APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY
Appl. Stochastic Models Bus. Ind., 2005; **21**:55–82
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asmb.551

Predictability and model selection in the context of ARCH models

Stavros Degiannakis[1] and Evdokia Xekalaki[1,2,*,†]

[1] *Department of Statistics, Athens University of Economics and Business, Athens, Greece*
[2] *Department of Statistics, Trinity College, Dublin, Ireland*

SUMMARY

Most of the methods used in the ARCH literature for selecting the appropriate model are based on evaluating the ability of the models to describe the data. An alternative model selection approach is examined based on the evaluation of the predictability of the models in terms of standardized prediction errors. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: ARCH models; model selection; predictability; correlated gamma ratio distribution; standardized prediction error criterion

1. INTRODUCTION

ARCH models have widely been used in financial time series analysis, particularly in analysing the risk of holding an asset, evaluating the price of an option, forecasting time-varying confidence intervals and obtaining more efficient estimators under the existence of heteroscedasticity.

In the recent literature, numerous parametric specifications of ARCH models have been considered for the description of the characteristics of financial markets. In the linear ARCH(q) model, originally introduced by Engle [1], the conditional variance is postulated to be a linear function of the past q squared innovations. Bollerslev [2] proposed the generalized ARCH, or GARCH(p,q), model, where the conditional variance is postulated to be a linear function of both the past q squared innovations and the past p conditional variances. Nelson [3] proposed the exponential GARCH, or EGARCH, model. The EGARCH model belongs to the family of asymmetric GARCH models, which capture the phenomenon that negative returns predict higher volatility than positive returns of the same magnitude. Other popular asymmetric models are the GJR model of Glosten *et al.* [4], the threshold GARCH, or TARCH, model, introduced by Zakoian [5] and the quadratic ARCH, or QGARCH, model, introduced by Sentana [6]. ARCH models go by such exotic names as AARCH, NARCH, PARCH, PNP-ARCH and

*Correspondence to: Evdokia Xekalaki, Department of Statistics, Athens University of Economics and Business and Department of Statistics, Trinity College, Dublin, Ireland.
† E-mail: exek@aueb.gr

Received 1 January 2003
Revised 8 October 2004
Accepted 4 November 2004

STARCH among others. For a comprehensive review of the literature on such models, the interested reader is referred to Degiannakis and Xekalaki [7].

The richness of the family of parametric ARCH models certainly complicates the search for the true model, and leaves quite a bit of arbitrariness in the model selection stage. The problem of selecting the model that describes best the movement of the series under study is, therefore, of practical importance.

The aim of this paper is to develop a model selection method based on the evaluation of the predictability of the ARCH models. In Section 2 of the paper, the ARCH process is presented. Section 3 provides a brief description of the methods used in the literature for selecting the appropriate model based on evaluating the ability of the models to describe the data. In Section 4, Xekalaki *et al.*'s [8] model selection method based on a standardized prediction error criterion is examined in the context of ARCH models. In Section 5, the suggested model selection method is applied using return data for the Athens stock exchange (ASE) index over the period August 30th, 1993 to November 4th, 1996, while, in Section 6, a selection method based on the ability of the models describing the data is investigated. Finally, in Section 7, a brief discussion of the results is provided.

2. THE ARCH PROCESS

Let $\{y_t(\theta)\}_{t\geqslant 1}$ refer to the univariate discrete time real-valued stochastic process to be predicted (e.g. the rate of return of a particular stock or market portfolio from time $t-1$ to t) where θ is a vector of unknown parameters and $E(y_t(\theta)|I_{t-1}) \equiv E_{t-1}(y_t(\theta)) \equiv \mu_t(\theta)$ denotes the conditional mean given the information set available at time $t-1$, I_{t-1}. The innovation process for the conditional mean, $\{\varepsilon_t(\theta)\}_{t\geqslant 1}$, is then represented by $\varepsilon_t(\theta) = y_t(\theta) - \mu_t(\theta)$ with corresponding unconditional variance $V(\varepsilon_t(\theta)) = E(\varepsilon_t^2(\theta)) \equiv \sigma^2(\theta)$, zero unconditional mean and $E(\varepsilon_t(\theta)\varepsilon_s(\theta)) = 0$, $\forall t \neq s$. The conditional variance of the process given I_{t-1} is defined by $V(y_t(\theta)|I_{t-1}) \equiv V_{t-1}(y_t(\theta)) \equiv E_{t-1}(\varepsilon_t^2(\theta)) \equiv \sigma_t^2(\theta)$. Since investors would know the information set I_{t-1} when they make their investment decisions at time $t-1$, the relevant expected return to the investors and volatility are $\mu_t(\theta)$ and $\sigma_t^2(\theta)$, respectively.

