t-s-1} (A_1 + B_1)^k\right] + (A_1 + B_1)^{t-s}\mathrm{vech}(\Omega_s), \tag{6.19}$$

where $(A_1 + B_1)^0$ is equal to the identity matrix by definition. Let $V\Lambda V^{-1}$ denote the Jordan decomposition of the matrix $A_1 + B_1$, so that $(A_1 + B_1)^{t-s} = V\Lambda^{t-s}V^{-1}$.[20] Thus, $E_s(\mathrm{vech}(\Omega_t))$ converges to the unconditional covariance matrix of the process, $W(I - A_1 - B_1)^{-1}$, for $t \to \infty$ a.s. if and only if the norm of the largest eigenvalue of $A_1 + B_1$ is strictly less than one. Similarly, by expressing the vector GARCH(p, q) model in companion first order form, it follows that the forecast moments converge, and that the process is covariance stationary if and only if the norm of the largest root of the characteristic equation $|I - A(x^{-1}) - B(x^{-1})| = 0$ is strictly less than one. A formal proof is given in Bollerslev and Engle (1993). This corresponds directly to the condition for the univariate GARCH(p, q) model in equation (1.9), where the persistence of a shock to the optimal forecast of the future conditional variances is determined by the largest root of the characteristic polynomial $\alpha(x^{-1}) + \beta(x^{-1}) = 1$. The conditions for strict stationarity and ergodicity for the multivariate GARCH(p, q) model have not yet been established.

[20] If the eigenvalues for $A_1 + B_1$ are all distinct, Λ equals the diagonal matrix of eigenvalues, and V the corresponding matrix of right eigenvectors. If some of the eigenvalues coincide, Λ takes the more general Jordan canonical form; see Anderson (1971) for further discussion.

Results for other multivariate formulations are scarce, although in some instances the appropriate conditions may be established by reference to the univariate results in Section 3. For instance, for the constant conditional correlations model in equation (6.13), the persistence of a shock to $E_s(\Omega_t)$, and conditions for the model to be covariance stationary are simply determined by the properties of each of the N univariate conditional variance processes; i.e., $E_s(\{\Omega_t\}_{ii})$ $i = 1, \ldots, N$. Similarly, for the factor ARCH model in equation (6.9), stationarity of the model depends directly on the properties of the univariate conditional variance processes for the factor-representing portfolios; i.e. $\{\lambda_{kt}\}$ $k = 1, \ldots, K$.

The empirical estimates for univariate and multivariate ARCH models often indicate a high degree of persistence in the forecast moments for the conditional variances; i.e. $E_s(\sigma_t^2)$ or $E_s(\{\Omega_t\}_{ii})$ $i = 1, \ldots, N$, for $t \to \infty$. At the same time, the commonality in volatility movements suggest that this persistence may be common across different series. More formally, Bollerslev and Engle (1993) define the multivariate ARCH process to be co-persistent in variance if at least one element in $E_s(\Omega_t)$ is non-convergent a.s. for increasing forecast horizons $t - s$, yet there exists a non-trivial linear combination of the process, $\gamma'\varepsilon_t$, such that for every forecast origin s, the forecasts of the corresponding future conditional variances, $E_s(\gamma'\Omega_t\gamma)$, converge to a finite limit independent of time s information. Exact conditions for this to occur within the context of the multivariate GARCH(p, q) model in equation (6.3) are presented in Bollerslev and Engle (1993). These results parallel the conditions for co-integration in the mean as developed by Engle and Granger (1987). Of course, as discussed in Section 3 above, for non-linear models different notions of convergence may give rise to different classifications in terms of the persistence of shocks. The focus on forecast second moments corresponds directly to the mean–variance trade-off relationship often stipulated by economic theory.

To further illustrate this notion of co-persistence, consider the K-factor GARCH(p, q) model defined in equation (6.12). If some of the factor-representing portfolios have persistent variance processes, then individual assets with non-zero factor loadings on such factors will have persistence in variance, also. However, there may be portfolios which have zero factor loadings on these factors. Such portfolios will not have persistence in variance, and hence the assets are co-persistent. This will generally be true if there are more assets than there are persistent factors. From a portfolio selection point of view such portfolios might be desirable as having only transitory fluctuations in variance. Engle and Lee (1993) explicitly test for such an effect between large individual stocks and a market index, but fail to find any evidence of co-persistence.

7. Model selection

Even in linear statistical models, the problem of selecting an appropriate model is non-trivial, to say the least. The usual model selection difficulties are further complicated in ARCH models by the uncountable infinity of functional forms allowed by equation (1.2), and the choice of an appropriate loss function.

Standard model selection criteria such as the Akaike (1973) and the Schwartz (1978) criterion have been widely used in the ARCH literature, though their statistical properties in the ARCH context are unknown. This is particularly true when the validity of the distributional assumptions underlying the likelihood is in doubt.

Most model selection problems focus on estimation of means and evaluate loss functions for alternative models using either in-sample criteria, possibly corrected for fitting by some form of cross-validation, or out-of-sample evaluation. The loss function of choice is typically mean squared error.

When the same strategy is applied to variance estimation, the choice of mean squared error is much less clear. A loss function such as

$$L_1 = \sum_{t=1,T} (\varepsilon_t^2 - \sigma_t^2)^2 \tag{7.1}$$

will penalize conditional variance estimates which are different from the realized squared residuals in a fully symmetrical fashion. However, this loss function does not penalize the method for negative or zero variance estimates which are clearly counterfactual. By this criterion, least squares regressions of squared residuals on past information will have the smallest in-sample loss.

More natural alternatives may be the percentage squared errors,

$$L_2 = \sum_{t=1,T} (\varepsilon_t^2 - \sigma_t^2)^2 \sigma_t^{-4}, \tag{7.2}$$

the percentage absolute errors, or the loss function implicit in the Gaussian likelihood

$$L_3 = \sum_{t=1,T} [\ln(\sigma_t^2) + \varepsilon_t^2 \sigma_t^{-2}]. \tag{7.3}$$

A simple alternative which exaggerates the interest in predicting when residuals are close to zero is[21]

$$L_4 = \sum_{t=1,T} [\ln(\varepsilon_t^2 \sigma_t^{-2})]^2. \tag{7.4}$$

The most natural loss function, however, may be one based upon the goals of the particular application. West et al. (1993) developed such a criterion from the portfolio decisions of a risk averse investor. In an expected utility comparison based on the

[21] Pagan and Schwert (1990) used the loss functions L_1 and L_4 to compare alternative parametric and nonparametric estimators with in-sample and out-of-sample data sets. As discussed in Section 1.5, the L_1 in-sample comparisons favored the nonparametric models, whereas the out-of-sample tests and the loss function L_4 in both cases favored the parametric models.

forecast of the return volatility, ARCH models turn out to fare very well. In a related context, Engle et al. (1993) assumed that the objective was to price options, and developed a loss function from the profitability of a particular trading strategy. They again found that the ARCH variance forecasts were the most profitable.

8. Alternative measures for volatility

Several alternative procedures for measuring the temporal variation in second order moments of time series data have been employed in the literature prior to the development of the ARCH methodology. This is especially true in the analysis of high frequency financial data, where volatility clustering has a long history as a salient empirical regularity.

One commonly employed technique for characterizing the variation in conditional second order moments of asset returns entails the formation of low frequency sample variance estimates based on a time series of high frequency observations. For instance, monthly sample variances are often calculated as the sum of the squared daily returns within the month[22]; examples include Merton (1980) and Poterba and Summers (1986). Of course, if the conditional variances of the daily returns differ within the month, the resulting monthly variance estimates will generally be inefficient; see French et al. (1987) and Chou (1988). However, even if the daily returns are uncorrelated and the variance does not change over the course of the month, this procedure tends to produce both inefficient and biased monthly estimates; see Foster and Nelson (1992).

A related estimator for the variability may be calculated from the inter-period highs and lows. Data on high and low prices within a day is readily available for many financial assets. Intuitively, the higher the variance, the higher the inter-period range. Of course, the exact relationship between the high–low distribution and the variance is necessarily dependent on the underlying distribution of the price process. Using the theory of range statistics, Parkinson (1980) showed that a high–low estimator for the variance of a continuous time random walk is more efficient than the conventional sample variance based on the same number of end-of-interval observations. Of course, the random walk model assumes that the variance remain constant within the sample period. Formal extensions of this idea to models with stochastic volatility are difficult; see also Wiggins (1991), who discusses many of the practical problems, such as sensitivity to data recording errors, involved in applying high–low estimators.

Actively traded options currently exist for a wide variety of financial instruments. A call option gives the holder the right to buy an underlying security at a pre-

[22]Since many high frequency asset prices exhibit low but significant first order serial correlation, two times the first order autocovariance is often added to the daily variance in order to adjust for this serial dependence.

specified price within a given time period. A put option gives the right to sell a security at a pre-specified price. Assuming that the price of the underlying security follows a continuous time random walk, Black and Scholes (1973) derived an arbitrage based pricing formula for the price of a call option. Since the only unknown quantity in this formula is the constant instantaneous variance of the underlying asset price over the life of the option, the option pricing formula may be inverted to infer the conditional variance, or volatility, implicit in the actual market price of the option. This technique is widely used in practice. However, if the conditional variance of the asset is changing through time, the exact arbitrage argument underlying the Black–Scholes formula breaks down. This is consistent with the evidence in Day and Lewis (1992) for stock index options which indicate that a simple GARCH(1, 1) model estimated for the conditional variance of the underlying index return provides statistically significant information in addition to the implied volatility estimates from the Black–Scholes formula. Along these lines Engle and Mustafa (1992) find that during normal market conditions the coefficients in the implied GARCH(1, 1) model which minimize the pricing error for a risk neutral stock option closely resemble the coefficients obtained using more conventional maximum likelihood estimation methods.[23] As mentioned in Section 4 above, much recent research has been directed towards the development of theoretical option pricing formulas in the presence of stochastic volatility; see, for instance, Amin and Ng (1993), Heston (1991), Hull and White (1987), Melino and Turnbull (1990), Scott (1987) and Wiggins (1987). While closed form solutions are only available for a few special cases, it is generally true that the higher the variance of the underlying security, the more valuable the option. Much further research is needed to better understand the practical relevance and quality of the implied volatility estimates from these new theoretical models, however.

Finance theory suggests a close relationship between the volume of trading and the volatility; see Karpoff (1987) for a survey of some of the earlier contributions to this literature. In particular, according to the mixtures of distributions hypothesis, associated with Clark (1973) and Tauchen and Pitts (1983), the evolution of returns and trading volume are both determined by the same latent mixing variable that reflects the amount of new information that arrives at the market. If the news arrival process is serially dependent, volatility and trading volume will be jointly serially correlated. Time series data on trading volume should therefore be useful in inferring the behavior of the second order moments of returns. This idea has been pursued by a number of empirical studies, including Andersen (1992b), Gallant et al. (1992) and Lamoureux and Lastrapes (1990). While the hypothesis that contemporaneous trading volume is positively correlated with financial market volatility is supported

[23]More specifically, Engle and Mustafa (1992) estimate the parameters for the implied GARCH(1, 1) model by minimizing the risk neutral option pricing error defined by the discounted value of the maximum of zero and the simulated future price of the underlying asset from the GARCH(1, 1) model minus the exercise price of the option.

in the data, the result that a single latent variable jointly determines both has been formally rejected by Lamoureux and Lastrapes (1994).

In a related context, modern market micro structure theories also suggest a close relationship between the behavior of price volatility and the distribution of the bid–ask spread though time. Only limited evidence is currently available on the usefulness of such a relationship for the construction of variance estimates for the returns; see, e.g. Bollerslev and Domowitz (1993), Bollerslev and Melvin (1994) and Brock and Kleidon (1992).

