JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **15**: 117–136 (2000)

AN EMPIRICAL ANALYSIS OF ALTERNATIVE PARAMETRIC ARCH MODELS

GEOFFREY F. LOUDON,[a] WING H. WATT[b] AND PRADEEP K. YADAV[c]*
[a] *Macquarie University, Australia*
[b] *DBS Bank, Singapore*
[c] *Department of Accounting and Finance, University of Strathclyde, 100 Cathedral Street, Glasgow G4 0LN, UK*

SUMMARY

This paper presents empirical evidence on the effectiveness of eight different parametric ARCH models in describing daily stock returns. Twenty-seven years of UK daily data on a broad-based value weighted stock index are investigated for the period 1971–97. Several interesting results are documented. Overall, the results strongly demonstrate the utility of parametric ARCH models in describing time-varying volatility in this market. The parameters proxying for asymmetry in models that recognize the asymmetric behaviour of volatility are highly significant in each and every case. However, the 'performance' of the various parameterizations is often fairly similar with the exception of the multiplicative GARCH model that performs qualitatively differently on several dimensions of performance. The outperformance of any model(s) is not consistent across different sub-periods of the sample, suggesting that the optimal choice of a model is period-specific. The outperformance is also not consistent as we change from in-sample inferences to out-of-sample inferences within the same period.

1. INTRODUCTION

There is extensive empirical evidence that stock market volatility varies systematically with time. The evidence dates back to the pioneering studies of Mandelbrot (1963) and Fama (1965) who found that large price changes tend to be followed by large price changes and small price changes by small price changes. More recent evidence is provided by Poterba and Summers (1986), French *et al*. (1987), Chou (1988) and Schwert (1990).

There is also strong evidence that ARCH models are good descriptions of this time-varying volatility in stock returns. Review articles such as Bollerslev *et al*. (1992) document the effective application of ARCH models to financial time series across a wide variety of markets.[1] Examples where significant ARCH effects are documented include Engle and Mustafa (1992) for individual US stocks, Akgiray (1989) for US stock indices, Poon and Taylor (1992) for a UK stock index, Corhay and Rad (1991) for a selection of international stock indices and Frennberg and Hansson (1992) for the Swedish stock market.

* Correspondence to: Professor Pradeep K. Yadav, Department of Accounting and Finance, University of Strathclyde, 100 Cathedral Street, Glasgow G4 0LN, UK; e-mail: p.k.yadav@strath.ac.uk

Contract/grant sponsor: UK Economic and Social Research Council; Contract/grant number: R000233234.
Contract/grant sponsor: Scottish Institute for Research in Investment and Finance.

[1] Other non-linear models have been used to describe specific types of financial time series. For example, Yadav *et al*. (1994) use a threshold autoregressive (TAR) framework to model the price difference between equivalent assets (like futures and equivalent spot prices) in the presence of market frictions.

Received 14 February 1994
Revised 23 May 1999

Systematic time variation in returns' volatility implies that the variance at time t can be broken up into predictable and unpredictable components. The predictable component is called the conditional variance and is a function of information available at time $t-1$. This information can include relevant firm-specific and economy-wide variables, and also the past history of the returns' series. Clearly, modelling conditional variance of stock returns is important because expected volatility is a fundamental input in portfolio selection decisions and in models of asset and option pricing.

In its most general form (see Engle, 1982, equations 1–5) a univariate ARCH model makes conditional variance, h_t, at time t a function of exogenous and lagged endogenous variables, time, the vector of parameters, and past residuals. A variety of different parameterizations for the functional dependence of h_t on the above variables have been proposed by econometricians. All these parameterizations of conditional heteroscedasticity have been motivated either by the need to more effectively model some specific empirical features of the underlying data[2] or by considerations of computational simplicity.[3] In contrast, there has been relatively little work on the economic foundations of ARCH models[4] or the formulation, from first principles, of equilibrium asset pricing models with ARCH features. Since the different parameterizations of ARCH models have been motivated essentially on the basis of data-specific empirical features, their relative usefulness can only be judged by the extent to which they are able to explain time variation in conditional volatility in actual data.

More recent ARCH models include additional parameters to those in the linear ARCH(q) of Engle (1982) and the linear GARCH(p, q) of Bollerslev (1986). Hence, clearly these recent models could be potentially better *ex post* descriptions of the return-generating process. But the *ex ante* usefulness of these models for portfolio selection and asset pricing decisions depends on the out-of-sample predictive ability of these models. It also depends on whether the outperformance is consistent across different sub-periods. Furthermore, it is important to appreciate the influence of market micro-structural factors. The results of Bollerslev and Domowitz (1991) show that the trade execution process can significantly alter the intertemporal dependence in conditional volatility of high-frequency returns because of the differences in serial correlation in market spreads across different trading systems.

Several studies have examined the effectiveness of different individual ARCH models in modelling conditional stock market volatility relative to the linear ARCH and the linear GARCH model. Comparisons of EGARCH with GARCH are reported *inter alia* by Nelson (1991) using daily CRSP index returns from July 1962 to December 1987; Kearns and Pagan (1993) using 88 years of monthly Australian stock index data; Poon and Taylor (1992) using daily, weekly, fortnightly and monthly UK stock index data over a period exceeding 20 years; and Zakoian (1994) using 2 years of daily French CAC40 index data. Sentana (1991) compares ARCH models with the linear GARCH using a century of US stock index data. Rabemananjara and Zakoian (1993) compare a threshold GARCH model with the linear GARCH using about 2 years of daily data on the French CAC40 index and some individual stocks comprising the index. Pagan and Schwert (1990) analyse several alternative conditional volatility models using monthly data from 1835 to 1925, but most of these are non-parametric models, the only parametric ARCH models being the EGARCH and the linear GARCH. Engle and Ng (1993) evaluate six

[2] For example, Black (1976) and Christie (1982) document the negative correlation between current returns and future volatility.
[3] For example, EGARCH models avoid the need for non-negativity constraints in estimation.
[4] This has been pointed out, for example, by Bollerslev *et al.* (1992, section 2.9).

parametric ARCH models using eight years of daily returns on the Japanese Topix Index. However, the main focus of their study is the development of new diagnostic tests and the asymmetry of the volatility response to news, and though they do a sub-sample robustness check, their empirical analysis does not include an evaluation of the out-of-sample predictive ability of the models.

This paper provides comprehensive empirical evidence on different parametric ARCH models using daily data on a value weighted stock index, covering the 27-year period 1971–97. It seeks to make a contribution in several directions. First, it documents for the same data, the relative effectiveness of most of the major parametric ARCH models that have been proposed in the literature. Second, the *ex ante* usefulness of these models for portfolio selection and asset pricing decisions is evaluated by quantifying their out-of-sample predictive ability. Third, it employs data from the UK market where reported prices represent truly tradable firm dealer quotes[5] (and hence economically significant prices) rather than last transaction prices[6] as in the USA or Japan on which most of the earlier work has been based.

This paper is organized as follows. Section 2 provides a specification of the parametric ARCH models examined; Section 3 describes the data and empirical results; and Section 4 summarizes the conclusions.

2. PARAMETRIC ARCH MODELS

Numerous parametric ARCH models have appeared in the literature. Given the complexity of financial markets and the absence of specific theoretical models for the appropriate functional form, model selection is a difficult issue. To demonstrate the issues involved we choose to estimate eight representative models that are specializations of the augmented GARCH(1,1)[7] process analysed in Duan (1997).

Since the focus of this study is on conditional volatility it is necessary to remove possible predictability in the conditional mean. Theoretical asset pricing models for expected returns do not explain *ex-post* daily return series particularly well.[8] However, it is well known that predictability in the conditional mean can potentially arise from two sources of empirical regularities: calendar-based day-of-the-week seasonalities[9] and the effects of infrequent trading of portfolio stocks.[10] We use linear models to filter out the predictability arising from these two sources in a way similar to that used by Pagan and Schwert (1990), Engle and Ng (1993) and Kearns and Pagan (1993).