An ARCH process, $\{\varepsilon_t(\theta)\}_{t\geqslant 1}$, can be presented as

$$\begin{gathered} y_t(\theta) = x_t'\beta + \varepsilon_t(\theta) \\ \varepsilon_t(\theta) = z_t\sigma_t(\theta) \\ z_t \overset{\text{i.i.d}}{\sim} f[E(z_t) = 0, V(z_t) = 1] \\ \sigma_t^2(\theta) = g(\sigma_{t-1}(\theta), \sigma_{t-2}(\theta), \ldots; \varepsilon_{t-1}(\theta), \varepsilon_{t-2}(\theta), \ldots; v_{t-1}, v_{t-2}, \ldots) \end{gathered} \tag{1}$$

where x_t is a $k \times 1$ vector of endogenous and exogenous explanatory variables included in the information set I_{t-1}, β is a $k \times 1$ vector of unknown parameters, $f(.)$ is the density function of z_t, $\sigma_t(\theta)$ is a time-varying, positive and measurable function of the information set at time $t-1$, v_t is a vector of predetermined variables included in I_t, and $g(.)$ is a linear or non-linear functional form. By definition, $\varepsilon_t(\theta)$ is serially uncorrelated with mean zero, but with a time-varying conditional variance equal to $\sigma_t^2(\theta)$. The standard ARCH models assume that $f(.)$ is the density function of the normal distribution. Bollerslev [9] proposed using the student t distribution with an estimated kurtosis regulated by the degrees of freedom parameter. Nelson [3] proposed the use of the generalized error distribution [10, 11], which is also referred to as the exponential power distribution. Other distributions, that have been employed, include the generalized t distribution [12], the normal Poisson mixture distribution [13], the normal lognormal mixture [14], and a serially dependent mixture of normally distributed variables [15] or student t distributed variables [16]. In the sequel, for notational convenience, no explicit indication of the dependence on the vector of parameters, θ, is given when obvious from the context.

Let us assume that the conditional mean, $\mu_t = E(y_t|I_{t-1})$, can be adequately described by a κth-order autoregressive [AR(κ)] model:

$$y_t = c_0 + \sum_{i=1}^{\kappa}(c_i y_{t-i}) + \varepsilon_t \tag{2}$$

Usually, the conditional mean is either the overall mean or a first-order autoregressive process. Theoretically, the AR(1) process allows for the autocorrelation induced by discontinuous (or non-synchronous) trading in the stocks making up an index [17, 18]. According to Campbell *et al.* [19], 'the non-synchronous trading arises when time series, usually asset prices, are taken to be recorded at time intervals of a fixed length when in fact they are recorded at time intervals of other, possible irregular lengths'. The Scholes and Williams model suggests the first-order moving average process for index returns, while the Lo and MacKinlay model suggests an AR(1) form. Higher orders of the autoregressive process are considered in order to investigate if they are adequate to produce more accurate predictions.

Engle [1] introduced the original form of $\sigma_t^2 = g(.)$ as a linear function of the past q squared innovations:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q}(a_i\varepsilon_{t-i}^2) \tag{3}$$

For the conditional variance to be positive, the parameters must satisfy $\alpha_0 > 0$, $a_i \geqslant 0$, for $i = 1, \ldots, q$. 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 [2] proposed the generalized ARCH, or GARCH(p,q), model:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q}(a_i\varepsilon_{t-i}^2) + \sum_{i=1}^{p}(b_i\sigma_{t-i}^2) \tag{4}$$

where $\alpha_0 > 0$, $a_i \geqslant 0$, for $i = 1, \ldots, q$, and $b_i \geqslant 0$, for $i = 1, \ldots, p$. Note that even though the innovation process for the conditional mean is serially uncorrelated, it is not independent through time. The innovations for the variance are denoted as

$$E_t(\varepsilon_t^2) - E_{t-1}(\varepsilon_t^2) = \varepsilon_t^2 - \sigma_t^2 \equiv v_t \tag{5}$$

The innovation process $\{v_t\}$ is a martingale difference sequence in the sense that it cannot be predicted from its past. However, its range may depend upon the past, making it neither serially independent nor identically distributed.