The use of the cross sectional variance from survey data to estimate the variance of the underlying time series has been advocated by a number of researchers. Zarnowitz and Lambros (1987) discuss a number of these studies with macroeconomic variables. Of course, the validity of the dispersion across forecasts as a proxy for the variance will depend on the theoretical connection between the degree of heterogeneity and uncertainty; see Pagan et al. (1983). Along these lines it is worth noting that Rich et al. (1992) only find a weak correlation between the dispersion across the forecasts for inflation and an ARCH based estimate for the conditional variance of inflation. The availability of survey data is also likely to limit the practical relevance of this approach in many applications.

In a related context, a number of authors have argued for the use of relative prices or returns across different goods or assets as a way of quantifying inflationary uncertainty or overall market volatility. Obviously, the validity of such cross sectional based measures again hinges on very stringent conditions about the structure of the market; see Pagan et al. (1983).

While all of the variance estimates discussed above may give some idea about the temporal dependencies in second order moments, any subsequent model estimates should be carefully interpreted. Analogously to the problems that arise in the use of generated regressors in the mean, as discussed by Pagan (1984, 1986) and Murphy and Topel (1985), the conventional standard errors for the coefficient estimates in a second stage model that involves a proxy for the variance will have to be adjusted to reflect the approximation error uncertainty. Also, if the conditional mean depends non-trivially on the conditional variance, as in the ARCH–M model discussed in Section 1.4, any two step procedure will generally result in inconsistent parameter estimates; for further analysis along these lines we refer to Pagan and Ullah (1988).

9. Empirical examples

9.1. U.S. Dollar/Deutschmark exchange rates

As noted in Section 1.2, ARCH models have found particularly wide use in the modeling of high frequency speculative prices. In this section we illustrate the empirical quasi-maximum likelihood estimation of a simple GARCH(1, 1) model for a time series of daily exchange rates. Our discussion will be brief. A more detailed and thorough discussion of the empirical specification, estimation and diagnostic testing of ARCH models is given in the next section, which analyzes the time series characteristics of more than one hundred years of daily U.S. stock returns.

The present data set consists of daily observations on the U.S. Dollar/Deutschmark exchange rate over the January 2, 1981 through July 9, 1992 period, for a total of 3006 observations.[24] A broad consensus has emerged that nominal exchange rates over the free float period are best described as non-stationary, or I(1), type processes; see, e.g. Baillie and Bollerslev (1989). We shall therefore concentrate on modeling the nominal percentage returns; i.e. $y_t \equiv 100[\ln(s_t) - \ln(s_{t-1})]$, where s_t denotes the spot Deutschmark/U.S. Dollar exchange rate at day t. This is the time series plotted in Figure 2 in Section 1.2 above. As noted in that section, the daily returns are clearly not homoskedastic, but are characterized by periods of tranquility followed by periods of more turbulent exchange rate movements. At the same time, there appears to be little or no own serial dependence in the levels of the returns. These visual observations are also borne out by more formal tests for serial correlation. For instance, the Ljung and Box (1978) portmanteau test for up to twentieth order serial correlation in y_t equals 19.1, whereas the same test statistic for twentieth order serial correlation in the squared returns, y_t^2, equals 151.9. Under the null of i.i.d. returns, both test statistics should asymptotically be the realization of a chi-square distribution with twenty degrees of freedom. Note that in the presence of ARCH, the portmanteau test for serial correlation in y_t will tend to over-reject.

As discussed above, numerous parametric and nonparametric formulations have been proposed for modeling the volatility clustering phenomenon. For the sake of brevity, we shall here concentrate on the results for the particularly simple MA(1)–GARCH(1, 1) model,

$$\begin{aligned} y_t &= \mu_0 + \theta_1 \varepsilon_{t-1} + \varepsilon_t, \\ \sigma_t^2 &= \omega_0 + \omega_1 W_t - \omega_1(\alpha_1 + \beta_1) W_{t-1} + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned} \tag{9.1}$$

where W_t denotes a weekend dummy equal to one following a closure of the market. The MA(1) term is included to take account of the weak serial dependence in the mean. Following Baillie and Bollerslev (1989), the weekend dummy is entered in the conditional variance to allow for an impulse effect.

The quasi-maximum likelihood estimates (QMLE) for this model, obtained by the numerical maximization of the normal likelihood function defined by equations (2.7), (2.8) and (2.12), are contained in Table 1. The first column in the table shows that the α_1 and β_1 coefficients are both highly significant at the conventional five percent level. The sum of the estimated GARCH parameters also indicates a fairly strong degree of persistence in the conditional variance process.[25]

[24] The rates were calculated from the ECU cross rates obtained through Datastream.

[25] Reparametrizing the conditional variance in terms of $(\alpha_1 + \beta_1)$ and α_1, the t-test statistic for the null hypothesis that $\alpha_1 + \beta_1 = 1$ equals 3.784, thus formally rejecting the IGARCH(1, 1) model at standard significance levels.

Table 1
Quasi-maximum likelihood estimates.[a]

Coefficient	Jan. 2, 1982 July 9, 1992	Jan. 2, 1982 Oct. 6, 1986	Oct. 7, 1986 July 9, 1992
μ_0	−0.002	0.014	−0.017
	(0.009)	(0.018)	(0.016)
	[0.009]	[0.018]	[0.016]
	{0.009}	{0.018}	{0.017}
θ_1	−0.056	−0.058	0.055
	(0.014)	(0.030)	(0.027)
	[0.013]	[0.027]	[0.027]
	{0.013}	{0.027}	{0.027}
ω_0	0.028	0.024	0.035
	(0.005)	(0.009)	(0.011)
	[0.004]	[0.007]	[0.011]
	{0.003}	{0.006}	{0.010}
ω_1	0.243	0.197	0.281
	(0.045)	(0.087)	(0.087)
	[0.031]	[0.062]	[0.061]
	{0.022}	{0.046}	{0.042}
α_1	0.068	0.076	0.063
	(0.009)	(0.022)	(0.017)
	[0.007]	[0.014]	[0.014]
	{0.005}	{0.010}	{0.011}
β_1	0.880	0.885	0.861
	(0.015)	(0.028)	(0.033)
	[0.012]	[0.020]	[0.031]
	{0.010}	{0.016}	{0.030}

[a]Robust standard errors based on equation (2.21) are reported in parentheses, $(\cdot)$. Standard errors calculated from the Hessian in equation (2.18) are reported in $[\cdot]$. Standard errors from on the outer product of the sample gradients in (2.19) are given in $\{\cdot\}$.

Consistent with the stylized facts discussed in Section 1.2.4, the conditional variance is also significantly higher following non-trading periods.

The second and third columns of Table 1 report the results with the same model estimated for the first and second half of the sample respectively; i.e. January 2, 1981 through October 6, 1986 and October 7, 1986 through July 9, 1992. The parameter estimates are remarkably similar across the two sub-periods.[26]

In summary, the simple model in equation (9.1) does a remarkably good job of capturing the own temporal dependencies in the volatility of the exchange rate series. For instance, the highly significant portmanteau test for serial correlation in

[26]Even though the assumption of conditional normality is violated empirically, it is interesting to note that the sum of the maximized normal quasi log likelihoods for the two sub-samples equals $-1727.750 - 1597.166 = -3324.916$, compared to -3328.984 for the model estimated over the full sample.

the squares of the raw series, y_t^2, drops to only 21.687 for the squared standardized residuals, $\hat{\varepsilon}_t\hat{\sigma}_t^{-2}$. We defer our discussion of other residual based diagnostics to the empirical example in the next section.

While the GARCH(1, 1) model is able to track the own temporal dependencies, the assumption of conditionally normally distributed innovations is clearly violated by the data. The sample skewness and kurtosis for $\hat{\varepsilon}_t\hat{\sigma}_t^{-1}$ equal -0.071 and 4.892, respectively. Under the null of i.i.d. normally distributed standardized residuals, the sample skewness should be the realization of a normal distribution with a mean of 0 and a variance of $6/\sqrt{3005} = 0.109$, while the sample kurtosis is asymptotically normally distributed with a mean of 3 and a variance of $24/\sqrt{3005} = 0.438$.

The standard errors for the quasi-maximum likelihood estimates reported in $(\cdot)$ in Table 1 are based on the asymptotic covariance matrix estimator discussed in Section 2.3. These estimates are robust to the presence of conditional excess kurtosis. The standard errors reported in $[\cdot]$ and $\{\cdot\}$ are calculated from the Hessian and the outer product of the gradients as in equations (2.18) and (2.19), respectively. For some of the conditional variance parameters, the non-robust standard errors are less than one half of their robust counterparts. This compares to the findings reported in Bollerslev and Wooldridge (1992), and highlights the importance of appropriately accounting for any conditional non-normality when conducting inference in ARCH type models based on a normal quasi-likelihood function.

9.2. *U.S. stock prices*

We next turn to modeling heteroskedasticity in U.S. stock index returns data. Drawing on the optimal filtering results of Nelson and Foster (1991, 1994) summarized in Section 4, as a guidance in model selection, new very rich parametrizations are introduced.

From 1885 on, the Dow Jones corporation has published various stock indices daily. In 1928, the Standard Statistics company began publishing daily a wider index of 90 utility, industrial and railroad stocks. In 1953, the Standard 90 index was replaced by an even broader index, the Standard and Poor's 500 composite. The properties of these indices are considered in some detail in Schwert (1990).[27] The Dow data has one substantial chronological break, from July 30, 1914, through December 11, 1914, when the financial markets were closed following the outbreak of the First World War. The first data set we analyze is the Dow data from its inception on February 16, 1885 until the market closure in 1914. The second data set is the Dow data from the December 1914 market reopening until January 3, 1928. The third data set is the Standard 90 capital gains series beginning in January 4, 1928 and extending to the end of May 1952. The Standard 90 index data is

[27] G. William Schwert kindly provided the data. Schwert's indices differ from ours aftér 1962, when he uses the CRSP value weighted market index. We continue to use the S&P 500 through 1990.

available through the end of 1956, but we end at the earlier date because that is when the New York Stock Exchange ended its Saturday trading session, which presumably shifted volatility to other days of the week. The final data set is the S&P 500 index beginning in January 1953 and continuing through the end of 1990.

9.2.1. Model specification

Our basic capital gains series, r_t, is derived from the price index data, P_t, as

$$r_t \equiv 100 \ln[P_t/P_{t-1}]. \tag{9.2}$$

Thus, r_t corresponds to the continuously compounded capital gain on the index measured in percent. Any ARCH formulation for r_t may be compactly written as

$$r_t = \mu(r_{t-1}, \sigma_t^2) + \varepsilon_t \tag{9.3}$$

and

$$\varepsilon_t = z_t \cdot \sigma_t, \quad z_t \sim \text{i.i.d.}, \quad E[z_t] = 0, \quad E[z_t^2] = 1, \tag{9.4}$$

where $\mu(r_{t-1}, \sigma_t^2)$ and σ_t denote the conditional mean and the conditional standard deviation, respectively.

In the estimation reported below we parametrized the functional form for the conditional mean by

$$\mu(r_{t-1}, \sigma_t^2) \equiv \mu_0 + r_{t-1}[\mu_1 + \mu_2 \exp(-\sigma_t^2/u^2)] + \mu_3 \sigma_t^2. \tag{9.5}$$

This is very close to the specification in LeBaron (1992). The μ_1 coefficient allows for first order autocorrelation. The u^2 term denotes the sample mean of r_t^2, which is essentially equal to the unconditional sample variance of r_t. As noted by LeBaron (1992), serial correlation seems to be a decreasing function of the conditional variance, which may be captured by equation (9.5) through $\mu_2 > 0$. The parameter μ_3 is an ARCH–M term.

We assume that the conditional distribution of ε_t given σ_t is generalized t; see, e.g. McDonald and Newey (1988). The density for the generalized t-distribution takes the form

$$f[\varepsilon_t \sigma_t^{-1}; \eta, \psi] = \frac{\eta}{2\sigma_t b \cdot \psi^{1/\eta} B(1/\eta, \psi) \cdot [1 + |\varepsilon_t|^\eta / (\psi b^\eta \sigma_t^\eta)]^{\psi + 1/\eta}} \tag{9.6}$$

where $B(1/\eta, \psi) \equiv \Gamma(1/\eta)\Gamma(\psi)\Gamma(1/\eta + \psi)$ denotes the beta function, $b \equiv [\Gamma(\psi)\Gamma(1/\eta)/\Gamma(3/\eta)\Gamma(\psi - 2/\eta)]^{1/2}$, and $\psi\eta > 2$, $\eta > 0$ and $\psi > 0$. The scale factor b makes $\mathrm{Var}(\varepsilon_t \sigma_t^{-1}) = 1$.