[5] These quotes have been firm typically up to much larger trade sizes than is common for NYSE specialist quotes or NASDAQ market maker quotes.
[6] Clearly, since trades do not necessarily take place at every instant of time, the last transaction price at the end of any time interval need not be the truly tradable price at which a trade can actually be done at that time. The true economic value of the asset could have changed since the time of the last trade.
[7] Corhay and Rad (1991), Chou (1988) and Poon and Taylor (1992) have found that lags of one are not only sufficient but often provide the best fit.
[8] See Fama (1991) for a review.
[9] Yadav and Pope (1992) provide *inter alia* a review of the literature on calendar-based seasonalities.
[10] See Cohen *et al.* (1986) for a review of the methods used to model the effects associated with infrequent trading of portfolio stocks.

We first regress the continuously compounded stock index returns, R_t, on five day-of-the-week dummies—$MON_t, TUE_t, WED_t, THU_t, FRI_t$—and a dummy variable HOL_t for the day following a holiday:

$$R_t = \lambda_1 MON_t + \lambda_2 TUE_t + \lambda_3 WED_t + \lambda_4 THU_t + \lambda_5 FRI_t + \lambda_6 HOL_t + \nu_t \quad \text{(1A)}$$

$\hat{\nu}_t$ is then regressed against a constant and ten autoregressive lags:

$$\hat{\nu}_t = \varphi_0 + \sum_{l=1}^{10} \varphi_l \hat{\nu}_{t-l} + \varepsilon_t \quad \text{(1B)}$$

Hereafter, we refer to the residuals ε_t from this autoregression as *linearly filtered raw returns* and use them as data in estimating the GARCH models analysed in this paper.

The augmented GARCH(1,1) process assumes that ε_t is conditionally normally distributed with mean zero and time-varying variance of h_t given by:

$$\begin{aligned} h_t &= \left\{ \begin{array}{ll} |\lambda\phi_t - \lambda + 1|^{1/\lambda} & \text{if } \lambda \neq 0, \\ \exp(\phi_t - 1) & \text{if } \lambda = 0 \end{array} \right\} \\ \phi_t &= \alpha_0 + \phi_{t-1}\xi_{1,t-1} + \xi_{2,t-1} \\ \xi_{1,t-1} &= \alpha_1 + \alpha_2|\varepsilon_t - c|^\delta + \alpha_3 \max(0, c - \varepsilon_t)^\delta \\ \xi_{2,t-1} &= \alpha_4 f(|\varepsilon_t - c|; \delta) + \alpha_5 f(\max(0, c - \varepsilon_t); \delta) \end{aligned} \quad \text{(2)}$$

where $f(z; \delta) = (z^\delta - 1)/\delta$ for any $z \geqslant 0$.

We choose the generalized model of Duan (1997) for the following reasons. The Box–Cox transformation in the augmented GARCH(1,1) model creates a continuum of both linear and logarithmic conditional variance functions. Like the earlier transformations in Higgins and Bera (1992) and Ding *et al*. (1993), it permits multiplicative shocks to the conditional variance process. Moreover, unlike these earlier transformations, Duan's model also permits additive shocks. It therefore encompasses many of the existing members of the GARCH family. In particular, the augmented GARCH(1,1) process contains the following eight representative specifications of the conditional variance.

(1) Linear GARCH (LGARCH) model (Bollerslev, 1986; Taylor, 1986)
(2) Multiplicative GARCH (MGARCH) model (Geweke, 1986; Pantula, 1986; Milhoj, 1987)
(3) Exponential GARCH (EGARCH) model (Nelson, 1991)
(4) Glosten, Jagganathan and Runkle GARCH (GJR-GARCH) model (Glosten *et al*., 1993)
(5) Non-linear asymmetric GARCH model (Engle and Ng, 1993)
(6) VGARCH model (Engle and Ng, 1993)
(7) TS-GARCH model (Taylor, 1986; Schwert, 1989)
(8) Threshold GARCH model (Zakoian, 1994)

All parameter restrictions required to yield the eight specialisations are summarized below:

LGARCH	$\lambda = 1$, $c = 0$, $\delta = 2$, $\alpha_3 = 0$, $\alpha_4 = 0$, $\alpha_5 = 0$, $\alpha_0 > 0$, $\alpha_1 \geqslant 0$ and $\alpha_2 \geqslant 0$
MGARCH	$\lambda = 0$, $c = 0$, $\delta = 0$, $\alpha_2 = 0$, $\alpha_3 = 0$, $\alpha_4 = 2\alpha_1$ and $\alpha_5 = 0$
EGARCH	$\lambda = 0$, $c = 0$, $\delta = 1$, $\alpha_2 = 0$ and $\alpha_3 = 0$
GJR-GARCH	$\lambda = 1$, $c = 0$, $\delta = 2$, $\alpha_4 = 0$, $\alpha_5 = 0$, $\alpha_0 > 0$, $\alpha_1 \geqslant 0$, $\alpha_2 \geqslant 0$ and $\alpha_2 + \alpha_3 \geqslant 0$

NGARCH $\lambda = 1, \delta = 2, \alpha_3 = 0, \alpha_4 = 0, \alpha_5 = 0, \alpha_0 > 0, \alpha_1 \geqslant 0$ and $\alpha_2 \geqslant 0$
VGARCH $\lambda = 1, \delta = 2, \alpha_2 = 0, \alpha_3 = 0, \alpha_5 = 0, \alpha_0 > 0, \alpha_1 \geqslant 0$ and $\alpha_4 \geqslant 0$
TS-GARCH $\lambda = 1/2, \quad c = 0, \quad \delta = 1, \quad \alpha_3 = 0, \quad \alpha_4 = \alpha_2, \quad \alpha_5 = 0, \quad \alpha_1 \geqslant 0, \quad \alpha_2 \geqslant 0$ and $\alpha_0 > \alpha_1 + \alpha_2 - 1$
TGARCH $\lambda = 1/2, \quad c = 0, \quad \delta = 1, \quad \alpha_4 = \alpha_2, \quad \alpha_5 = \alpha_3, \quad \alpha_1 \geqslant 0, \quad \alpha_2 \geqslant 0, \quad \alpha_2 + \alpha_3 \geqslant 0$ and $\alpha_0 > \alpha_1 + \alpha_2 + \alpha_3 - 1$

Although the LGARCH(1,1) model has been widely used to model financial time series, it has several shortcomings. First, there are features of the data which the model is not capable of describing. In particular, Black (1976) and Christie (1982) document a negative correlation between current returns and future volatility. There is an asymmetry in the impact of news on volatility. Negative news surprises increase predictable volatility more than positive news surprises. The GARCH model fails to capture this 'leverage' effect because it is a symmetric (i.e. quadratic) function of ε_t. Second, the restrictions required in estimation to ensure non-negativity of the conditional variances makes computation more difficult. Furthermore, as highlighted by Rabemananjara and Zakoian (1993), the impact of past volatility shocks, irrespective of sign, always increases with the magnitude of the shock in the LGARCH model making it incapable of describing cyclical or non-linear behaviour in the volatility. These limitations have motivated many of the alternative specializations to LGARCH.