The GARCH(p,q) model successfully captures several characteristics of financial time series, such as thick-tailed returns and volatility clustering first noted by Mandelbrot [20]: '...large changes tend to be followed by large changes of either sign, and small changes tend to be followed by small changes...'. On the other hand, the GARCH structure imposes important limitations. The variance only depends on the magnitude and not on the sign of ε_t, which is somewhat at odds with the empirical behaviour of stock market prices where a *leverage effect* may be present. The term *leverage effect*, first noted by Black [21], refers to the tendency for changes in stock returns to be negatively correlated with changes in returns volatility, i.e. volatility tends to rise in response to *bad news* ($\varepsilon_t < 0$), and to fall in response to *good news* ($\varepsilon_t > 0$).

In order to capture the asymmetry exhibited by the data, a new class of models was introduced, termed the *asymmetric ARCH models*. The most popular model proposed to capture the asymmetric effects is Nelson's [3] exponential GARCH, or EGARCH(p,q), model:

$$\ln(\sigma_t^2) = a_0 + \sum_{i=1}^{q} \left(a_i \left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| + \gamma_i \left(\frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right) \right) + \sum_{i=1}^{p} (b_i \ln(\sigma_{t-i}^2)) \tag{6}$$

Because of the logarithmic transformation, the forecasts of the variance are guaranteed to be non-negative. Thus, in contrast to the GARCH model, no restrictions need to be imposed on the model estimation. The number of possible conditional volatility formulations is vast. The threshold GARCH, or TARCH(p,q), model is one of the widely used models:

$$\sigma_t^\delta = a_0 + \sum_{i=1}^{q} (a_i |\varepsilon_{t-i}|^\delta) + \gamma |\varepsilon_{t-1}|^\delta d(\varepsilon_{t-1} \leqslant 0) + \sum_{i=1}^{p} (b_i \sigma_{t-i}^\delta) \tag{7}$$

where $d(\varepsilon_t \leqslant 0) = 1$ if $\varepsilon_t \leqslant 0$, and $d(\varepsilon_t \leqslant 0) = 0$ otherwise. Zakoian's [5] model is a special case of the TARCH model with $\delta = 1$, while Glosten *et al.* [4] consider a version of the TARCH model with $\delta = 2$. The TARCH model allows a response of volatility to news with different coefficients for good and bad news.

A wide range of ARCH models proposed in the literature has been reviewed by Bera and Higgins [22], Bollerslev *et al.* [23], Bollerslev *et al.* [12], Degiannakis and Xekalaki [7], Gourieroux [24] and Hamilton [25].

3. MODEL SELECTION METHODS

Most of the methods used in the literature for selecting the appropriate model are based on evaluating the ability of the models to describe the data. Standard model selection criteria such as the Akaike information criterion (AIC) [26] and the Schwarz Bayesian criterion (SBC) [27] have widely been used in the ARCH literature, despite the fact that their statistical properties in the ARCH context are unknown. These are defined in terms of $l_T(\hat{\theta})$, the maximized value of the log-likelihood function of a model, where $\hat{\theta}$ is the maximum likelihood estimator of θ based on a sample of size T and $\breve{\theta}$ denotes the dimension of θ, thus:

$$\text{AIC} = l_T(\hat{\theta}) - \breve{\theta} \tag{8}$$

$$\text{SBC} = l_T(\hat{\theta}) - 2^{-1}\breve{\theta}\ln(T) \tag{9}$$

In addition, the evaluation of loss functions for alternative models is mainly used in model selection. When we focus on estimation of means, the loss function of choice is typically the mean squared error (MSE):

$$\mathrm{MSE} = T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 \tag{10}$$

When the same strategy is applied to variance estimation, the choice of the mean squared error is much less clear. Because of high non-linearity in volatility models, a number of researchers constructed heteroscedasticity-adjusted loss functions. Bollerslev *et al.* [12] present four types of loss functions:

$$L_1 = \sum_{t=1}^{T} (\varepsilon_t^2 - \sigma_t^2)^2 \tag{11}$$

$$L_2 = \sum_{t=1}^{T} \ln\left(\frac{\varepsilon_t^2}{\sigma_t^2}\right)^2 \tag{12}$$

$$L_3 = \sum_{t=1}^{T} \frac{(\varepsilon_t^2 - \sigma_t^2)^2}{\sigma_t^4} \tag{13}$$

$$L_4 = \sum_{t=1}^{T} \left(\frac{\varepsilon_t^2}{\sigma_t^2} + \ln(\sigma_t^2)\right) \tag{14}$$