One advantage of this specification is that it nests both the Student's t and the GED distributions discussed in Section 2.2 above. In particular, the Student's t-distribution sets $\eta = 2$ and $\psi = \frac{1}{2}$ times the degrees of freedom. The GED is obtained for $\psi = \infty$. Nelson (1989, 1991) fit EGARCH models to U.S. Stock index returns assuming a GED conditional distribution, and found that there were many more large standardized residuals $z_t \equiv \varepsilon_t \sigma_t^{-1}$ than would be expected if the returns were actually conditionally GED with the estimated η. The GED has only one "shape" parameter η, which is apparently insufficient to fit both the central part and the tails of the conditional distribution. The generalized t-distribution has two shape parameters, and may therefore be more successful in parametrically fitting the conditional distribution.

The conditional variance function, σ_t^2, is parametrized using a variant of the EGARCH formulation in equation (1.11),

$$\ln(\sigma_t^2) = \omega_t + \frac{(1 + \alpha_1 L + \cdots + \alpha_q L^q)}{(1 - \beta_1 L - \cdots - \beta_p L^p)} g(z_{t-1}, \sigma_{t-1}^2), \tag{9.7}$$

where the deterministic component is given by

$$\omega_t \equiv \omega_0 + \ln[1 + \omega_1 W_t + \omega_2 S_t + \omega_3 H_t]. \tag{9.8}$$

As noted in Section 1.2, trading and non-trading periods contribute differently to volatility. To also allow for differences between weekend and holiday non-trading periods W_t gives the number of weekend non-trading days between trading days t and $t-1$, while H_t denotes the number of holidays. Prior to May 1952, the NYSE was open for a short trading session on Saturday. Since Saturday may have been a "slow" news day and the Saturday trading session was short, we would expect low average volatility on Saturdays. The S_t dummy variable equals one if trading day t is a Saturday and zero otherwise.

Our specification of the news impact function, $g(\cdot,\cdot)$, is a generalization of EGARCH inspired by the optimal filtering results of Nelson and Foster (1994). In the EGARCH model in equation (1.11) $\ln(\sigma_{t+1}^2)$ is homoskedastic conditional on σ_t^2, and the partial correlation between z_t and $\ln(\sigma_{t+1}^2)$ is constant conditional on σ_t^2. These assumptions may well be too restrictive, and the optimal filtering results indicate the importance of correctly specifying these moments. Our specification of $g(z_t, \sigma_t^2)$ therefore allows both moments to vary with the level of σ_t^2.

Several recent papers, including Engle and Ng (1993), have suggested that GARCH, EGARCH and similar formulations may make σ_t^2 or $\ln(\sigma_t^2)$ too sensitive to outliers. The optimal filtering results discussed in Section 4 lead to the same conclusion when ε_t is drawn from a conditionally heavy tailed distribution. The final form that we assume for $g(\cdot,\cdot)$ was also motivated by this observation:

$$g(z_t, \sigma_t^2) \equiv \sigma_t^{-2\theta_0} \frac{\theta_1 z_t}{1 + \theta_2 |z_t|} + \sigma_t^{-2\gamma_0} \left[\frac{\gamma_1 |z_t|^\rho}{1 + \gamma_2 |z_t|^\rho} - E_t\left(\frac{\gamma_1 |z_t|^\rho}{1 + \gamma_2 |z_t|^\rho} \right) \right]. \tag{9.9}$$

The γ_0 and θ_0 parameters allow both the conditional variance of $\ln(\sigma_{t+1}^2)$ and its conditional correlation with z_t to vary with the level of σ_t^2. If $\theta_1 < 0, \ln(\sigma_{t+1}^2)$ and z_t are negatively correlated: the "leverage effect". The EGARCH model constrains $\theta_0 = \gamma_0 = 0$, so that the conditional correlation is constant, as is the conditional variance of $\ln(\sigma_t^2)$. The ρ, γ_2, and θ_2 parameters give the model flexibility in how much weight to assign to the tail observations. For example, if γ_2 and θ_2 are both positive, the model downweights large $|z_t|$'s. The second term on the right hand side of equation (9.9) was motivated by the optimal filtering results in Nelson and Foster (1994), designed to make the ARCH model serve as a robust filter.

The orders of the ARMA model for $\ln(\sigma_t^2)$, p and q, remain to be determined. Table 2 gives the maximized values of the log likelihoods from (2.7), (2.8) and (9.6) for ARMA models of order up to ARMA(3, 5). For three of the four data sets, the information criterion of Schwartz (1978) selects an ARMA(2, 1) model, the exception being the Dow data for 1914–1928, for which an AR(1) is selected. For linear time series models, the Schwartz criterion has been shown to consistently estimate the order of an ARMA model. As noted in Section 7, it is not known whether this result carries over to the ARCH class of models. However, guided by the results in Table 2,

Table 2
Log likelihood values for fitted models.[a]

Fitted model	Dow 1885–1914	Dow 1914–1928	Standard 90 1928–1952	S&P 500 1953–1990
White Noise	−10036.188	−4397.693	−11110.120	−10717.199
MA(1)	−9926.781	−4272.639	−10973.417	−10658.775
MA(2)	−9848.319	−4241.686	−10834.937	−10596.849
MA(3)	−9779.491	−4233.371	−10765.259	−10529.688
MA(4)	−9750.417	−4214.821	−10740.999	−10463.534
MA(5)	−9718.642	−4198.672	−10634.429	−10433.631
AR(1)	−9554.352	−4164.093[SC]	−10275.294	−10091.450
ARMA(1, 1)	−9553.891	−4164.081	−10269.771	−10076.775
ARMA(1, 2)	−9553.590	−4160.671	−10265.464	−10071.040
ARMA(1,3)	−9552.148	−4159.413	−10253.027	−10070.587
ARMA(1,4)	−9543.855	−4158.836	−10250.446	−10064.695
ARMA(1,5)	−9540.485	−4158.179	−10242.833	−10060.336
AR(2)	−9553.939	−4164.086	−10271.732	−10083.442
ARMA(2, 1)	−9529.904[SC]	−4159.011[AIC]	−10237.527[SC]	−10052.322[SC]
ARMA(2, 2)	−9529.642	−4158.428	−10235.724	−10049.237
ARMA(2, 3)	−9526.865	−4157.731	−10234.556	−10049.129
ARMA(2, 4)	−9525.683	−4157.569	−10234.429	−10047.962
ARMA(2, 5)	−9525.560	−4155.071	−10230.418	−10046.343
AR(3)	−9553.787	−4159.227	−10270.685	−10075.441
ARMA(3, 1)	−9529.410	−4158.608	−10237.462	−10049.833
ARMA(3,2)	−9526.089	−4158.230	−10228.701[AIC]	−10049.044
ARMA(3, 3)	−9524.644[AIC]	−4157.730	−10228.263	−10042.710
ARMA(3, 4)	−9524.497	−4156.823	−10227.982	−10042.284
ARMA(3, 5)	−9523.375	−4154.906	−10227.958	−10040.547[AIC]

[a]The AIC and SC indicators denote the models selected by the information criteria of Akaike (1973) and Schwartz (1978), respectively.

Table 3
Maximum likelihood estimates.[a]

Coefficient	Dow 1885–1914 ARMA(2, 1)	Dow 1914–1928 AR(1)	Standard 90 1928–1952 ARMA(2, 1)	S&P 500 1953–1990 ARMA(2, 1)
ω_0	−0.6682 (0.1251)	−0.6228 (0.0703)	−1.2704 (2.5894)	−0.7899 (0.2628)
ω_1	0.2013 (0.0520)	0.3059 (0.0904)	0.1011 (0.0518)	0.1286 (0.0295)
ω_2	−0.4416 (0.0270)	−0.5557 (0.0328)	−0.6534 (0.0211)	*
ω_3	0.5099 (0.1554)	0.3106 (0.1776)	0.6609 (0.1702)	0.1988 (0.1160)
ψ	3.6032 (0.8019)	2.5316 (0.5840)	4.0436 (0.9362)	3.5437 (0.7557)
η	2.2198 (0.1338)	2.4314 (0.2041)	1.7809 (0.1143)	2.1844 (0.1215)
μ_0	0.0280 (0.0112)	0.0642 (0.0222)	0.0725 (0.1139)	0.0259 (0.0113)
μ_1	−0.0885 (0.0270)	−0.0920 (0.0418)	−0.0914 (0.0243)	0.0717 (0.0260)
μ_2	0.2206 (0.0571)	0.3710 (0.0828)	0.2990 (0.0387)	0.2163 (0.0532)
μ_3	0.0006 (0.0209)	0.0316 (0.0442)	0.0285 (0.0102)	0.0050 (0.0213)
γ_0	−0.1058 (0.0905)	0.0232 (0.1824)	−0.0508 (0.0687)	0.1117 (0.0908)
γ_1	0.1122 (0.0256)	0.0448 (0.0478)	0.1356 (0.0327)	0.0658 (0.0157)
γ_2	0.0245 (0.0178)	0.0356 (0.0316)	0.0168 (0.0236)	0.0312 (0.0080)
ρ	2.1663 (0.3119)	3.2408 (1.5642)	1.6881 (0.3755)	2.2477 (0.3312)
θ_0	−0.6097 (0.0758)	−0.5675 (0.1232)	−0.1959 (0.0948)	−0.1970 (0.1820)
θ_1	−0.1509 (0.0258)	−0.3925 (0.1403)	−0.1177 (0.0271)	−0.1857 (0.0287)
θ_2	0.0361 (0.0828)	0.3735 (0.3787)	−0.0055 (0.0844)	0.2286 (0.1241)
Δ_1	0.9942 (0.0033)	0.9093 (0.0172)	0.9994 (0.0009)	0.9979 (0.0011)
Δ_2	0.8759 (0.0225)	*	0.8303 (0.0282)	0.8945 (0.0258)
α_1	−0.9658 (0.0148)	*	−0.9511 (0.0124)	−0.9695 (0.0010)

[a] Standard errors are reported in parentheses. The parameters indicated by a * were not estimated. The AR coefficients are decomposed as $(1-\Delta_1 L)(1-\Delta_2 L) \equiv (1-\beta_1 L-\beta_2 L^2)$, where $|\Delta_1| \geqslant |\Delta_2|$.

Table 4
Wald hypothesis tests.

Test	Dow 1885–1914 ARMA(2, 1)	Dow 1914–1928 AR(1)	Standard 90 1928–1952 ARMA(2, 1)	S&P 500 1953–1990 ARMA(2, 1)
$\gamma_2 = \theta_2 = \gamma_0 = \theta_0 = \rho - 1 = 0$: χ^2_5	97.3825 (0.0000)	63.4545 (0.0000)	10.1816 (0.0703)	51.8152 (0.0000)
$\omega_1 = \omega_3$: χ^2_1	3.3867 (0.0657)	0.0006 (0.9812)	9.8593 (0.0017)	0.3235 (0.5695)
$\theta_0 = \gamma_0 = 0$: χ^2_2	67.4221 (0.0000)	21.3146 (0.0000)	4.4853 (0.1062)	2.2024 (0.3325)
$\theta_0 = \gamma_0$: χ^2_1	17.2288 (0.0000)	7.4328 (0.0064)	1.7718 (0.1832)	1.7844 (0.1816)
$\eta = \rho$: χ^2_1	0.0247 (0.8751)	0.2684 (0.6044)	0.0554 (0.8139)	0.0312 (0.8598)
$\gamma_2 = b^{-\eta}\psi^{-1}$: χ^2_1	14.0804 (0.0002)	10.0329 (0.0015)	14.1293 (0.0002)	14.6436 (0.0001)
$\eta = \rho, \gamma_2 = b^{-\eta}\psi^{-1}$:$\chi^2_2$	18.4200 (0.0001)	10.4813 (0.0053)	22.5829 (0.0000)	16.9047 (0.0002)

Table 5
Conditional moment specification tests.