For instance, the EGARCH model avoids the need to restrict $\alpha_1 \geqslant 0$ and $\alpha_2 \geqslant 0$ to ensure non-negativity of the conditional variances. The EGARCH, GJR-GARCH, NGARCH, VGARCH and TS-GARCH models all account for the asymmetric relationship between returns and volatility. While each of these models permit asymmetry, they each do so in a particular way. To analyse such variation in how new information affects the next period variance, the concept of the *News Impact Curve* introduced by Engle and Ng (1993) is particularly revealing. This curve represents the relationship between innovations in returns, ε_{t-1}, and the conditional volatility, h_t, estimated at the sample mean and holding constant information dated $t-2$ and earlier. For the LGARCH model a quadratic curve, symmetric at $\varepsilon_{t-1} = 0$, is obtained. The curves for EGARCH and TGARCH have their minimum also at $\varepsilon_{t-1} = 0$, but they are asymmetric. The NGARCH and VGARCH model not only capture the asymmetric relationship between ε_{t-1} and h_t, but they also allow the asymmetric impact response curve to be centred at a non-zero ε_{t-1}. We therefore conduct the full set of Engle–Ng news impact curve tests as part of our empirical comparison of these models.

Duan (1997) also distinguishes these models in terms of the nature of the implied autoregressive system. The LGARCH, GJR-GARCH, NGARCH, TS-GARCH and TGARCH models have a stochastic autoregressive coefficient in contrast to the constant coefficient implied by the other three.

Given the differences of these eight models, we believe that they are representative of the wide variety of possible ARCH models.

3. EMPIRICAL RESULTS

3.1 Data

The empirical analysis is based on 27 years of daily data on the FT All Share Index of the London Stock Exchange from January 1971 to October 1997.[11] Daily values of the FT All Share Index for the sample period were collected from Datastream and used to calculate daily log returns.

This long sample period is divided into three sub periods — January 1971 to December 1980, January 1981 to December 1990 and January 1991 to October 1997. The second period is of special interest as it contains the major market crash on Black Monday 19 October 1987 and the mini-crash on Black Friday 16 October 1989. Both these episodes involved high volatility, localized in time, following high negative returns.

3.2 Model Estimates

Table I reports estimates of the eight parameterizations of the generalized augmented GARCH process outlined in the previous section. Estimation is by means of maximum likelihood using the algorithm of Berndt *et al.* (1974) assuming conditional normality.[12] Parameters for each of the variance equations (corresponding to the different parametric ARCH models) are estimated separately for each of the three periods (1971–80, 1981–90, 1991–7) for the FT All Share Index. These parameters are used to estimate the daily conditional volatility and together with the diagnostics constitute the in-sample set of results. To estimate the *ex-ante* out-of-sample predictive power of the different models, the parameters of the relevant variance equation estimated from an earlier sub-period are used to compute the conditional volatility in the following period. Therefore, parameters estimated from the 1971–80 data are used to predict conditional volatility for the 1981–90 and the parameters estimated from the 1981–90 data are used to predict conditional volatility for the 1991–7 period.[13]

Table I reports the estimated coefficients. Respective standard errors are given in parentheses below the parameter estimates. Superscripts next to the standard error denote the level of significance (where [a] denotes 10%, [b] denotes 5% and [c] denotes 1%). Table I clearly demonstrates that the conditional variance is strongly related to its previous level and to past surprises in returns. This is seen in that the estimates for the ARCH parameters across all models in each of the periods are highly significant, except for the constant term, α_0, for the VGARCH model in the third period. With the exception of the leverage parameter, α_3, in the GJR-GARCH model for the first period, all significant conditional variance parameters have *p*-values less than 0·01. However, even for that exception, the coefficient is still significant at the 5% level.

Volatility measures exhibit a high degree of persistence. In the first and last periods, the parameter linking the level of the current and previous variance, α_1, lies between 0·899 and 0·992, except for the case of the MGARCH specification. The value of this coefficient is consistently lower in the middle period which contains the large negative shocks occurring in October 1987

[11] In an earlier version of this paper, we also analysed ten size-based portfolios of UK stocks. Qualitatively, the results relating to the relative performance of the different parametric ARCH models investigated in that version were largely similar to those obtained for the FT All Share Index with no apparent differences across portfolios. Hence, these were not investigated further and reported in this version of the paper.

[12] Several other forms of the conditional density function have been suggested in the literature, e.g. the Student-*t* distribution in Bollerslev (1987), the normal-Poisson mixture distribution in Jorion (1988), the power exponential distribution in Baillie and Bollerslev (1989), the normal-lognormal mixture distribution in Hsieh (1989), and the generalized exponential distribution in Nelson (1990). The results (on other financial assets) of McCurdy and Morgan (1987), Milhøj (1987), Hsieh (1989) and Baillie and Bollerslev (1989) suggest that the conditional normality assumption does not capture all the observed skewness and excess kurtosis. In this context, the use, for example, of likelihood ratio tests is not entirely appropriate. We are grateful to an anonymous referee for pointing this out to us.

[13] The out-of-sample analysis could better highlight the usefulness of the models in portfolio management if we re-estimated each ARCH model every day and then did the next day analysis. We are grateful to an anonymous referee for pointing this out to us. However, given the long data period, such an exercise would have been very computer intensive.

Table I. Model estimates

Continuously compounded returns on the FTSE100 index, R_t, are pre-filtered for day of the week, holiday and serial correlation effects using equations (1A) and (1B) below:

$$R_t = \lambda_1 MON_t + \lambda_2 TUE_t + \lambda_3 WED_t + \lambda_4 THU_t + \lambda_5 FRI_t + \lambda_6 HOL_t + \nu_t \tag{1A}$$

$$\hat{\nu}_t = \varphi_0 + \sum_{l=1}^{10} \varphi_l \hat{\nu}_{t-l} + \varepsilon_t \tag{1B}$$

The residuals from equation (1B) are labelled as linearly filtered raw returns, and form the data for estimating eight specializations of the generalized augmented GARCH(1,1) process where ε_t is assumed to be conditionally normally distributed with mean zero and time varying variance of h_t given by:

$$h_t = \begin{cases} |\lambda\phi_t - \lambda + 1|^{1/\lambda} & \text{if } \lambda \neq 0, \\ \exp(\phi_t - 1) & \text{if } \lambda = 0 \end{cases} \qquad \begin{aligned} \xi_{1,t-1} &= \alpha_1 + \alpha_2|\varepsilon_t - c|^\delta + \alpha_3 \max(0, c - \varepsilon_t)^\delta \\ \xi_{2,t-1} &= \alpha_4 f(|\varepsilon_t - c|; \delta) + \alpha_5 f(\max(0, c - \varepsilon_t); \delta) \end{aligned}$$

$$\phi_t = \alpha_0 + \phi_{t-1}\xi_{1,t-1} + \xi_{2,t-1} \qquad \text{where } f(z;\delta) = (z^\delta - 1)/\delta \text{ for any } z \geqslant 0$$

Models are estimated for three sub-periods: 1971–80; 1981–90; and 1991–October 1997. The table reports estimated coefficients and standard errors (in parentheses). Superscripts next to the standard error denote the level of significance ([a] denotes 10%, [b] denotes 5% and [c] denotes 1%). Values are also displayed for the log likelihood function (Log-L). For the GJR-GARCH and NGARCH models, significance of the likelihood ratio test (LRT) of the restriction in the nested LGARCH model is denoted by the superscript next to the Log-L values; for the TGARCH model, the LRT test is for restriction in the nested TS-GARCH model.