Pagan and Schwert [28] used the first two of the loss functions to compare alternative estimators with in-sample and out-of-sample data sets. Andersen *et al.* [29], Heynen and Kat [30], Hol and Koopman [31], are some examples from the literature that applied loss functions to compare the forecast performance of various volatility models.

Moreover, loss functions have been constructed, based upon the goals of the particular application. West *et al.* [32] developed such a criterion based on the portfolio decisions of a risk averse investor. Engle *et al.* [33] assumed that the objective was to price options and developed a loss function from the profitability of a particular trading strategy.

4. MODEL SELECTION BASED ON THE STANDARDIZED PREDICTION ERROR CRITERION (SPEC)

Let us assume that a researcher is interested in evaluating the ability of the ARCH models to forecast the conditional variance. Consider the simple case of a regression model: $y_t = x_t'\beta + \varepsilon_t$ where β is a vector of k unknown parameters to be estimated, x_t is a vector of explanatory variables included in the information set at time $t-1$ and $\varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$. At time $t-1$, the expected value μ_t of y_t is estimated on the basis of the information available at time $t-1$, i.e. $\hat{y}_{t|t-1} = \hat{\mu}_t = x_t'\hat{\beta}_{t-1}$, where $\hat{\beta}_{t-1} = (\mathbf{X}_{t-1}'\mathbf{X}_{t-1})^{-1}(\mathbf{X}_{t-1}'\mathbf{Y}_{t-1})$ is the least square estimator of β at time $t-1$, $\mathbf{Y}_t$ is the $(l_t \times 1)$ vector of l_t observations on the dependent variable y_t, and $\mathbf{X}_t$ is the $(l_t \times k)$ matrix whose rows comprise the k-dimensional vectors x_t of the explanatory variables included in the information set, so that $\mathbf{X}_t = \begin{bmatrix} \mathbf{X}_{t-1} \\ x_t' \end{bmatrix}$, $\mathbf{Y}_t = \begin{bmatrix} \mathbf{Y}_{t-1} \\ y_t \end{bmatrix}$. Here $l_0 > k$, $l_{t+1} = l_t + 1$ and $|\mathbf{X}_t'\mathbf{X}_t| \neq 0$, $t = 0, 1, \ldots$. In a manner of speaking, $\hat{y}_{t|t}$ and $\hat{y}_{t|t-1}$ can be considered as in-sample and out-of-sample forecasts, respectively. In other words, $\hat{y}_{t|t}$ is measured on the basis of I_t, the information set available at time t, while $\hat{y}_{t|t-1}$ is measured on the basis of I_{t-1}, the information set available at time $t-1$.

In the sequel, the density function $f(.)$, in Equation (1), is assumed to be that of the normal distribution and $\hat{z}_{t|t-1} \equiv \hat{\varepsilon}_{t|t-1}\hat{\sigma}^{-1}_{t|t-1}$ denotes the standardized one step ahead prediction errors[‡]. The most commonly used way to model the conditional variance is the GARCH(p,q) process in (4). The GARCH(p,q) process may be rewritten as[§]

$$\sigma_t^2 = (u_t', \eta_t', w_t')(v, \zeta, \omega)$$

where $u_t' = (1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-q}^2)$, $\eta_t' = 0$, $w_t' = (\sigma_{t-1}^2, \ldots, \sigma_{t-p}^2)$, $v' = (a_0, a_1, \ldots, a_q)$, $\zeta' = 0$, $\omega' = (b_1, \ldots, b_p)$.

The vector $\theta = (\beta', v', \zeta', \omega')$ denotes the set of parameters to be estimated for both the conditional mean and the conditional variance at time t.