Orthogonality Condition	Dow 1885–1914 ARMA(2, 1)	Dow 1914–1928 AR(1)	Standard 90 1928–1952 ARMA(2, 1)	S&P 500 1953–1990 ARMA(2, 1)
(1) $E_t[z_t] = 0$	−0.0147 (0.0208)	−0.0243 (0.0319)	−0.0275 (0.0223)	−0.0110 (0.0202)
(2) $E_t[z_t^2] = 1$	0.0007 (0.0382)	0.0007 (0.0613)	0.0083 (0.0503)	0.0183 (0.0469)
(3) $E_t[z_t \cdot \lvert z_t \rvert] = 0$	−0.0823 (0.0365)	−0.1122 (0.0564)	−0.1072 (0.0414)	−0.0658 (0.0410)
(4) $E_t[g(z_t, \sigma_t)] = 0$	0.0007 (0.0046)	0.0013 (0.0080)	0.0036 (0.0051)	0.0003 (0.0035)
(5) $E_t[(z_t^2 - 1)(z_{t-1}^2 - 1)] = 0$	−0.0050 (0.0714)	−0.0507 (0.0695)	−0.0105 (0.0698)	0.1152 (0.0930)
(6) $E_t[(z_t^2 - 1)(z_{t-2}^2 - 1)] = 0$	−0.0047 (0.0471)	0.0399 (0.0606)	−0.0358 (0.0815)	−0.0627 (0.0458)
(7) $E_t[(z_t^2 - 1)(z_{t-3}^2 - 1)] = 0$	0.0037 (0.0385)	−0.0365 (0.0521)	0.0373 (0.0583)	−0.0171 (0.0611)
(8) $E_t[(z_t^2 - 1)(z_{t-4}^2 - 1)] = 0$	0.0950 (0.0562)	−0.0658 (0.0403)	−0.0018 (0.0543)	−0.0312 (0.0426)
(9) $E_t[(z_t^2 - 1)(z_{t-5}^2 - 1)] = 0$	0.0165 (0.0548)	0.0195 (0.0486)	0.0710 (0.0565)	0.0261 (0.0731)
(10) $E_t[(z_t^2 - 1)(z_{t-6}^2 - 1)] = 0$	−0.0039 (0.0309)	0.0343 (0.0602)	0.0046 (0.0439)	−0.0557 (0.0392)

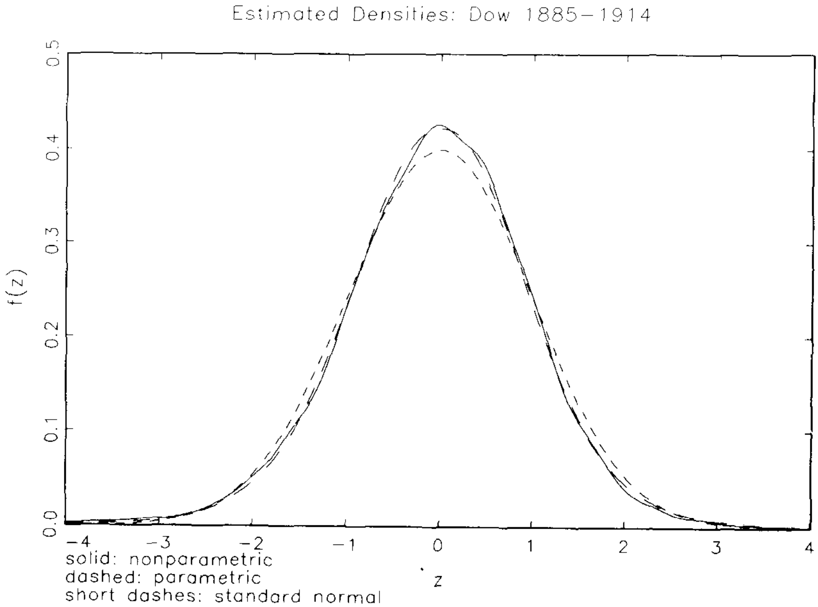
Table 5 (*continued*)

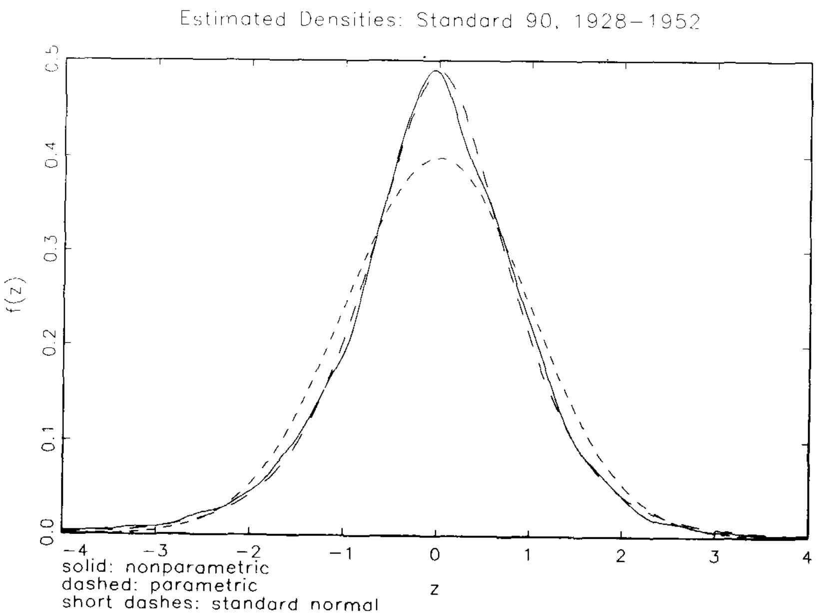
Orthogonality Condition	Dow 1885–1914 ARMA(2, 1)	Dow 1914–1928 AR(1)	Standard 90 1928–1952 ARMA(2, 1)	S&P 500 1953–1990 ARMA(2, 1)
(11) $E_t[(z_t^2-1)z_{t-1}]=0$	−0.0338 (0.0290)	−0.0364 (0.0414)	−0.0253 (0.0367)	−0.0203 (0.0413)
(12) $E_t[(z_t^2-1)z_{t-2}]=0$	0.0069 (0.0251)	−0.0275 (0.0395)	−0.0434 (0.0315)	−0.0378 (0.0278)
(13) $E_t[(z_t^2-1)z_{t-3}]=0$	0.0110 (0.0262)	0.0290 (0.0352)	0.0075 (0.0306)	0.0292 (0.0357)
(14) $E_t[(z_t^2-1)z_{t-4}]=0$	−0.0296 (0.0275)	0.0530 (0.0340)	−0.0103 (0.0292)	−0.0137 (0.0238)
(15) $E_t[(z_t^2-1)z_{t-5}]=0$	−0.0094 (0.0240)	0.0567 (0.0342)	0.0153 (0.0287)	0.0064 (0.0238)
(16) $E_t[(z_t^2-1)z_{t-6}]=0$	0.0281 (0.0216)	0.0038 (0.0350)	−0.0170 (0.0253)	0.0417 (0.0326)
(17) $E_t[z_t\cdot z_{t-1}]=0$	0.0265 (0.0236)	0.0127 (0.0346)	0.0383 (0.0243)	0.0188 (0.0226)
(18) $E_t[z_t\cdot z_{t-2}]=0$	0.0133 (0.0157)	−0.0176 (0.0283)	−0.0445 (0.0174)	−0.0434 (0.0158)
(19) $E_t[z_t\cdot z_{t-3}]=0$	0.0406 (0.0158)	0.0012 (0.0262)	0.0019 (0.0175)	0.0140 (0.0152)
(20) $E_t[z_t\cdot z_{t-4}]=0$	0.0580 (0.0161)	0.0056 (0.0253)	0.0211 (0.0172)	0.0169 (0.0153)
(21) $E_t[z_t\cdot z_{t-5}]=0$	0.0516 (0.0163)	0.0164 (0.0251)	0.0250 (0.0174)	0.0121 (0.0158)
(22) $E_t[z_t\cdot z_{t-6}]=0$	−0.0027 (0.0158)	0.0081 (0.0261)	−0.0040 (0.0172)	−0.0211 (0.0150)
(1)–(16): χ^2_{16}	39.1111 (0.0010)	45.1608 (0.0001)	31.7033 (0.011)	25.1116 (0.0679)
(1)–(22): χ^2_{22}	94.0156 (0.0000)	52.1272 (0.0003)	67.1231 (0.0000)	63.6383 (0.0000)

Table 3 reports the maximum likelihood estimates (MLE) for the models selected by the Schwartz criterion. Various Wald and conditional moment specification tests are given in Tables 4 and 5.

9.2.2. *Persistence of shocks to volatility*

As in Nelson (1989, 1991), the ARMA(2, 1) models selected for three of the four data sets can be decomposed into the product of two AR(1) components, one of which has very long-lived shocks, with an AR root very close to one, the other of which exhibits short-lived shocks, with an AR root very far from one; i.e. $(1-\beta_1L-\beta_2L^2)\equiv$


Estimated Densities: Dow 1885–1914
f(z)
0.0
0.1
0.2
0.3
0.4
0.5
−4
−3
−2
−1
0
1
2
3
4
solid: nonparametric
dashed: parametric
short dashes: standard normal
z


Estimated Densities: Standard 90, 1928–1952
f(z)
0.0
0.1
0.2
0.3
0.4
0.5
−4
−3
−2
−1
0
1
2
3
4
solid: nonparametric
dashed: parametric
short dashes: standard normal
z