	LGARCH	MGARCH	EGARCH	GJR-GARCH	NGARCH	VGARCH	TS-GARCH	TGARCH
Panel A: 1971–80								
α_0	0·011	1·27	0·065	0·01	0·01	0·01	0·035	0·044
	(0·002)[c]	(0·017)[c]	(0·008)[c]	(0·002)[c]	(0·002)[c]	(0·003)[c]	(0·003)[c]	(0·005)[c]
α_1	0·922	0·162	0·991	0·926	0·926	0·992	0·916	0·923
	(0·005)[c]	(0·006)[c]	(0·002)[c]	(0·005)[c]	(0·005)[c]	(0·001)[c]	(0·005)[c]	(0·005)[c]
α_2	0·07			0·056	0·066	0·136	0·09	0·072
	(0·007)[c]			(0·01)[c]	(0·007)[c]	(0·009)[c]	(0·007)[c]	(0·008)[c]
α_3				0·021				0·025
				(0·011)[a]				(0·008)[c]
α_4			0·135					
			(0·015)[c]					
α_5			0·042					
			(0·015)[c]					
c					0·161	0·119		
					(0·073)[b]	(0·031)[c]		
Log-L	−3664·52	−3968·51	−3671·43	−3663·05[a]	−3662·51[b]	−3672·11	−3673·76	−3670·58[b]
Panel B: 1981–1990								
α_0	0·048	0·805	0·158	0·05	0·061	0·053	0·082	0·129
	(0·007)[c]	(0·013)[c]	(0·009)[c]	(0·007)[c]	(0·008)[c]	(0·008)[c]	(0·007)[c]	(0·009)[c]
α_1	0·826	0·128	0·928	0·828	0·782	0·917	0·814	0·817
	(0·016)[c]	(0·004)[c]	(0·01)[c]	(0·017)[c]	(0·018)[c]	(0·009)[c]	(0·014)[c]	(0·016)[c]
α_2	0·108			0·058	0·11	0·114	0·141	0·079
	(0·011)[c]			(0·014)[c]	(0·013)[c]	(0·011)[c]	(0·007)[c]	(0·013)[c]

Table continued over page

Table I (Continued)

	LGARCH	MGARCH	EGARCH	GJR-GARCH	NGARCH	VGARCH	TS-GARCH	TGARCH
α_3				0·085 (0·012)[c]				0·092 (0·009)[c]
α_4			0·153 (0·023)[c]					
α_5			0·136 (0·017)[c]					
c					0·46 (0·07)[c]	0·338 (0·059)[c]		
Log-L	−3125·86	−3316·19	−3109·99	−3115·43[c]	−3108·62[c]	−3123·19	−3132·64	−3113·92[c]
Panel C: 1991–October 1997								
α_0	0·014 (0·004)[c]	0·258 (0·02)[c]	0·074 (0·011)[c]	0·01 (0·003)[c]	0·01 (0·003)[c]	0·003 (0·004)	0·022 (0·004)[c]	0·044 (0·006)[c]
α_1	0·899 (0·017)[c]	0·049 (0·007)[c]	0·982 (0·005)[c]	0·923 (0·015)[c]	0·906 (0·015)[c]	0·977 (0·006)[c]	0·929 (0·012)[c]	0·939 (0·011)[c]
α_2	0·071 (0·01)[c]			0·027 (0·009)[c]	0·052 (0·009)[c]	0·029 (0·005)[c]	0·067 (0·009)[c]	0·031 (0·01)[c]
α_3				0·061 (0·014)[c]				0·051 (0·01)[c]
α_4			0·061 (0·02)[c]					
α_5			0·095 (0·019)[c]					
c					0·659 (0·136)[c]	0·703 (0·165)[c]		
Log-L	−1742·34	−1812·08	−1729·95	−1735·17[c]	−1730·84[c]	−1741·97	−1738·34	−1729·41[c]

and October 1989. During that period, the coefficient is around 0·8 for most models with the lowest value of 0·782 being for the NGARCH model.

Strong asymmetry in the behaviour of volatility is consistently revealed from the coefficients on the respective ‘leverage terms’, both across models and time periods. The parameters proxying for asymmetry in models that recognize the asymmetric behaviour of volatility are highly significant in each and every case. However the temporal pattern of the relative size of the coefficients differs across specifications. In Table I, the parameters indicating asymmetry in different models are $\alpha_4\alpha_5$ for EGARCH, α_3 for GJR-GARCH and TGARCH, and c for NGARCH and VGARCH. The values of these parameters in Table I show that for the EGARCH, GJR-GARCH and T-GARCH models, the effect of asymmetry was much stronger in the middle period, weakening again in the last period but falling not as far as the levels in the first period. In contrast, the NGARCH and VGARCH specifications suggest that the impact of asymmetry grew greater across successive periods, being strongest in the final period.

While the sign and significance of all ARCH parameters within each model are consistent throughout the three periods, the observed differences in magnitudes, such as those highlighted in the previous paragraph, raise an important question. Do these differences impinge on the usefulness of these models to predict volatility out-of-sample? We address this important question as part of our evaluation of these eight parameterizations in Section 3.3 below.

To further assess the appropriateness of these extensions to the benchmark LGARCH model, Table I also reports the values for the log-likelihood function (Log-L) for each estimation. While formal likelihood ratio tests are not possible where these models are non-nested, a comparison of the log-likelihood function value is nevertheless fairly instructive. For instance, in each period, the value for the MGARCH model is substantially more negative than for all the other models. Within the set of models which permit asymmetry, the maximum log-likelihood value is not completely consistent intertemporally — the maximum is achieved by NGARCH for the first two periods, but TGARCH performs 'best' in the final period. However, VGARCH does 'worst' in each of the three periods.

Clearly, the GARCH model is nested within both the GJR-GARCH and NGARCH models. Similarly the TS-GARCH model is nested within the T-GARCH model. Therefore, formal likelihood ratio tests (LRT) are conducted in these cases. Where applicable, Table I reports the significance of the likelihood ratio tests of the respective restriction in the nested models by the superscript next to the Log-L values for the unrestricted model. In all cases, the restriction is rejected. For the latter two periods, the likelihood ratio tests all reject at p-values less than 1%. However, in the first period, the test of LGARCH against GJR-GARCH is only marginally rejected at 10% while the other two tests still reject at 5%.

3.3 Model Evaluation

Benchmarks and diagnostics

To evaluate the performance of our different models, several benchmarks and diagnostics are examined. Table II reports the following tests and statistics for each of the eight GARCH specifications during each of the three estimation periods:

- Root mean squared error in conditional variance forecasts (RMSE)
- West–Cho (1995) test for equality of the conditional variance forecast error (WC)
- Diebold–Mariano (1995) test for equality of the conditional variance forecast error (DM)
- Skewness and kurtosis in standardized residuals;
- LM tests for serial correlation up to ten lags in the levels of the standardized residuals (LM(levels))
- LM tests for serial correlation up to ten lags in the squares of the standardized residuals (LM(squares))
- Engle–Ng (1993) news impact curve tests of the standardized residuals (EN)
- Pagan–Sabau (1987) consistency tests of the squared residuals (PS)
- Pagan–Sabau (1987) consistency tests of the squared residuals using the logarithmic form (PS(log)).

The RMSE, West–Cho and Diebold–Mariano tests are computed using the conditional variance forecast errors, $\varepsilon_t^2 - h_{jt}$, where ε_t is the linearly filtered raw return in period t, and h_{jt} is the period t conditional variance obtained from the jth parameterization ($j = 1, \ldots, 8$) of the generalized augmented GARCH process. The skewness and kurtosis statistics, the LM tests and the Engle–Ng tests are computed for the standardized residuals, $\varepsilon_t/\sqrt{h_{jt}}$. The Pagan–Sabau tests employ squared residuals, ε_t^2. For benchmarking purposes, corresponding tests are also reported for the linearly filtered raw returns, ε_t, wherever that is applicable. For the last two periods, the table also reports the same set of diagnostics using out-of-sample forecasts. Here the estimated ARCH parameters from the previous period's data are applied to the current period data to

Table II. Diagnostics

This table reports the following diagnostic tests for each of the eight GARCH specifications during each of the three estimation periods — root mean squared error in conditional variance forecasts (RMSE); West–Cho (1995) test for equality of the conditional variance forecast error (WC); Diebold–Mariano (1995) test for equality of the conditional variance forecast error (DM); skewness and kurtosis in standardized residuals; LM tests for serial correlation up to ten lags in the levels of the standardized residuals (LM(levels)); LM tests for serial correlation up to ten lags in the squares of the standardized residuals (LM(squares)); Engle–Ng (1993) news impact curve tests of the standardized residuals (EN); Pagan–Sabau (1987) consistency tests of the squared residuals (PS); Pagan–Sabau (1987) consistency tests of the squared residuals using the logarithm form (PS(log)).