The residual $\hat{\varepsilon}_{t|t-1} \equiv y_t - \hat{y}_{t|t-1}$ reflects the difference between the forecast and the observed value of the stochastic process. Xekalaki *et al.* [8] suggested measuring the predictive behaviour of linear regression models on the basis of the standardized distance between the predicted and the observed value of the dependent random variable. The estimate of the standardized distance was defined by

$$r_t = \frac{y_t - \hat{y}_{t|t-1}}{\sqrt{V(\hat{y}_{t|t-1})}}$$

where $V(\hat{y}_{t|t-1}) = (\mathbf{Y}_{t-1} - \mathbf{X}_{t-1}\hat{\beta}_{t-1})'(\mathbf{Y}_{t-1} - \mathbf{X}_{t-1}\hat{\beta}_{t-1})(1 + x_t(\mathbf{X}_{t-1}'\mathbf{X}_{t-1})^{-1}x_t')(l_{t-1} - k)^{-1}$. A scoring rule to rate the performance of the model at time t for a series of T points in time, $(t = 1, \ldots, T)$, was defined by

$$R_T = T^{-1} \sum_{t=1}^{T} r_t^2$$

the average of the squared standardized residuals. As an ARCH model estimates simultaneously the conditional mean and the conditional variance, its evaluation is two fold. In the sequel, this approach is adopted using the average of the squared standardized one step ahead prediction errors as a scoring rule in order to rate the performance of an ARCH model to forecast both the conditional mean and the conditional variance, in particular,

$$R_T = \frac{\sum_{t=1}^{T} \hat{z}_{t|t-1}^2}{T} \tag{15}$$

$\hat{z}_{t|t-1} \equiv \hat{\varepsilon}_{t|t-1}\hat{\sigma}^{-1}_{t|t-1}$ is the estimated standardized distance between the predicted and the observed value of the dependent random variable, when the conditional standard deviation of the dependent variable given I_{t-1} is defined by an ARCH model, $V(y_t|I_{t-1}) \equiv \sigma_t^2$.

Let (θ_t) denote the vector of unknown parameters to be estimated at time t. Under the assumption of constancy of parameters over time, $(\theta_1) = (\theta_2) = \cdots = (\theta_T) = (\theta)$, the estimated

[‡]Consider the case of the AR(1)GARCH(1,1) model as defined by Equations (2) and (4), for $\kappa = 1$ and $p = q = 1$, respectively. The estimators of the one step ahead prediction error and its variance conditional on the information set available at time $t-1$ are given by $\hat{\varepsilon}_{t|t-1} = y_t - \hat{c}_{0,t-1} - \hat{c}_{1,t-1}y_{t-1}$ and $\hat{\sigma}^2_{t|t-1} = \hat{a}_{0,t-1} + \hat{a}_{1,t-1}\hat{\varepsilon}^2_{t-1|t-1} + \hat{b}_{1,t-1}\hat{\sigma}^2_{t-1|t-1}$, respectively. The estimated parameters are indexed by the subscript t to indicate that they may vary with time.

[§]The conditional variance is written in the form: $(u_t', \eta_t', w_t')(v, \zeta, \omega)$, which includes the most widely used ARCH models such as the TARCH and the EGARCH processes.

standardized one step ahead prediction errors $\hat{z}_{t|t-1}, \hat{z}_{t+1|t}, \ldots, \hat{z}_{T|T-1}$ are asymptotically independently standard normally distributed. Symbolically,

$$\hat{z}_{t|t-1} \equiv (y_t - \hat{y}_{t|t-1})\hat{\sigma}^{-1}_{t|t-1} \sim N(0,1), \quad t = 1, 2, \ldots, T \tag{16}$$

To verify this, observe that at time $t-1$, the expected value of y_t is estimated on the basis of the information available at time $t-1$, i.e. $\hat{y}_{t|t-1} = x_t'\hat{\beta}_{t-1}$ and the expected value of the conditional variance is estimated on the basis of the information available at time $t-1$, i.e. $\hat{\sigma}^2_{t|t-1} = (u_t', \eta_t', w_t')(\hat{v}_{t-1}, \hat{\zeta}_{t-1}, \hat{\omega}_{t-1})$. Note that the elements of the vector (u_t', η_t', w_t') belong to the I_{t-1}, so are considered as known values. The $\hat{z}_{t|t-1}$ can be written as