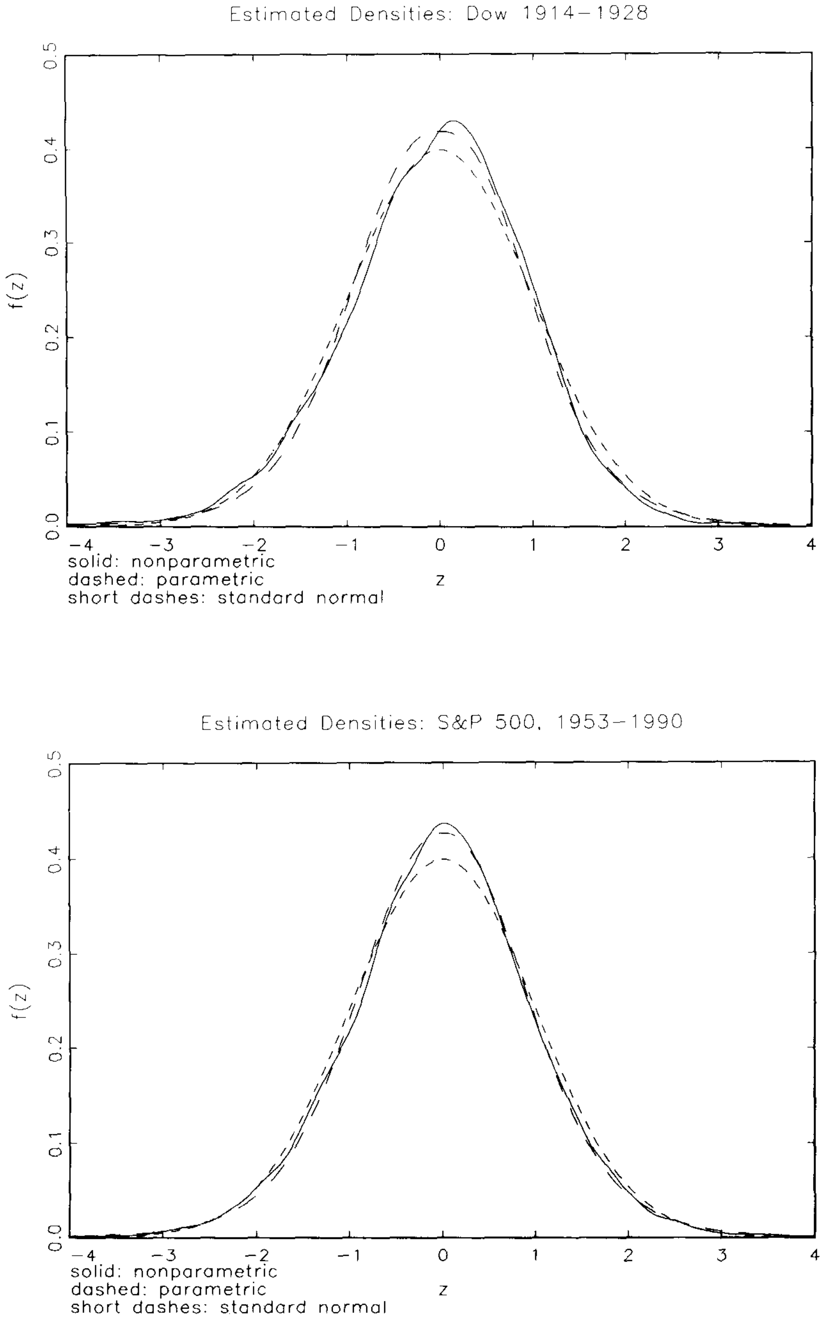


Figure 3. Conditional Distribution of Returns

$(1-\Delta_1 L)(1-\Delta_2 L)$, where $|\Delta_1| \geqslant |\Delta_2|$. When the estimated AR roots are real, a useful gauge of the persistence of shocks in an AR(1) model is the estimated "half life"; that is the value of n for which $\Delta^n = \frac{1}{2}$. For the Dow 1885–1914, the Standard 90 and the S&P 500 the estimated half lives of the long-lived components are about 119 days, $4\frac{1}{2}$ years and 329 days respectively. The corresponding estimated half lives of the short-lived components are only 5.2, 3.7 and 6.2 days, respectively.[28,29]

9.2.3. *Conditional mean of returns*

The estimated μ_i terms strongly support the results of LeBaron (1992) of a negative relationship between the conditional variance and the conditional serial correlation in returns. In particular, μ_2 is significantly positive in each data set, both statistically and economically. For example, for the Standard 90 data, the fitted conditional first order correlation in returns is 0.17 when σ_t^2 is at the 10th percentile of its fitted sample values, but equals -0.07 when σ_t^2 is at the 90th percentile. The implied variation in returns serial correlation is similar in the other data sets. The relatively simple specification of $\mu(r_{t-1}, \sigma_t^2)$ remains inadequate, however, as can be seen from the conditional moment tests reported in Table 5. The 17th through 22nd conditions test for serial correlation in the fitted z_t's at lags one through six. In each data set, significant serial correlation is found at the higher lags.

9.2.4. *Conditional distribution of returns*

Figure 3 plots the fitted generalized t density of the z_t's against both a standard normal and a nonparametric density estimate constructed from the fitted z_t's using a Gaussian kernel with the bandwidth selection method of Silverman (1986, pp. 45–48). The parametric and nonparametric densities appear quite close, with the exception of the Dow 1914–1928 data, which exhibits strong negative skewness in $\hat{z}_t$. Further aspects of the fitted conditional distribution are checked in the first three conditional moment specification tests reported in Table 5. These three orthogonality conditions test that the standardized residuals $\hat{z}_t \equiv \hat{\varepsilon}_t \hat{\sigma}_t^{-1}$ have mean zero, unit variance, and no skewness.[30] In the first three data sets the $\hat{z}_t$ series exhibit statistically significant, though not overwhelmingly so, negative skewness.

[28] This is consistent with recent work by Ding et al. (1993), in which the empirical autocorrelations of absolute returns from several financial data sets are found to exhibit rapid decay at short lags but much slower decay at longer lags. This is also the motivation behind the permanent/transitory components ARCH model introduced by Engle and Lee (1992, 1993), and the fractionally integrated ARCH models recently proposed by Baillie et al. (1993).

[29] Volatility in the Dow 1914–1928 data shows much less persistence. The half life asociated with the AR(1) model selected by the Schwartz (1978) criterion is only about 7.3 days. For the ARMA(2,1) model selected by the AIC for this data set, the half lives associated with the two AR roots are only 24 and 3.3 days, respectively.

[30] More precisely, the third orthogonality condition tests that $E_t[z_t \cdot |z_t|] = 0$ rather than $E_t[z_t^3] = 0$. We use this test because it requires only the existence of a fourth conditional moment for z_t rather than a sixth conditional moment.

Table 6
Frequency of tail events.[a]

	Dow 1885–1914 ARMA(2, 1)		Dow 1914–1928 AR(1)		Standard 90 1928–1952 ARMA(2, 1)		S&P 500 1953–1990 ARMA(2, 1)	
N	Expected	Actual	Expected	Actual	Expected	Actual	Expected	Actual
2	421.16	405	180.92	177	369.89	363	458.85	432
3	63.71	74	31.11	33	76.51	81	72.60	57
4	11.54	12	6.99	10	18.76	23	13.83	14
5	2.61	4	2.01	3	5.47	4	3.27	6
6	0.72	2	0.70	1	1.86	1	0.94	5
7	0.23	1	0.28	0	0.71	1	0.31	3
8	9.56×10^{-6}	0	0.13	0	0.30	1	0.12	2
9	3.89×10^{-7}	0	0.06	0	0.14	0	0.05	2
10	1.73×10^{-7}	0	0.03	0	0.07	0	0.02	2
11	8.25×10^{-8}	0	0.01	0	0.04	0	0.01	1

[a]The table reports the expected and the actual number of observations exceeding N conditional standard deviations.

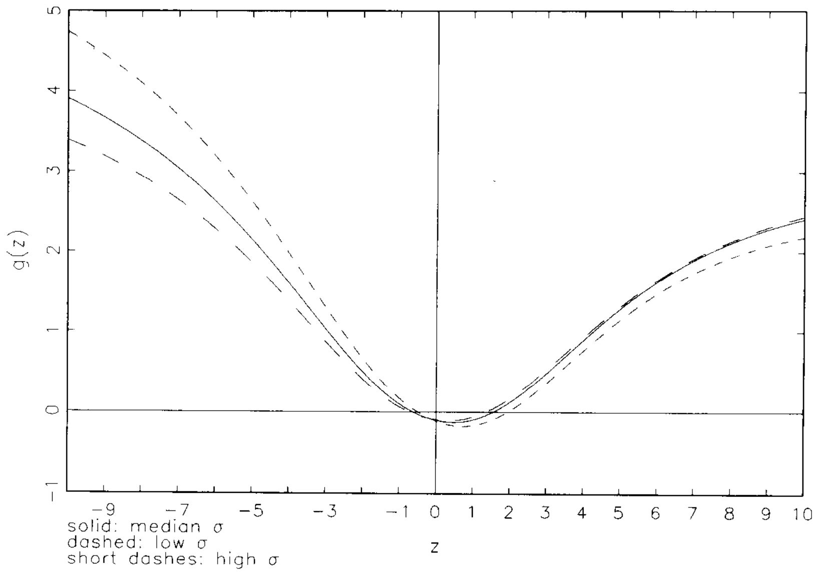
The original motivation for adopting the generalized t-distribution was that the two shape parameters η and ψ would allow the model to fit both the tails and the central part of the conditional distribution. Table 6 gives the expected and the actual number of z_t's in each data set exceeding N standard deviations. In the S&P 500 data, the number of outliers is still too large. In the other data sets, the tail fit seems adequate.

As noted above, the generalized t-distribution nests both the Student's t ($\eta = 2$) and the GED ($\psi = \infty$). Interestingly, in only two of the data sets does a t-test for the null hypothesis that $\eta = 2$ reject at standard levels, and then only marginally. Thus, the improved fit appears to come from the t component rather than the GED component of the generalized t-distribution. In total, the generalized t-distribution is a marked improvement over the GED, though perhaps not over the usual Student's t-distribution. Nevertheless, the generalized t is not entirely adequate, as it does not account for the fairly small skewness in the fitted z_t's, and also appears to have insufficiently thick tails for the S&P 500 data.

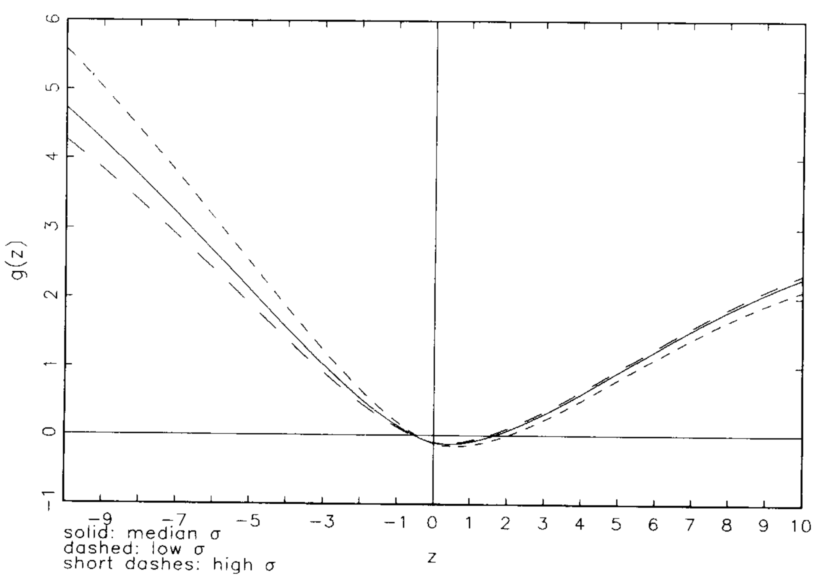
9.2.5. *News impact function*

In line with the results for the EGARCH model reported in Nelson (1989, 1991), the "leverage effect" term θ_1 in the $g(\cdot, \cdot)$ function is significantly negative in each of the data sets, while the "magnitude effect" term γ_1 is always positive, and significantly so except in the Dow 1914–1928 data. There are important differences, however. The EGARCH parameter restrictions that $\rho = 1$, $\gamma_0 = \gamma_2 = \theta_0 = \theta_2 = 0$ are decisively rejected in three of the four data sets. The estimated $g(z_t, \sigma_t^2)$ functions are plotted in Figure 4, from which the differences with the piecewise linear EGARCH $g(z_t)$ formulation are apparent.