Superscripts next to the test statistic denote the level of significance ([a] denotes 10%, [b] denotes 5% and [c] denotes 1%). Wherever applicable, tests are also reported for the linearly filtered raw returns for benchmark purposes. For the last two periods, the table also reports the same set of diagnostics using out-of-sample forecasts. Here the estimated ARCH parameters from the previous period's data are applied to the current period data to generate out-of-sample volatility forecasts.

For the West–Cho tests, the test for equality of the conditional variance forecast errors across all eight models is reported in the LGARCH column. For the remaining columns, the pairwise test for equality across the LGARCH and the alternative model is reported under the respective column. For the Diebold–Mariano test, the null is that there is no difference between the conditional variance forecast errors for the LGARCH and the alternative model for that column. For the Engle–Ng news impact curve tests, we report robust tests for sign bias, negative size bias, positive size bias and the joint bias test. For the Pagan–Sabau consistency tests, we report the chi-squared joint test that the intercept is zero and the slope is unity (PS χ^2); the R^2 (PS R^2) and the heteroscedasticity adjusted Ljung–Box test of the residuals up to ten lags (PS LB).

	Linearly filtered raw returns	LGARCH	MGARCH	EGARCH	GJR-GARCH	NGARCH	VGARCH	TS-GARCH	TGARCH
Panel A: 1971–80									
RMSE									
in sample		3·245	3·548	3·271	3·26	3·261	3·331	3·257	3·278
WC χ^2									
in sample		13·51[a]	2·53	0·85	0·72	0·61	0·70	1·69	1·15
DM									
in sample			0·47	0·89	0·73	0·87	1·02	1·23	1·51
Skewness									
in sample	0·00	−0·27[c]	−0·13[c]	−0·28[c]	−0·26[c]	−0·26[c]	−0·15[c]	−0·30[c]	−0·29[c]
Kurtosis									
in sample	5·09[c]	1·41[c]	3·23[c]	1·47[c]	1·39[c]	1·41[c]	0·87[c]	1·55[c]	1·57[c]
LM (levels)									
in sample	0·32	14·85	3·55	14·03	14·32	13·98	10·65	14·54	13·63
LM (squares)									
in sample	560·3[c]	7·67	306·9[c]	9·54	7·30	7·44	14·44	8·81	7·88
EN sign bias									
in sample	3·81[c]	−0·84	2·19[b]	−0·73	−0·69	−0·76	−0·68	−0·75	−0·59
EN neg. size bias									
in sample	−8·39[c]	−2·74[c]	−0·99	−2·37[b]	−2·43[b]	−2·42[b]	−2·17[b]	−2·55[b]	−2·08[b]

EN pos. size bias									
in sample	16·13[c]	0·14	4·57[c]	0·44	0·48	0·56	1·07	−0·10	0·44
EN Joint bias									
in sample	293·6[c]	8·23[b]	21·86[c]	6·01	6·38[a]	6·30[a]	5·87	7·58[a]	4·66
PS χ^2									
in sample		0·06	24·50[c]	0·09	0·02	0·01	6·11[b]	0·17	0·17
PS R^2									
in sample		0·199	0·019	0·186	0·192	0·191	0·165	0·193	0·183
PS LB									
in sample		68·00[c]	103·6[c]	86·31[c]	76·81[c]	78·97[c]	162·2[c]	79·02[c]	92·41[c]
PS(log) χ^2									
in sample		945·9[c]	1057·6[c]	949·2[c]	944·1[c]	944·3[c]	945·3[c]	944·6[c]	941·8[c]
PS(log) R^2									
in sample		0·092	0·026	0·09	0·092	0·092	0·088	0·09	0·09
PS(log) LB									
in sample		12·68	105·1[c]	11·54	13·26	13·09	11·93	10·95	11·00
Panel B: 1981–90									
RMSE									
in sample		3·175	3·47	3·112	3·124	3·151	3·276	3·293	3·246
out of sample		3·216	3·433	3·258	3·215	3·229	3·264	3·308	3·302
WC χ^2									
in sample		5·22	1·29	1·32	0·95	1·01	0·99	0·98	0·80
out of sample		63·05[c]	1·42	0·43	0·01	1·64	1·09	0·61	0·60
DM									
in sample			0·03	1·16	−0·72	−0·26	1·13	0·79	1·06
out of sample			−3·25[c]	1·43	−1·23	0·74	−2·03[b]	1·19	1·43
Skewness									
in sample	−1·2[c]	−0·94[c]	−0·88[c]	−0·64[c]	−0·80[c]	−0·69[c]	−0·72[c]	−0·80[c]	−0·62[c]
out of sample		−0·97[c]	−1·22[c]	−0·82[c]	−0·93[c]	−0·91[c]	−0·79[c]	−0·91[c]	−0·84[c]
Kurtosis									
in sample	17[c]	8·65[c]	8·14[c]	4·82[c]	6·93[c]	5·35[c]	6·05[c]	6·32[c]	4·36[c]
out of sample		9·33[c]	12·68[c]	7·02[c]	8·80[c]	8·45[c]	8·40[c]	7·87[c]	7·05[c]
LM (levels)									
in sample	0·10	14·89	3·78	14·39	14·67	14·05	12·40	15·77	14·49
out of sample		17·49[a]	31·27[c]	17·78[a]	17·82[a]	17·65[a]	13·65	16·33[a]	16·91[a]
LM (squares)									
in sample	1027·8[c]	8·72	490·8[c]	7·29	5·69	8·40	26·04[c]	58·59[c]	36·72[c]
out of sample		20·16[b]	539·5[c]	51·72[c]	18·17[a]	23·12[b]	9·95	140[c]	123·5[c]

Table continued over page

Table II (Continued)

	Linearly filtered raw returns	LGARCH	MGARCH	EGARCH	GJR-GARCH	NGARCH	VGARCH	TS-GARCH	TGARCH
EN sign bias									
in sample	−6·32[c]	−0·06	−3·71[c]	0·23	0·39	−0·10	−0·38	−1·28	−0·71
out of sample		−0·26	−0·85	−0·82	−0·09	−0·28	−0·01	−2·13[b]	−1·84[a]
EN neg. size bias									
in sample	−23·19[c]	−1·49	−11·42[c]	−0·52	−0·72	−0·84	−2·24[b]	−3·57[c]	−2·11[b]
out of sample		−2·74[c]	−4·56[c]	−3·90[c]	−2·50[b]	−2·82[c]	−1·49	−6·60[c]	−5·98[c]
EN pos. size bias									
in sample	5·82[c]	−0·63	0·97	0·01	0·01	−0·04	0·16	−0·98	0·04
out of sample		0·01	0·09	0·30	0·21	0·23	−0·10	0·00	0·28
EN Joint bias									
in sample	473·4[c]	5·47	129·4[c]	0·90	1·96	1·04	6·30[a]	16·73[c]	4·81
out of sample		10·66[b]	24·19[c]	17·75[c]	9·12[b]	10·60[b]	3·76	46·53[c]	38·19[c]
PS χ^2									
in sample		0·41	4·37	2·61	0·14	0·24	4·75[a]	1·42	1·24
out of sample		1·06	85·94[c]	0·46	1·26	0·83	13·01[c]	0·23	0·09
PS R^2									
in sample		0·182	0·012	0·281	0·204	0·192	0·247	0·129	0·156
out of sample		0·135	0·03	0·111	0·137	0·129	0·112	0·084	0·087
PS LB									
in sample		668·7[c]	933·2[c]	470·2[c]	631·0[c]	627·6[c]	606·8[c]	704·2[c]	650·1[c]
out of sample		793·6[c]	829·6[c]	820·1[c]	793·7[c]	805·5[c]	878·9[c]	857·6[c]	857·1[c]
PS(log) χ^2									
in sample		972·2[c]	1386·6[c]	957·6[c]	963·9[c]	959·9[c]	987·2[c]	978·3[c]	962·8[c]
out of sample		1072·4[c]	1947·2[c]	1052·8[c]	1070·0[c]	1056·8[c]	1156·3[c]	1080·9[c]	1058·8[c]
PS(log) R^2									
in sample		0·044	0·01	0·043	0·046	0·046	0·044	0·04	0·043
out of sample		0·041	0·006	0·039	0·04	0·04	0·031	0·04	0·039
PS(log) LB									
in sample		9·02	47·57[c]	9·31	9·01	9·02	8·79	11·09	9·63
out of sample		12·86	70·72[c]	10·84	14·47	14·60	22·54[b]	8·24	9·57
Panel C: 1991–October 1997									
RMSE									
in sample		1·125	1·141	1·117	1·118	1·117	1·124	1·12	1·117
out of sample		1·118	1·166	1·107	1·109	1·11	1·118	1·112	1·108