$$\begin{aligned}
\hat{z}_{t|t-1} &= \frac{(y_t - \hat{y}_{t|t-1})}{\sqrt{\hat{\sigma}^2_{t|t-1}}} \\
&= \frac{(x_t'\beta + \varepsilon_t - x_t'\hat{\beta}_{t-1})}{\sqrt{\hat{\sigma}^2_{t|t-1}}} \\
&= \frac{\varepsilon_t}{\sqrt{\hat{\sigma}^2_{t|t-1}}} + \frac{(x_t'(\beta - \hat{\beta}_{t-1}))}{\sqrt{\hat{\sigma}^2_{t|t-1}}} \\
&= \frac{z_t\sqrt{\sigma_t^2}}{\sqrt{\hat{\sigma}^2_{t|t-1}}} + \frac{(x_t'(\beta - \hat{\beta}_{t-1}))}{\sqrt{\hat{\sigma}^2_{t|t-1}}} \\
&= \frac{z_t((u_t', \eta_t', w_t')(v, \zeta, \omega))^{1/2}}{((u_t', \eta_t', w_t')(\hat{v}_{t-1}, \hat{\zeta}_{t-1}, \hat{\omega}_{t-1}))^{1/2}} + \frac{(x_t'(\beta - \hat{\beta}_{t-1}))}{((u_t', \eta_t', w_t')(\hat{v}_{t-1}, \hat{\zeta}_{t-1}, \hat{\omega}_{t-1}))^{1/2}}
\end{aligned}$$

We assume that a sample of T observations has been used to estimate the vector of unknown parameters. According to Bollerslev [2], the maximum likelihood estimate $\hat{\theta}_t$ is strongly consistent for θ and asymptotically normal with mean θ. In other words, $p\lim(\hat{\theta}_t) = \theta \Leftrightarrow p\lim(\hat{\beta}_t', \hat{v}_t', \hat{\zeta}_t', \hat{\omega}_t') = (\beta', v', \zeta', \omega')$, where $p\lim$ denotes limit in probability as the size of the sample, T, goes to infinity. By Slutsky's theorem (see, e.g. Reference [34, p. 118]), for any continuous function $g(x_T)$ that is not a function of T, $p\lim g(x_T) = g(p\lim x_T)$. Hence

$$\begin{aligned}
&p\lim(\hat{z}_{t|t-1}) \\
&= p\lim\left(\frac{z_t(u_t', \eta_t', w_t')(v, \zeta, \omega))^{1/2}}{((u_t', \eta_t', w_t')(\hat{v}_{t-1}, \hat{\zeta}_{t-1}, \hat{\omega}_{t-1}))^{1/2}}\right) + p\lim\left(\frac{(x_t'(\beta - \hat{\beta}_{t-1}))}{((u_t', \eta_t', w_t')(\hat{v}_{t-1}, \hat{\zeta}_{t-1}, \hat{\omega}_{t-1}))^{1/2}}\right)
\end{aligned}$$

Using Slutsky's theorem, the right-hand side of this relationship can be written as

$$\frac{z_t((u_t', \eta_t', w_t')(v, \zeta, \omega))^{1/2}}{(u_t', \eta_t', w_t')(p\lim(\hat{v}_{t-1}, \hat{\zeta}_{t-1}, \hat{\omega}_{t-1}))^{1/2}} + \frac{(x_t' p\lim(\beta - \hat{\beta}_{t-1}))}{(p\lim((u_t', \eta_t', w_t')(\hat{v}_{t-1}, \hat{\zeta}_{t-1}, \hat{\omega}_{t-1})))^{1/2}}$$

$$= \frac{z_t((u_t', \eta_t', w_t')(v, \zeta, \omega))^{1/2}}{(u_t', \eta_t', w_t')(v, \zeta, \omega)^{1/2}} + \frac{(x_t' p\lim(\beta - \hat{\beta}_{t-1}))}{((u_t', \eta_t', w_t')p\lim(\hat{v}_{t-1}, \hat{\zeta}_{t-1}, \hat{\omega}_{t-1}))^{1/2}}$$

$$= z_t + \frac{(x_t')(0)}{((u_t', \eta_t', w_t')(v, \zeta, \omega))^{1/2}}$$

$$= z_t$$

As convergence in probability implies convergence in distribution, the $\hat{z}_{t|t-1}, \hat{z}_{t+1|t}, \ldots, \hat{z}_{T|T-1}$ are asymptotically standard normally distributed:

$$\hat{z}_{t|t-1} \xrightarrow{\mathrm{p}} z_t \Rightarrow \hat{z}_{t|t-1} \xrightarrow{\mathrm{d}} z_t \sim N(0, 1)$$