g(z): Dow 1885–1914



g(z): Standard 90, 1928–1952



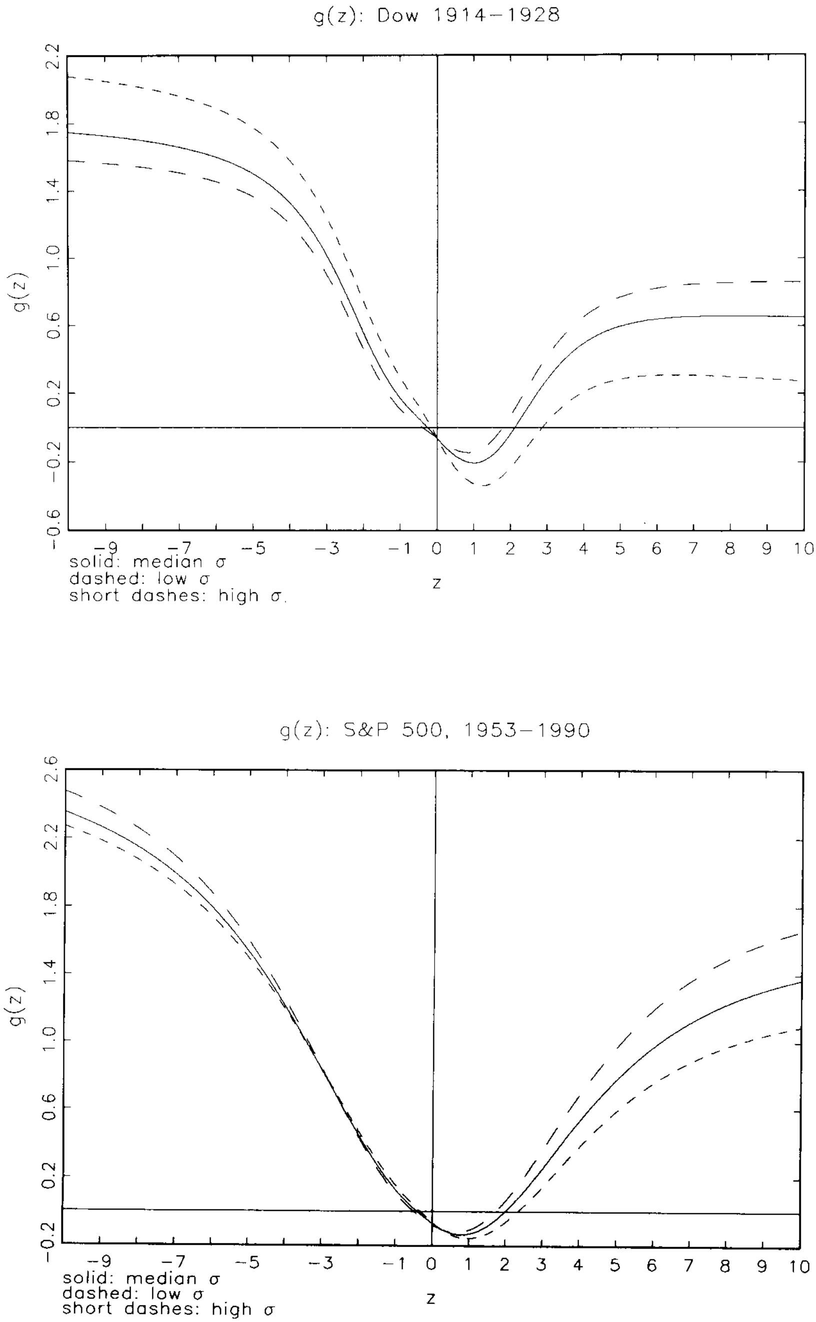


Figure 4. News Impact Functions

To better understand why the standard EGARCH model is rejected, consider more closely the differences between the specification of the $g(z_t, \sigma_t^2)$ function in equation (9.9) and the EGARCH formulation in equation (1.11). Firstly, the parameters γ_0 and θ_0 allow the conditional variance of $\ln(\sigma_t^2)$ and the conditional correlation between $\ln(\sigma_t^2)$ and r_t to change as functions of σ_t^2. Secondly, the parameters ρ, γ_2, and θ_2 give the model an added flexibility in how much weight to assign to large versus small values of z_t.

As reported in Table 4, the EGARCH assumption that $\gamma_0 = \theta_0 = 0$ is decisively rejected in the Dow 1885–1914 and 1914–1928 data sets, but not for either the Standard 90 or the S&P 500 data sets. For none of the four data sets is the estimated value of γ_0 significantly different from 0 at conventional levels. The estimated value of θ_0 is always negative, however, and very significantly so in the first two data sets, indicating that the "leverage effect" is more important in periods of high volatility than in periods of low volatility.

The intuition that the influence of large outliers should be limited by setting $\theta_2 > 0$ and $\gamma_2 > 0$ receives mixed support from the data. The estimated values of γ_2 and three of the estimated θ_2's are positive, but only the estimate of γ_2 for the S&P 500 data is significantly positive at standard levels. We also note that if the data is generated by a stochastic volatility model, as opposed to an ARCH model, with conditionally generalized t-distributed errors, the asymptotically optimal ARCH filter would set $\eta = \rho$ and $\gamma_2 = \psi^{-1} b^{-\eta}$. The results in Table 4 indicate that the $\eta = \rho$ restriction is not rejected, but that $\gamma_2 = \psi^{-1} b^{-\eta}$ is not supported by the data. The estimated values of γ_2 are "too low" relative to the asymptotically optimal filter for the stochastic volatility model.

10. Conclusion

This chapter has focused on a wide range of theoretical properties of ARCH models. It has also presented some new important empirical results, but has not attempted to survey the literature on applications, a recent survey of which can be found in Bollerslev et al. (1992).[31] Three of the most active lines of inquiry are prominently surveyed here, however. The first concerns the general parametrizations of univariate discrete time models of time-varying heteroskedasticity. From the original ARCH model, the literature has focused upon GARCH, EGARCH, IGARCH, ARCH-M, AGARCH, NGARCH, QARCH, QTARCH, STARCH, SWARCH and many other formulations with particular distinctive properties. Not only has this literature been surveyed here, but it has been expanded by the analysis of variations in the EGARCH model. Second, we have explored the relations between the discrete time models and the very popular continuous time diffusion processes that are widely used in

[31] Other recent surveys of the ARCH methodology are given in Bera and Higgins (1993) and Nijman and Palm (1993).

finance. Very useful approximation theorems have been developed, which hold with increasing accuracy when the length of the sampling interval diminishes. The third area of important investigation concerns the analysis of multivariate ARCH processes. This problem is more complex than the specification of univariate models because of the interest in simultaneously modeling a large number of variables, or assets, without having to estimate an intractable large number of parameters. Several multivariate formulations have been proposed, but no clear winners have yet emerged, either from a theoretical or an empirical point of view.

References

Akaike, H. (1973) "Information Theory and an Extension of the Maximum Likelihood Principle", in: B.N. Petrov and F. Csáki, eds., *Second International Symposium on Information Theory.* Akadémiai Kiadó: Budapest.

Amemiya, T. (1985) *Advanced Econometrics.* Harvard University Press: Cambridge, MA.

Amin, K.I. and V.K. Ng (1993) "Equilibrium Option Valuation with Systematic Stochastic Volatility", *Journal of Finance*, 48, 881–910.

Andersen, T.G. (1992a) Volatility, unpublished manuscript, J.L. Kellogg Graduate School of Management, Northwestern University.

Andersen, T.G. (1992b) Return Volatility and Trading Volume in Financial Markets: An Information Flow Interpretation of Stochastic Volatility, unpublished manuscript, J.L. Kellogg Graduate School of Management, Northwestern University.

Anderson, T.W. (1971) *The Statistical Analysis of Time Series.* John Wiley and Sons: New york, NY.

Andrews, D.W.K. and W. Ploberger (1992) Optimal Tests when a Nuisance Parameter Is Present only under the Alternative, unpublished manuscript, Department of Economics, Yale University.

Andrews, D.W.K. and W. Ploberger (1993) Admissibility of the Likelihood Ratio Test when a Nuisance Parameter Is Present only under the Alternative, unpublished manuscript, Department of Economics, Yale University.

Attanasio, O. (1991) "Risk, Time-Varying Second Moments and Market Efficiency", *Review of Economic Studies*, 58, 479–494.

Baek, E.G. and W.A. Brock (1992) "A Nonparametric Test for Independence of a Multivariate Time Series", *Statistica Sinica*, 2, 137–156.

Baillie, R.T. and T. Bollerslev (1989) "The Message in Daily Exchange Rates: A Conditional Variance Tale", *Journal of Business and Economic Statistics*, 7, 297–305.

Baillie, R.T. and T. Bollerslev (1990) "A Multivariate Generalized ARCH Approach to Modeling Risk Premia in Forward Foreign Exchange Rate Markets", *Journal of International Money and Finance*, 9, 309–324.

Baillie, R.T. and T. Bollerslev (1991) "Intra Day and Inter Day Volatility in Foreign Exchange Rates", *Review of Economic Studies*, 58, 565–585.

Baillie, R.T. and T. Bollerslev (1992) "Prediction in Dynamic Models with Time Dependent Conditional Variances", *Journal of Econometrics*, 52, 91–113.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1993), Fractionally Integrated Autoregressive Conditional Heteroskedasticity, unpublished manuscript, J.L. Kellogg Graduate School of Management, Northwestern University.

Bekaert, G. and R.J. Hodrick (1993) "On Biases in the Measurement of Foreign Exchange Risk Premiums", *Journal of International Money and Finance*, 12, 115–138.

Bera, A.K. and M.L. Higgins (1993) "ARCH Models: Properties, Estimation and Testing", *Journal of Economic Surveys*, 7, 305–366.

Bera, A.K. and S. Lee (1992) "Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis", *Review of Economic Studies*, 60, 229–240.

Bera, A.K. and J-S. Roh (1991) A Moment Test of the Constancy of the Correlation Coefficient in the Bivariate GARCH Model, unpublished manuscript, Department of Economics, University of Illinois, Urbana–Champaign.

Bera, A.K., M.L. Higgins and S. Lee (1993) "Interaction Between Autocorrelation and Conditional Heteroskedasticity: A Random Coefficients Approach", *Journal of Business and Economic Statistics*, 10, 133–142.

Berndt E.R., B.H. Hall, R.E. Hall, and J.A. Haussman (1974) "Estimation and Inference in Nonlinear Structural Models", *Annals of Economic and Social Measurement*, 4, 653–665.

Black, F. (1976) "Studies of Stock Price Volatility Changes", *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, 177–181.

Black, F. and M. Scholes (1973) "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81, 637–659.

Blattberg, R.C. and N.J. Gonedes (1974) "A Comparison of the Stable and Student Distribution of Statistical Models for Stock Prices", *Journal of Business*, 47, 244–280.

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

Bollerslev, T. (1987) "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return", *Review of Economics and Statistics*, 69, 542–547.

Bollerslev, T. (1988) "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process", *Journal of Time Series Analysis*, 9, 121–131.

Bollerslev, T. (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 I. Domowitz (1993) "Trading Patterns and the Behavior of Prices in the Interbank Foreign Exchange Market," *Journal of Finance*, 48, 1421–1443.

Bollerslev, T. and R.F. Engle (1993) "Common Persistence in Conditional Variances", *Econometrica*, 61, 166–187.

Bollerslev, T. and M. Melvin (1994) "Bid-Ask Spreads in the Foreign Exchange Market: An Empirical Analysis", *Journal of International Economics*, forthcoming.

Bollerslev, T. and J.M. Wooldridge (1992) "Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances", *Econometric Reviews*, 11, 143–172.

Bollerslev, T., R.F. Engle and J.M. Wooldridge (1988) "A Capital Asset Pricing Model with Time Varying Covariances", *Journal of Political Economy*, 96, 116–131.

Bollerslev, T., R.Y. Chou and K.F. Kroner (1992) "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics*, 52, 5–59.

Bougerol, P. and N. Picard (1992) "Stationarity of GARCH Processes and of Some Non-Negative Time Series", *Journal of Econometrics*, 52, 115–128.

Box, G.E.P., and G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control.* Holden Day: San Francisco, CA. Second Edition.

Braun, P.A., D.B. Nelson and A.M. Sunier (1992) Good News, Bad News, Volatility, and Betas, unpublished manuscript, Graduate School of Business, University of Chicago.

Breusch, T. and A.R. Pagan (1979) "A Simple Test for Heteroskedasticity and Random Coefficient Variation", *Econometrica*, 47, 1287–1294.

Brock, W.A. and A. Kleidon (1992) "Periodic Market Closure and Trading Volume: A Model of Intra Day Bids and Asks", *Journal of Economic Dynamics and Control*, 16, 451–489.

Brock, W.A., and S.M. Potter (1992), Nonlinear Time Series and Macroeconometrics, unpublished manuscript, Department of Economics, University of Wisconsin, Madison.

Brock, W.A., W.D. Dechert and J.A. Scheinkman (1987) A Test for independence Based on the Correlation Dimension, unpublished manuscript, Department of Economics, University of Wisconsin, Madison.

Brock, W.A., D.A. Hsieh and B. LeBaron (1991). *Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence.* MIT Press: Cambridge, MA.

Cai, J. (1994) "A Markov Model of Unconditional Variance in ARCH", *Journal of Business and Economic Statistics*, forthcoming.

Campbell, J.Y. and L. Hentschel (1992) "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns", *Journal of Financial Economics*, 31, 281–318.