WC χ^2									
in sample		33·49[c]	1·30	1·85	2·21	4·16[b]	0·04	1·05	1·99
out of sample		157·9[c]	13·71[c]	2·34	2·79[a]	2·59	0·01	0·95	1·74
DM									
in sample			0·16	0·74	0·71	0·60	1·10	0·76	0·69
out of sample			−9·31[c]	0·69	0·73	−1·68[a]	−2·36[b]	0·57	−0·21
Skewness									
in sample	0·34[c]	0·04	0·26[c]	0·06	0·06	0·05	0·09	0·06	0·06
out of sample		0·03	0·22[c]	0·04	0·07	0·07	0·06	0·00	0·06
Kurtosis									
in sample	4·42[c]	2·35[c]	3·71[c]	2·31[c]	2·40[c]	2·39[c]	2·62[c]	2·41[c]	2·31[c]
out of sample		1·79[c]	4·33[c]	1·92[c]	2·01[c]	2·02[c]	2·03[c]	1·77[c]	2·01[c]
LM (levels)									
in sample	0·32	1·83	0·64	2·59	2·29	2·59	1·88	2·02	2·66
out of sample		15·97	20·15[b]	16·53[a]	16·73[a]	17·18[a]	16·71[a]	15·30	16·70[a]
LM (squares)									
in sample	60·3[c]	4·12	51·6[c]	5·16	3·74	3·88	5·74	5·48	5·46
out of sample		6·21	36·1[c]	7·22	6·61	7·78	5·11	7·84	9·09
EN sign bias									
in sample	1·33	0·75	1·46	0·69	0·93	0·70	0·81	0·57	0·66
out of sample		1·20	0·79	1·28	1·52	1·26	1·35	1·01	1·24
EN neg. size bias									
in sample	−2·33[b]	0·43	−0·75	0·80	0·92	0·97	0·31	0·16	0·80
out of sample		0·91	1·44	1·92[a]	1·46	1·70[a]	1·42	1·36	2·03[b]
EN pos. size bias									
in sample	4·41[c]	0·45	2·81[c]	0·75	0·74	0·79	1·44	0·50	0·74
out of sample		0·66	−0·39	0·72	1·27	1·28	1·22	−0·13	0·84
EN Joint bias									
in sample	24·8[c]	0·57	8·79[b]	1·25	1·39	1·73	2·23	0·39	1·27
out of sample		1·57	2·36	4·27	3·76	4·87	3·54	2·17	5·06
PS χ^2									
in sample		1·01	6·27[b]	1·39	0·19	0·18	1·87	0·23	1·13
out of sample		14·45[c]	104·3[c]	23·29[c]	20·35[c]	24·00[c]	29·44[c]	18·57[c]	27·19[c]
PS R^2									
in sample		0·032	0·002	0·047	0·043	0·045	0·035	0·039	0·046
out of sample		0·036	0·005	0·053	0·05	0·049	0·037	0·043	0·052
PS LB									
in sample		17·0[a]	1·15	16·6[a]	15·4	16·3[a]	18·9[b]	18·3[a]	17·0[a]
out of sample		14·9	52·3[c]	14·2	15·9	16·8[a]	14·8	15·4	13·6

Table continued over page

Table II (Continued)

	Linearly filtered raw returns	LGARCH	MGARCH	EGARCH	GJR-GARCH	NGARCH	VGARCH	TS-GARCH	TGARCH
PS(log) χ^2									
[c]in sample		679·6[c]	1178·3[c]	673·7[c]	675·2[c]	674·6[c]	681·8[c]	678·4[c]	673·4[c]
out of sample		870·3[c]	1222·0[c]	857·5[c]	872·9[c]	901·5[c]	906·2[c]	859·2[c]	882·4[c]
PS(log) R^2									
in sample		0·026	0·003	0·038	0·033	0·036	0·032	0·031	0·038
out of sample		0·027	0·005	0·035	0·031	0·032	0·028	0·03	0·034
PS(log) LB									
in sample		3·92	30·06[c]	2·47	2·69	2·26	3·43	2·43	2·56
out of sample		8·55	33·32[c]	9·45	8·24	8·12	8·87	9·18	9·85

generate out-of-sample volatility forecasts. In all cases, superscripts next to the test statistic denote the level of significance (where ^{a} denotes 10%, ^{b} denotes 5% and ^{c} denotes 1%).

Linearly filtered raw returns

Statistics for the linearly filtered raw returns ε_t in each period demonstrate the need for, and the potential utility of, the ARCH type models considered in this paper. Evidence of skewness is mixed and period-specific — not significantly different from zero in period 1; significantly negative in period 2; and significantly positive in period 3. Returns are characteristically leptokurtic — tests for excess kurtosis consistently reject the null at very high significance levels in each of our test periods.

The non-significance of the LM tests for serial correlation up to ten lags in the linearly filtered raw returns verifies that the procedure we use to remove calendar and thin trading effects from actual returns is successful. In contrast, the large and highly significant LM test statistics for serial correlation up to ten lags in the squares of linearly filtered raw returns indicates that there is significant and substantial predictability in volatility explained by the ARCH parameterizations which we employ. This evidence is weaker for the third period, yet the tests are still significant with p-values less than 1%.

The Engle–Ng news impact curve tests on the linearly filtered raw returns consistently reveal that the behaviour of squared returns is systematically related to the sign and size of the previous return. Again, this evidence is weaker in the third period, where there is no evidence of sign bias and the test for negative size bias is significant at 5% only. Overall this evidence suggests that models that permit asymmetry are likely to outperform those that ignore this pervasive characteristic of financial market price behaviour.

The strength of this evidence of kurtosis in linearly filtered raw returns, on serial correlation in the square of these returns, and on the sign and size biases is consistent with the previous research which motivated the original ARCH models and led to the continued development and use of ARCH specifications. Not only do these results for the linearly filtered raw returns clearly motivate the ARCH family of models, the size of the various statistics form an objective benchmark. They facilitate assessing the degree of success that the various models have in modelling these important characteristics of the behaviour of the UK FT-All Share Price Index.