This result implies that the $\hat{z}_{t|t-1}, \hat{z}_{t+1|t}, \ldots, \hat{z}_{T|T-1}$ are asymptotically independently standard normally distributed, since, from the definition of convergence in probability

$$P(\|(X_{1T}, X_{2T}, \ldots, X_{nT}) - (W_1, W_2, \ldots, W_n)\| > \varepsilon)$$

$$\leqslant P\left(|X_{1T} - W_1| > \sqrt{\varepsilon^2/n}\right) + P\left(|X_{2T} - W_2| > \sqrt{\varepsilon^2/n}\right) + \cdots + P\left(|X_{nT} - W_n| > \sqrt{\varepsilon^2/n}\right)$$

which asserts that component wise convergence in probability always implies convergence of vectors, i.e.

$$\hat{z}_{t|t-1} \xrightarrow{\mathrm{d}} z_t \overset{\text{i.i.d.}}{\sim} N(0, 1)$$

Hence, (16) has been established.

The result of formula (16) is valid for all the conditional variance functions with consistent estimators of the parameters.

Remark

As concerns the EGARCH and the TARCH models, the maximum likelihood estimator $\hat{\theta}_t = (\hat{\beta}_t', \hat{v}_t', \hat{\zeta}_t', \hat{\omega}_t')$ is consistent and asymptotically normal. In particular, the EGARCH(p,q) model can be written as

$$\ln \sigma_t^2 = (u_t', \eta_t', w_t')(v, \zeta, \omega)$$

where $u_t' = (1, |\varepsilon_{t-1}/\sigma_{t-1}|, \ldots, |\varepsilon_{t-q}/\sigma_{t-q}|)$, $\eta_t' = ([\varepsilon_{t-1}/\sigma_{t-1}], \ldots, [\varepsilon_{t-q}/\sigma_{t-q}])$, $w_t' = (\ln \sigma_{t-1}^2, \ldots, \ln \sigma_{t-p}^2)$, $v' = (a_0, a_1, \ldots, a_q)$, $\zeta' = (\gamma_1, \ldots, \gamma_q)$, $\omega' = (b_1, \ldots, b_p)$.

According to Nelson [3], under sufficient regularity conditions, the maximum likelihood estimator $\hat{\theta}_t = (\hat{\beta}_t', \hat{v}_t', \hat{\zeta}_t', \hat{\omega}_t')$ is consistent and asymptotically normal. Also, for the Glosten *et al.*'s

[4] TARCH(p,q) process, the conditional variance can be written as

$$\sigma_t^2 = (u_t', \eta_t', w_t')(v, \zeta, \omega)$$

where $u_t' = (1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-q}^2)$, $\eta_t' = (d(\varepsilon_{t-1} \leqslant 0)\varepsilon_{t-1}^2)$, $w_t' = (\sigma_{t-1}^2, \ldots, \sigma_{t-p}^2)$, $v' = (a_0, a_1, \ldots, a_q)$, $\zeta' = (\gamma)$, $\omega' = (b_1, \ldots, b_p)$, $d(\varepsilon_t \leqslant 0) = 1$ if $\varepsilon_t \leqslant 0$, and $d(\varepsilon_t \leqslant 0) = 0$ otherwise.

As pointed out by Glosten *et al.* [4], as long as the conditional mean and variance are correctly specified, the maximum likelihood estimates will be consistent and asymptotically normal.

According to Slutsky's theorem, if $p\lim \hat{z}_{t|t-1} = z_t \sim N(0,1)$ and $g(\hat{z}_{t|t-1}) = \sum_{t=1}^{T} (\hat{z}_{t|t-1}^2)$, which is a continuous function, then $p\lim \sum_{t=1}^{T} (\hat{z}_{t|t-1}^2) = \sum_{t=1}^{T} (z_t^2)$. As convergence in probability implies convergence in distribution, $\sum_{t=1}^{T} (\hat{z}_{t|t-1}^2) \xrightarrow{d} \sum_{t=1}^{T} (z_t^2) \sim \chi_T^2$. Hence, as $\hat{z}_{t|t-1}$ are asymptotically standard normal variables, the variable TR_T is asymptotically χ^2 distributed with T degrees of freedom, i.e.