Chou, R.Y. (1988) "Volatility Persistence and Stock Valuations: Some Empirical Evidence Using GARCH", *Journal of Applied Econometrics*, 3, 279–294.

Christie, A.A. (1982) "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects", *Journal of Financial Economics*, 10, 407–432.

Clark, P.K. (1973) "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", *Econometrica*, 41, 135–156.

Cornell, B. (1978) "Using the Options Pricing Model to Measure the Uncertainty Producing Effect of Major Announcements", *Financial Management*, 7, 54–59.

Crowder, M.J. (1976) "Maximum Likelihood Estimation with Dependent Observations", *Journal of the Royal Statistical Society*, 38, 45–53.

Danielson, J. and J.-F. Richard (1993) "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models", *Journal of Applied Econometrics*, 8, S153–S173.

Davidian, M. and R.J. Carroll (1987) "Variance Function Estimation", *Journal of the American Statistical Association*, 82, 1079–1091.

Davies, R.B. (1977) "Hypothesis Testing when a Nuisance Parameter is Present only under the Null Hypothesis", *Biometrika*, 64, 247–254.

Day, T.E. and C.M. Lewis (1992) "Stock Market Volatility and the Information Content of Stock Index Options", *Journal of Econometrics*, 52, 267–288.

Demos, A. and E. Sentana (1991) Testing for GARCH Effects: A One-Sided Approach, unpublished manuscript, London School of Economics.

Diebold, F.X. (1987) "Testing for Serial Correlation in the Presence of ARCH", *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, 323–328.

Diebold, F.X. (1988) *Empirical Modeling of Exchange Rate Dynamics*. Springer Verlag: New York, NY.

Diebold, F.X. and M. Nerlove (1989) "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model", *Journal of Applied Econometrics*, 4, 1–21.

Ding, Z., R.F. Engle, and C.W.J. Granger (1993) "Long Memory Properties of Stock Market Returns and a New Model", *Journal of Empirical Finance*, 1, 83–106.

Drost, F.C. and T.E. Nijman (1993) "Temporal Aggregation of GARCH Processes", *Econometrica*, 61, 909–927.

Engle, R.F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica*, 50, 987–1008.

Engle, R.F. (1984) "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics", in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics, Vol. II*. North-Holland: Amsterdam.

Engle, R.F. (1987) Multivariate GARCH with Factor Structures – Cointegration in Variance, unpublished manuscript, Department of Economics, UCSD.

Engle, R.F. (1990) "Discussion: Stock Market Volatility and the Crash of 87", *Review of Financial Studies*, 3, 103–106.

Engle, R.F. and T. Bollerslev (1986) "Modelling the Persistence of Conditional Variances", *Econometric Reviews*, 5, 1–50, 81–87.

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

Engle, R.F. and C.W.J. Granger (1987) "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251–276.

Engle, R.F. and S. Kozicki (1993) "Testing for Common Features", *Journal of Business and Economic Statistics*, 11, 369–379.

Engle, R.F. and K.F. Kroner (1993) Multivariate Simultaneous Generalized ARCH, unpublished manuscript, Department of Economics, UCSD.

Engle, R.F. and G.G.J. Lee (1992) A Permanent and Transitory Component Model of Stock Return Volatility, unpublished manuscript, Department of Economics, UCSD.

Engle, R.F. and G.G.J. Lee (1993) Long Run Volatility Forecasting for Individual Stocks in a One Factor Model, unpublished manuscript, Department of Economics, UCSD.

Engle, R.F. and C. Mustafa (1992) "Implied ARCH Models from Options Prices", *Journal of Econometrics*, 52, 289–311.

Engle, R.F. and V.K. Ng (1993) "Measuring and Testing the Impact of News on Volatility", *Journal of Finance*, 48, 1749–1778.

Engle R.F. and R. Susmel (1993) "Common Volatility in International Equity Markets", *Journal of Business and Economic Statistics*, 11, 167–176.

Engle, R.F., D.F. Hendry and D. Trumble (1985) "Small Sample Properties of ARCH Estimators and Tests", *Canadian Journal of Economics*, 18, 66–93.

Engle, R.F., D.M. Lilien and R.P. Robins (1987) "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model", *Econometrica*, 55, 391–407.

Engle, R.F., T. Ito and W-L. Lin (1990a) "Meteor Showers or Heat Waves? Heteroskedastic Intra Daily Volatility in the Foreign Exchange Market", *Econometrica*, 58, 525–542.

Engle, R.F., V. Ng and M. Rothschild (1990b) "Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills", *Journal of Econometrics*, 45, 213–238.
Engle, R.F., C-H. Hong, A. Kane, and J. Noh (1993) "Arbitrage Valuation of Variance Forecasts with Simulated Options", in: D.M. Chance and R.R. Trippi, eds., *Advances in Futures and Options Research*. JAI Press: Greenwich, Connecticut.
Ethier, S.N. and T.G. Kurtz (1986) *Markov processes: Characterization and Convergence*. John Wiley: New York, NY.
Fama, E.F. (1963) "Mandelbrot and the Stable Paretian Distribution", *Journal of Business*, 36, 420–429.
Fama, E.F. (1965) "The Behavior of Stock Market Prices", *Journal of Business*, 38, 34–105.
Foster, D.P. and D.B. Nelson (1992) Rolling Regressions, unpublished manuscript, Graduate School of Business, University of Chicago.
French, K.R. and R. Roll (1986) "Stock Return Variances: The Arrival of Information and the Reaction of Traders", *Journal of Financial Economics*, 17, 5–26.
French, K.R., G.W. Schwert and R.F. Stambaugh (1987) "Expected Stock Returns and Volatility", *Journal of Financial Economics*, 19, 3–30.
Gallant, A.R. and G. Tauchen (1989) "Semi Non-Parametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications", *Econometrica*, 57, 1091–1120.
Gallant, A.R., D.A. Hsieh and G. Tauchen (1991) "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate 1974–83", in: W.A. Barnett, J. Powell and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press: Cambridge.
Gallant, A.R., P.E. Rossi and G. Tauchen (1992) "Stock Prices and Volume", *Review of Financial Studies*, 5, 199–242.
Gallant, A.R., P.E. Rossi and G. Tauchen (1993) "Nonlinear Dynamic Structures", *Econometrica*, 61, 871–907.
Gennotte, G. and T.A. Marsh (1991) Variations in Economic Uncertainty and Risk Premiums on Capital Assets, unpublished manuscript, Department of Finance, University of California, Berkeley.
Gerity, M.S. and J.H. Mulherin (1992) "Trading Halts and Market Activity: An Analysis of Volume at the Open and the Close", *Journal of Finance*, 47, 1765–1784.
Geweke, J. (1989a) "Exact Predictive Densities in Linear Models with ARCH Disturbances", *Journal of Econometrics*, 44, 307–325.
Geweke, J. (1989b) "Bayesian Inference in Econometric Models Using Monte Carlo Integration", *Econometrica*, 57, 1317–1339.
Glosten, L.R., R. Jagannathan and D. Runkle (1993) "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance*, 48, 1779–1801.
Gourieroux, C. and A. Monfort (1992) "Qualitative Threshold ARCH Models", *Journal of Econometrics*, 52, 159–199.
Gourieroux, C., A. Holly and A. Monfort (1982) "Likelihood Ratio Test, Wald Test and Kuhn–Tucker Test in Linear Models with Inequality Constraints on Regression Parameters", *Econometrica*, 50, 63–80.
Granger, C.W.J., R.F. Engle, and R.P. Robins (1986) "Wholesale and Retail Prices: Bivariate Time-Series Modelling with Forecastable Error Variances", in: D. Belsley and E. Kuh, eds., *Model Reliability*. MIT Press: Massachusetts. pp 1–17.
Hamao, Y., R.W. Masulis and V.K. Ng (1990) "Correlations in Price Changes and Volatility Across International Stock Markets", *Review of Financial Studies*, 3, 281–307.
Hamilton, J.D. and R. Susmel (1992), Autoregressive Conditional Heteroskedasticity and Changes in Regime, unpublished manuscript, Department of Economics, UCSD.
Harris, L. (1986) "A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns", *Journal of Financial Economics*, 16, 99–117.
Harvey, C.R. and R.D. Huang (1991) "Volatility in the Foreign Currency Futures Market", *Review of Financial Studies*, 4, 543–569.
Harvey, C.R. and R.D. Huang (1992) Information Trading and Fixed Income Volatility, unpublished manuscript, Department of Finance, Duke University.
Harvey, A.C., E. Ruiz and E. Sentana (1992) "Unobserved Component Time Series Models with ARCH Disturbances", *Journal of Econometrics*, 52, 129–158.
Harvey, A.C., E. Ruiz and N. Shephard (1994) "Multivariate Stochastic Volatility Models", *Review of Economic Studies*, forthcoming.
Heston, S.L. (1991) A Closed Form Solution for Options with Stochastic Volatility, unpublished manuscript, Department of Finance, Yale University.

Higgins, M.L. and A.K. Bera (1992) "A Class of Nonlinear ARCH Models", *International Economic Review*, 33, 137–158.
Hong, P.Y. (1991) "The Autocorrelation Structure for the GARCH-M Process", *Economics Letters*, 37, 129–132.
Hsieh, D.A. (1991) "Chaos and Nonlinear Dynamics: Applications to Financial Markets", *Journal of Finance*, 46, 1839–1878.
Huber, P.J. (1977). *Robust Statistical Procedures*. SIAM: Bristol, United Kingdom.
Hull, J. and A. White (1987) "The Pricing of Options on Assets with Stochastic Volatilities". *Journal of Finance*, 42, 281–300.
Jacquier, E., N.G. Polson and P.E. Rossi (1994) "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics*, forthcoming.
Karatzas, I. and S.E. Shreve (1988) *Brownian Motion and Stochastic Calculus*. Springer–Verlag: New York, NY.
Karpoff, J.M. (1987) "The Relation Between Price Changes and Trading Volume: A Survey", *Journal of Financial and Quantitative Analysis*, 22, 109–126.
Kim, C.M. (1989) Nonlinear Dependence of Exchange Rate Changes, unpublished Ph.D. dissertation, Graduate School of Business, University of Chicago.
King, M., E. Sentana and S. Wadhwani (1994) "Volatility and Links Between National Stock Markets", *Econometrica*, forthcoming.
Kitagawa, G. (1987) "Non-Gaussian State Space Modelling of Nonstationary Time Series", *Journal of the American Statistical Association*, 82, 1032–1063.
Kodde, D.A. and F.C. Palm (1986) "Wald Criterion for Jointly Testing Equality and Inequality Restrictions", *Econometrica*, 54, 1243–1248.
Kraft, D.F. and R.F. Engle (1982) Autoregressive Conditional Heteroskedasticity in Multiple Time Series, unpublished manuscript, Department of Economics, UCSD.
Krengel, U. (1985). *Ergodic Theorems*. Walter de Gruyter: Berlin, Germany.
Kroner, K.F. and S. Claessens (1991) "Optimal Dynamic Hedging Portfolios and the Currency Composition of External Debt", *Journal of International Money and Finance*, 10, 131–148.
Kroner, K.F. and J. Sultan (1991) "Exchange Rate Volatility and Time Varying Hedge Ratios", in: S.G. Rhee and R.P. Chang, eds., *Pacific-Basin Capital Markets Research, Vol. II*. North-Holland: Amsterdam.
Lamoureux, C.G. and W.D. Lastrapes (1990) "Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects", *Journal of Finance*, 45, 221–229.
Lamoureux, C.G. and W.D. Lastrapes (1994) "Endogenous Trading Volume and Momentum in Stock Return Volatility", *Journal of Business and Economic Statistics*, forthcoming.
LeBaron, B. (1992) "Some Relations Between Volatility and Serial Correlation in Stock Market Returns", *Journal of Business*, 65, 199–220.
Lee, J.H.H. and M.L. King (1993) "A Locally Most Mean Powerful Based Score Test for ARCH and GARCH Regression Disturbances", *Journal of Business and Economic Statistics*, 7, 259–279.
Lee, S.W. and B.E. Hansen (1993) Asymptotic Theory for the GARCH(1, 1) Quasi-Maximum Likelihood Estimator, unpublished manuscript, Department of Economics, University of Rochester.
Lin, W.L. (1992) "Alternative Estimators for Factor GARCH Models – A Monte Carlo Comparison", *Journal of Applied Econometrics*, 7, 259–279.
Lin, W.L., R.F. Engle and T. Ito (1994) "Do Bulls and Bears Move Across Borders? International Transmission of Stock Returns and Volatility as the World Turns", *Review of Financial Studies*, forthcoming.
Ljung, G.M. and G.E.P. Box (1978) "On a Measure of Lag of Fit in Time Series Models", *Biometrika*, 67, 297–303.
Lumsdaine, R.L. (1992a) Asymptotic Properties of the Quasi-Maximum Likelihood Estimator in GARCH(1, 1) and IGARCH(1, 1) Models, unpublished manuscript, Department of Economics, Princeton University.
Lumsdaine, R.L. (1992b) Finite Sample Properties of the Maximum Likelihood Estimator in GARCH(1, 1) and IGARCH(1, 1) Models: A Monte Carlo Investigation, unpublished manuscript, Department of Economics, Princeton University.
MacKinnon, J.G. and H. White (1985) "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties", *Journal of Econometrics*, 29, 305–325.
Mandelbrot, B. (1963) "The Variation of Certain Speculative Prices", *Journal of Business*, 36, 394–419.