Different GARCH parameterizations

In assessing the reported RMSE's of the conditional variance forecasts h_{jt} ($j = 1, \ldots, 8$) from the different GARCH parameterizations, it needs to be borne in mind that the levels are substantially lower in the final period, overall, dropping from around 3·2 on average to around 1·1. Within each test period, the in-sample RMSE statistics and the out-of-sample RMSE statistics are both quite similar across the eight models, except for the MGARCH model which are always highest, and usually by a relatively large margin. For example, in period 1, the RMSE for the MGARCH is 3·548 compared to a range from 3·245 (LGARCH) to 3·31 (VGARCH) for the other seven models. Rankings are inconsistent, both as we move from in-sample to out-of-sample, and as we move from period to period.

To compare the predictability of these models more formally, we conduct a series of tests used in West and Cho (1995). These examine whether there are significant differences in the conditional variance forecast errors $\varepsilon_t^2 - h_{jt}$ ($j = 1, \ldots, 8$) across the model specifications. The West–Cho statistic caters for serial correlation with the lag length automatically chosen by a data-dependent rule. In the LGARCH column, a test of equality across all eight specifications is reported. Both out-of-sample tests reject the null at the 1% significance level. Evidence is mixed in-sample — the test does not reject in the middle period, marginally rejects for the earliest period but strongly rejects in the final period.

To examine this issue further, we conduct pairwise tests of the equality of the conditional variance forecast errors $\varepsilon_t^2 - h_{jt}$ for the benchmark LGARCH model ($j = 1$) and each of the alternative models ($j = 2, \ldots, 8$). These tests are reported in the respective column for each alternative model. For instance, in the third period, the West–Cho statistic for equality between the conditional variance forecast errors of the LGARCH and of the NGARCH model is 4·16 which is significant at the 5% level. No other individual West–Cho test in-sample is significant in any period. Nor are there any significant rejections out-of-sample for the middle period. The West–Cho tests do, however, distinguish the out-of-sample performance of the models estimated in the second period for the third period. In that case, the conditional variance forecast errors are significantly different between the LGARCH model and the MGARCH, GJR-GARCH and NGARCH models. Of course, these parameter estimates are influenced by the impact of the crash and mini-crash.

Another comparative test of forecast error behaviour is due to Diebold and Mariano (1995). Here we form the time series of the difference between the conditional variance forecast errors $\varepsilon_t^2 - h_{jt}$ for the LGARCH model ($j = 1$) and the alternative model ($j = 2, \ldots, 8$) in order to test the null that the time series mean is zero. In computing the standard error of the test statistic we allow for serial correlation up to a lag of ten. Under each column for the non-LGARCH models, the Diebold–Mariano test statistic is reported for the comparison between the LGARCH and that particular alternative model. None of the mean differences are found to be significantly non-zero for any model or any time period on an in-sample basis. There is some evidence of differences occurring out-of-sample. For example, in respect of the LGARCH and MGARCH comparison, the Diebold–Mariano test statistic for the second (third) out-of-sample period is −3·25 (−9·31). Similarly, for the LGARCH and VGARCH comparison, the Diebold–Mariano test statistic for the second (third) out-of-sample period is −2·03 (−2·26), rejecting equality at 5% level.

Tests for skewness in the standardized residuals $\varepsilon_t/\sqrt{h_{jt}}$ ($j = 1, \ldots, 8$) provide evidence of substantial negative skewness in the first two time periods. This skewness occurs consistently across all specifications within each time period. Excess kurtosis exists in the standardized residuals for all the models. Although the kurtosis coefficients are always lower in the standardized residuals of the different GARCH models than in the linearly filtered raw returns, they remain significant in all models in all time periods. Hence, no specification demonstrates superior ability in meeting this challenge. These tests for skewness and excess kurtosis in the standardized residuals show that the ARCH models analysed in this paper are only able to capture some, but not all, of the observed skewness and excess kurtosis. Our evidence in regard to UK stock returns is therefore consistent with earlier evidence in other markets and financial assets, e.g. McCurdy and Morgan (1987), Milhøj (1987), Hsieh (1989) and Baillie and Bollerslev (1989).

As was the case for linearly filtered raw returns, the in-sample LM tests for serial correlation up to ten lags in the standardized residuals $\varepsilon_t/\sqrt{h_{jt}}$ ($j = 1, \ldots, 8$) do not reject the null in any case. Conversely, out-of-sample tests show that there exists substantial predictability when the parameters used to purge out the effects of calendar seasonalities and thin trading on actual returns are not updated. This suggests intertemporal instability in these parameters.

The results of the LM tests for serial correlation up to ten lags in the squares of the standardized residuals indicate that not all the volatility specifications are able to fully capture this aspect of the heteroscedastic behaviour of financial risk. While the LM test statistics, both in-sample and out-of-sample, are always substantially lower for the standardized residuals of the different GARCH models than for the linearly filtered raw returns, a few still remain significant. In-sample LM tests are significant at the 1% level in periods 1 and 3 for the MGARCH model only; in period 2 they are significant for MGARCH, VGARCH, TGARCH and TS-GARCH. Out-of-sample tests are significant in the second period for all models, except the VGARCH model. Conversely, none but MGARCH are significant in the later out-of-sample period.

For the Engle–Ng news impact curve tests, we report robust tests for sign bias, negative size bias, positive size bias and the joint bias test. The sign-bias test indicates whether *positive* and *negative* return innovations have different impact on the volatility not predicted by the null volatility model. The positive-size-bias test indicates the difference in impact between *large* and *small* positive innovations on volatility not explained by the null volatility model. The negative size bias test indicates the difference in impact between *large* and *small* negative innovations on volatility not predicted by the null volatility model. Finally, we also conduct the robust chi-squared joint test for all three forms of bias.

Overall these tests show that as a group, all the ARCH parameterizations go a long way towards capturing the new structure in returns. For the three individual tests, as well as the joint test, the test statistics are often insignificant. They are invariably substantially lower for the standardized residuals of the different GARCH models than for the linearly filtered raw returns across all specifications and in each time period.

Virtually none of the sign bias tests or positive size bias tests are significant in-sample. With the sign bias tests, exceptions are the MGARCH model in the first two periods; for the positive size bias tests, exceptions are the MGARCH model in the first and last periods. In contrast, there is some evidence of negative size bias in the first two periods, although none in the last. The pattern of incidence of negative size bias is inconsistent across models, both in-sample/out-of-sample and over time. This suggests that the appropriate choice of model is period-specific.

Our final set of diagnostics are the consistency tests in Pagan and Sabau (1987, unpublished manuscript). These are obtained from a regression of the squared residuals, ε_t^2, on the conditional variance from the *j*th model, h_{jt}, for that period. We report the chi-squared joint test that the intercept is zero and the slope is unity (PS $\chi 2$); the R^2 from the regression (PS R2); and the heteroscedasticity adjusted Ljung–Box test of the regression residuals up to ten lags (PS LB). We also report the same results for the logarithmic version of this regression, (PS(log)). Both versions are informative since the linear model assumes a quadratic loss function whereas the logarithmic specification infers a proportional one. The regression with logs implies that mistakes in predicting small variances are given more weight than in the linear regression.[14]

With the linear specification for the Pagan–Sabau regression, joint tests for bias and efficiency do not reject the null in-sample (except for the MGARCH and the VGARCH models in some periods) showing that the conditional variance is an unbiased predictor of the actual variance. However, there is evidence of serial correlation in the residuals, suggesting additional persistence in volatility that is not captured by these parametric ARCH models. In contrast, with the

[14] If volatility estimates are for option pricing, the correct loss function should be given by the gamma of the option, i.e. the derivative of the call price with respect to the standard deviation. We are grateful to an anonymous referee for pointing this out to us.

logarithmic specification for the Pagan–Sabau regression, the joint tests for bias and efficiency strongly reject the null for all models in each of the periods analysed. Conversely, there is little evidence of serial correlation in the residuals, except for the MGARCH parameterization.

Results with respect to the R^2 obtained from these regressions are mixed. Generally, R^2's tend to take on higher values in the middle data period, both in-sample and out-of-sample; tend to take higher values in-sample as opposed to out-of-sample across all periods; and for the linear specification relative to the logarithmic specification. The variation across models is not consistent in so far as relative rankings can change across time periods, and in moving from in-sample inferences to out-of-sample inferences within a common period.

4. SUMMARY AND CONCLUSIONS

This paper presents empirical evidence on the relative effectiveness of eight different parametric ARCH models in describing daily stock returns. Twenty-seven years of UK daily data on a broad-based value weighted stock index are investigated for the period 1971–97. In addition to in-sample comparison, the *ex ante* sample usefulness of these models, for example for portfolio selection and asset pricing decisions, is evaluated by quantifying their out-of-sample predictive ability. Several interesting results are documented. First, estimates for the ARCH parameters across all models in each of the periods are highly significant. Second, ARCH models with conditionally normal density functions are able to capture some, but not all, of the observed skewness and excess kurtosis. Third, the conditional variance is an unbiased predictor of the actual variance for all models except the MGARCH and VGARCH; but there is additional persistence in volatility that is not captured by the models. Fourth, modelling asymmetry is important since the parameters proxying for asymmetry in models that recognize the asymmetric behaviour of volatility are all highly significant. Fifth, the 'performance' of the various parameterizations is often fairly similar; however, an obvious exception is the multiplicative GARCH model that performs qualitatively differently on several dimensions of performance. Sixth, outperformance of any model(s) is not consistent across different sub-periods of the sample, suggesting that the optimal choice of a model is period-specific. Finally, outperformance is not always consistent as we change from in-sample inferences to out-of-sample inferences within the same period, suggesting that model parameters are not always stable across time.

ACKNOWLEDGEMENTS

Pradeep Yadav is grateful to the UK Economic and Social Research Council (ESRC) for financial support through grant #R000233234, and to the Scottish Institute for Research in Investment and Finance (SIRIF) for infrastructural support. This research was completed while Geoffrey Loudon was Visiting Scholar at the University of Strathclyde. The authors thank the Co-Editor Professor Mark Watson, three anonymous referees, the participants of the 1994 Western Finance Association Annual Conference, the participants of the 1994 INQUIRE Europe Annual Conference, and the participants of the 1995 World Econometric Congress for helpful comments and suggestions. The authors remain responsible for all errors.

REFERENCES

Akgiray, V. (1989), 'Conditional heteroskedasticity in time series of stock returns: evidence and forecast', *Journal of Business*, **62**, 55–80.

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.

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

Black, F. (1976), 'Studies in stock price volatility changes', *Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section*, American Statistical Association, 177–181.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics*, **31**, 307–327.

Bollerslev, T. (1987), 'A conditionally heteroskedastic time series model for speculative prices and rates of return', *Review of Economics and Statistics*, **69**, 542–547.

Bollerslev, T., R. Y. Chou and K. F. Kroner (1992), 'ARCH modelling in finance: a review of the theory and empirical evidence', *Journal of Econometrics*, **52**, 5–59.

Bollerslev, T. and I. Domowitz (1991), 'Price volatility, spread variability and the role of alternative market mechanisms', *Review of Futures Markets*, **10**(1).

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.

Cohen, K. J., S. F. Maier, R. A. Schwartz and D. K. Whitcomb (1986), *The Microstructure of Securities Markets*, Prentice Hall, Englewood Cliffs, NJ.

Corhay, A. and A. T. Rad (1991), 'Conditional heteroscedasticity in stock returns: international evidence', Faculty of Economics Research Memorandum 91-014, University of Limburg.

Diebold, F. X. and R. S. Mariano (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics*, **13**, 253–263.

Ding, Z., C. W. Granger and R. Engle (1993), 'A long memory property of stock market returns and a new model', *Journal of Empirical Finance*, **1**, 83–106.

Duan, J. C. (1997), 'Augmented GARCH(p,q) process and its diffusion limit', *Journal of Econometrics*, **79**, 97–127.

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

Engle, R. F. and C. Mustafa (1992), 'Implied ARCH models from option 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.

Fama, E. F. (1965), 'The behaviour of stock market prices', *Journal of Business*, **38**, 34–105.

Fama, E. F. (1991), 'Efficient capital markets II', *Journal of Finance*, **46**, 1575–1617.

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

Frennberg, P. and B. Hansson (1992), 'An Evaluation of Alternative Models for Predicting Stock Volatility', European Finance Association Annual Meeting 1992.

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

Glosten, L. R., R. Jagannathan and D. Runkle (1993), 'Relationship between the expected value and the volatility of the nominal excess return on stocks', *Journal of Finance*, **48**, 1779–1801.

Higgins, M. L. and A. K. Bera (1992), 'A class of nonlinear ARCH models', *International Economic Review*.

Hsieh, D. A. (1989), 'Modelling heteroskedasticity in daily foreign exchange rates', *Journal of Business and Economic Statistics*, **7**, 307–317.

Jorion, P. (1988), 'On jump processes in the foreign exchange and stock markets', *Review of Financial Studies*, **1**, 427–445.

Kearns, P. and A. R. Pagan (1993), 'Australian stock market volatility', *The Economic Record*, **69**, 163–178.

Ljung, G. M. and G. E. P. Box (1978), 'On a measure of lack of fit in time series models', *Biometrika*, **65**, 297–303.
Mandelbrot, B. (1963), 'The variation of certain speculative prices', *Journal of Business*, **36**, 394–419.
McCurdy, T. H. and I. Morgan (1987), 'Tests of the Martingale hypothesis for foreign currency futures with time varying volatility', *International Journal of Forecasting*, **3**, 131–148.
Milhøj, A. (1987), 'A conditional variance model for daily observations of an exchange rate', *Journal of Business and Economic Statistics*, **5**, 99–103.
Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: a new approach', *Econometrica*, **59**, 347–370.
Pagan, A. R. and G. W. Schwert (1990), 'Alternative models for conditional stock volatility', *Journal of Econometrics*, **45**, 267–290.
Pantula, S. G. (1986), 'Comment: modelling the persistence of conditional variances', *Econometric Reviews*, **5**, 71–73.
Poon, S. H. and S. J. Taylor (1992), 'Stock returns and volatility: an empirical study of the UK stock market', *Journal of Banking and Finance*, **16**, 37–59.
Poterba, J. and L. Summers (1986), 'The persistence of volatility and stock market fluctuations', *American Economic Review*, **76**, 1142–1151.
Rabemananjara, R. and J-M Zakoian (1993), 'Threshold ARCH models and asymmetries in volatility', *Journal of Applied Econometrics*, **8**, 31–50.
Schwert, G. W. (1989), 'Why does stock market volatility change over time?', *Journal of Finance*, **44**, 1115–1153.
Schwert, G. W. (1990), 'Stock volatility and the crash of 87', *Review of Financial Studies*, **3**, 77–102.
Sentana, E. (1991), 'Quadratic ARCH models: a potential re-interpretation of ARCH models', Financial Markets Group Discussion Paper No. 122, London School of Economics.
Taylor, S. (1986), *Modelling Financial Time Series*, Wiley, New York.
West, K. D. and D. Cho (1995), 'The predictive ability of several models of exchange rate volatility', *Journal of Econometrics*, **69**, 367–391.
Yadav, P. K. and P. F. Pope (1992), 'Intraweek and intraday seasonalities in stock market risk premia: cash vs futures', *Journal of Banking and Finance*, **16**, 233–270.
Yadav, P. K., P. F. Pope and K. Paudyal (1994), 'Threshold autoregressive modelling in finance: the pricing of equivalent assets', *Mathematical Finance*, **4**, 205–221.
Zakoian, J. M. (1994), 'Threshold heteroskedastic models', *Journal of Economic Dynamics and Control*, **18**, 931–955.