Marcus, M. and H. Minc (1964) *A Survey of Matrix Theory and Matrix Inequalities.* Prindle, Weber and Schmidt: Boston, MA.

McCurdy, T.H. and T. Stengos (1992) "A Comparison of Risk Premium Forecasts Implied by Parametric and Nonparametric Conditional Mean Estimators", *Journal of Econometrics*, 52, 225–244.

McDonald, J.B. and W.K. Newey (1988) "Partially Adaptive Estimation of Regression Models via the Generalized t Distribution", *Econometric Theory*, 4, 428–457.

Melino, A. and S. Turnbull (1990) "Pricing Foreign Currency Options with Stochastic Volatility", *Journal of Econometrics*, 45, 239–266.

Merton, R.C. (1973) "An Intertemporal Capital Asset Pricing Model", *Econometrica*, 42, 867–887.

Merton, R.C. (1980) "On Estimating the Expected Return on the Market", *Journal of Financial Economics*, 41, 867–887.

Milhøj, A. (1985) "The Moment Structure of ARCH Processes", *Scandinavian Journal of Statistics*, 12, 281–292.

Murphy, K. and R. Topel (1985) "Estimation and Inference in Two-Step Econometric Models", *Journal of Business and Economic Statistics*, 3, 370–379.

Nelson, D.B. (1989) "Modeling Stock Market Volatility Changes", *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, 93–98.

Nelson, D.B. (1990a) "ARCH Models as Diffusion Approximations", *Journal of Econometrics*, 45, 7–38.

Nelson, D.B. (1990b) "Stationarity and Persistence in the GARCH(1, 1) Model", *Econometric Theory*, 6, 318–334.

Nelson, D.B. (1991) "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, 59, 347–370.

Nelson, D.B. (1992) "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model", *Journal of Econometrics*, 52, 61–90.

Nelson, D.B. and C.Q. Cao (1992) "Inequality Constraints in the Univariate GARCH Model", *Journal of Business and Economic Statistics*, 10, 229–235.

Nelson, D.B. and D.P. Foster (1991) Filtering and Forecasting with Misspecified ARCH Models II: Making the Right Forecast with the Wrong Model, unpublished manuscript, Graduate School of Business, University of Chicago.

Nelson, D.B. and D.P. Foster (1994) "Asymptotic Filtering Theory for Univariate ARCH Models", *Econometrica*, 62, 1–41.

Newey, W.K. (1985) "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica*, 53, 1047–1070.

Ng, V., R.F. Engle and M. Rothschild (1992) "A Multi-Dynamic Factor Model for Stock Returns", *Journal of Econometrics*, 52, 245–265.

Nijman, T.E. and F.C. Palm (1993) "GARCH Modelling of Volatility: An Introduction to Theory and Applications", in: A.J. de Zeeuw, ed., *Advanced Lectures in Quantitative Economics.* Academic Press: London.

Nijman, T.E. and E. Sentana (1993) Marginalization and Contemporaneous Aggregation in Multivariate GARCH Processes, unpublished manuscript, Center for Economic Research, Tilburg University.

Nummelin, E. and P. Tuominen (1982) "Geometric Ergodicity of Harris Recurrent Markov Chains with Applications to Renewal Theory," *Stochastic Processes and Their Applications*, 12, 187–202.

Pagan, A.R. (1984) "Econometric Issues in the Analysis of Regressions with Generated Regressors", *International Economic Review*, 25, 221–247.

Pagan, A.R. (1986) "Two Stage and Related Estimators and their Applications", *Review of Economic Studies*, 53, 517–538.

Pagan, A.R. and Y.S. Hong (1991) "Nonparametric Estimation and the Risk Premium", in: W.A. Barnett, J. Powell and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics.* Cambridge University Press: Cambridge.

Pagan, A.R. and H.C.L. Sabau (1987a), On the Inconsistency of the MLE in Certain Heteroskedastic Regression Models, unpublished manuscript, University of Rochester.

Pagan, A.R. and H.C.L. Sabau (1987b) Consistency Tests for Heteroskedasticity and Risk Models, unpublished manuscript, Department of Economics, University of Rochester.

Pagan, A.R. and G.W. Schwert (1990) "Alternative Models for Conditional Stock Volatility", *Journal of Econometrics*, 45, 267–290.

Pagan, A.R. and A. Ullah (1988) "The Econometric Analysis of Models with Risk Terms", *Journal of Applied Econometrics*, 3, 87–105.

Pagan, A.R., A.D. Hall and P.K. Trivedi (1983) "Assessing the Variability of Inflation", *Review of Economic Studies*, 50, 585–596.

Pardoux, E. and D. Talay (1985) "Discretization and Simulation of Stochastic Differential Equations", *Acta Applicandae Mathematica*, 3, 23–47.

Parkinson, M. (1980) "The Extreme Value Method for Estimating the Variance of the Rate of Return", *Journal of Business*, 53, 61–65.

Patell, J.M. and M.A. Wolfson (1979) "Anticipated Information Releases Reflected in Call Option Prices", *Journal of Accounting and Economics*, 1, 117–140.

Patell, J.M. and M.A. Wolfson (1981) "The Ex-Ante and Ex-Post Price Effects of Quarterly Earnings Announcement Reflected in Option and Stock Price", *Journal of Accounting Research*, 19, 434–458.

Poterba, J. and L. Summers (1986) "The Persistence of Volatility and Stock Market Fluctuations", *American Economic Review*, 76, 1142–1151.

Rich, R.W., J.E. Raymond, and J.S. Butler (1992) "The Relationship between Forecast Dispersion and Forecast Uncertainty: Evidence from a Survey Data-ARCH Model", *Journal of Applied Econometrics*, 7, 131–148.

Royden, H.L. (1968) *Real Analysis*. Macmillan Publishing Co.: New York, NY.

Scheinkman, J., and B. LeBaron (1989) "Nonlinear Dynamics and Stock Returns", *Journal of Business*, 62, 311–337.

Schwarz, G. (1978) "Estimating the Dimension of a Model", *Annals of Statistics*, 6, 461–464.

Schwert, G.W. (1989a) "Why Does Stock Market Volatility Change Over Time", *Journal of Finance*, 44, 1115–1153.

Schwert, G.W. (1989b) "Business Cycles, Financial Crises, and Stock Volatility", *Carnegie–Rochester Conference Series on Public Policy*, 39, 83–126.

Schwert, G.W. (1990) "Indexes of U.S. Stock Prices from 1802 to 1987", *Journal of Business*, 63, 399–426.

Schwert, G.W. and P.J. Seguin (1990) "Heteroskedasticity in Stock Returns", *Journal of Finance*, 45, 1129–1155.

Scott, L.O. (1987) "Option Pricing when the Variance Changes Randomly: Theory, Estimation and an Application", *Journal of Financial and Quantitative Analysis*, 22, 419–438.

Sentana, E. (1991) Quadratic ARCH Models: A Potential Re-Interpretation of ARCH Models, unpublished manuscript, London School of Economics.

Shephard, N. (1993) "Fitting Nonlinear Time Series Models with Applications to Stochastic Variance Models", *Journal of Applied Economics*, 8, S135–S152.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London, United Kingdom.

Stambaugh, R.F. (1993) Estimating Conditional Expectations When Volatility Fluctuates, unpublished manuscript, The Wharton School, University of Pennsylvania.

Stroock, D.W. and S.R.S. Varadhan (1979) *Multidimensional Diffusion Processes*. Springer–Verlag: Berlin, Germany.

Tauchen, G. (1985) "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *Journal of Econometrics*, 30, 415–443.

Tauchen, G. and M. Pitts (1983) "The Price Variability–Volume Relationship on Speculative Markets", *Econometrica*, 51, 485–505.

Taylor, S. (1986) *Modeling Financial Time Series*. Wiley and Sons: New York, NY.

Tsay, R.S. (1987) "Conditional Heteroskedastic Time Series Models", *Journal of the American Statistical Association*, 82, 590–604.

Tweedie, R.L. (1983a) "Criteria for Rates of Convergence of Markov Chains, with Application to Queuing and Storage Theory", in: J.F.C. Kingman and G.E.H. Reuter, eds., *Probability, Statistics, and Analysis, London Mathematical Society Lecture Note Series No. 79*. Cambridge University Press: Cambridge.

Tweedie, R.L. (1983b) "The Existence of Moments for Stationary Markov Chains", *Journal of Applied Probability*, 20, 191–196.

Watson, M.W. and R.F. Engle (1985) "Testing for Regression Coefficient Stability with a Stationary AR(1) Alternative", *Review of Economics and Statistics*, 67, 341–346.

Weiss, A.A. (1984) "ARMA Models with ARCH Errors", *Journal of Time Series Analysis*, 5, 129–143.

Weiss, A.A. (1986) "Asymptotic Theory for ARCH Models: Estimation and Testing", *Econometric Theory*, 2, 107–131.

West, K.D., H.J. Edison and D. Cho (1993) "A Utility Based Comparison of Some Models for Exchange Rate Volatility", *Journal of International Economics*, 35, 23–45.

White, H. (1980) "A Heteroskedastic-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity", *Econometrica*, 48, 421–448.

White, H. (1987) "Specification Testing in Dynamic Models", in: T.F. Bewley, ed., *Advances in Econometrics: Fifth World Congress, Vol. I*. Cambridge University Press: Cambridge.

White, H. (1994) *Estimation, Inference and Specification Analysis*, forthcoming.

Wiggins, J.B. (1987) "Option Values under Stochastic Volatility: Theory and Empirical Estimates", *Journal of Financial Economics*, 19, 351–372.

Wiggins, J.B. (1991) "Empirical Tests of the Bias and Efficiency of the Extreme-Value Variance Estimator for Common Stocks", *Journal of Business*, 64, 417–432.

Wolak, F.A. (1991) "The Local Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models", *Econometrica*, 59, 981–995.

Wooldridge, J.M. (1990) "A Unified Approach to Robust Regression Based Specification Tests", *Econometric Theory*, 6, 17–43.

Wooldridge, J.M. (1994) "Estimation and Inference for Dependent Processes", in: R.F. Engle and D. McFadden, eds., *Handbook of Econometrics, Vol. IV*. North-Holland: Amsterdam, Chapter 45.

Zakoian, J.-M. (1990) Threshold Heteroskedastic Models, unpublished manuscript, CREST, INSEE.

Zarnowitz, V. and L.A. Lambros (1987) "Consensus and Uncertainty in Economic Prediction", *Journal of Political Economy*, 95, 591–621.