$$TR_T \xrightarrow{d} \chi_T^2 \tag{17}$$

According to Kibble [35], if, for $t = 1, 2, \ldots, T$, $\hat{z}_{t|t-1}^{(A)}$ and $\hat{z}_{t|t-1}^{(B)}$ are standard normally distributed variables, following jointly the bivariate standard normal distribution, then the joint distribution of $(T/2\, R_T^{(A)}, T/2\, R_T^{(B)})$ is the bivariate gamma distribution with probability density function (p.d.f) given by

$$f_{T/2R_T^{(A)},T/2R_T^{(B)}}(x,y)$$
$$= \frac{\exp(-(x+y)/1-\rho^2)}{\Gamma(T/2)(1-\rho^2)^{T/2}} \sum_{i=0}^{\infty} \left(\frac{(\rho/(1-\rho^2))^{2i}}{\Gamma(i+1)\Gamma(i+(T/2))} (xy)^{(T/2)-1-i} \right), \quad x, y > 0 \tag{18}$$

where $\Gamma(.)$ is the gamma function and ρ is the correlation coefficient between $\hat{z}_{t|t-1}^{(A)}$ and $\hat{z}_{t|t-1}^{(B)}$, i.e. $\rho \equiv \mathrm{Cor}(\hat{z}_{t|t-1}^{(A)}, \hat{z}_{t|t-1}^{(B)})$. Xekalaki *et al.* [8] showed that, when the joint distribution of $(T/2R_T^{(A)}, T/2R_T^{(B)})$ is Kibble's bivariate gamma, the distribution of the ratio $Z_T^{(A,B)} \equiv R_T^{(A)}/R_T^{(B)}$ is defined by the following p.d.f.:

$$f_{Z_T^{(A,B)}}(z) = \frac{(1-\rho^2)^{T/2}}{B(T/2,T/2)} z^{T/2-1}(1+z)^{-T} \left[1 - \left(\frac{2\rho}{z+1} \right)^2 z \right]^{-(T+1)/2}, \quad z > 0 \tag{19}$$

where $B(T/2, T/2) = \Gamma(T/2)^2/\Gamma(T)$. Symbolically,

$$Z_T^{(A,B)} \equiv \sum_{t=1}^{T} \hat{z}_{t|t-1}^{2(B)} / \sum_{t=1}^{T} \hat{z}_{t|t-1}^{2(A)} \sim \mathrm{CGR}(k, \rho) \tag{20}$$

where $k = T/2$. Xekalaki *et al.* [8] referred to the distribution in (19) as the correlated gamma ratio (CGR) distribution. (A sample of tables of the percentage points of this distribution and of graphs depicting its probability density function is given in Appendix A, Table AIII and Figure A5, respectively.) Full tables of percentage points and graphs for various values of k and ρ can be found in [8].

As pointed out by Xekalaki *et al.* [8], $R_T^{(A)}$ and $R_T^{(B)}$ could represent the sum of the squared standardized prediction errors from two regression models (not necessarily nested) but with a common dependent variable. Thus, two regression models can be compared through testing a null hypothesis of equivalence of the models in their predictability against the alternative that model A produces 'better' predictions. Here, the notion of the equivalence of two models with respect to their predictive ability is considered in Reference [8] sense to be defined implicitly through their mean squared prediction errors. Following Xekalaki *et al.*'s [8] rationale, the closest description of the hypothesis to be tested is

H_0: Models A and B have equal mean squared prediction errors

Versus

H_1: Model A has lower mean squared prediction error than model B

using $Z_T^{(A,B)}$ as a test statistic, i.e. using the ratio of the sum of the squared standardized one step ahead prediction errors $\hat{z}_{t|t-1}$ of the two competing models. The null hypothesis is rejected if $Z_T^{(A,B)} > \mathrm{CGR}(k, \rho, a)$, where $\mathrm{CGR}(k, \rho, a)$ is the $100(1-a)$ percentile of the CGR distribution.

Since very few financial time series have a constant conditional mean of zero, in order to estimate the conditional variance, the conditional mean should have been defined. Thus, both the conditional mean and variance are estimated simultaneously. According to the SPEC model selection algorithm, the models that are considered as having a 'better' ability to predict future values of the dependent variable, are those with the lowest sum of squared standardized one step ahead prediction errors. It becomes evident, therefore, that these models can potentially be regarded as the most appropriate to use for volatility forecasts too.

5. EMPIRICAL RESULTS

The suggested model selection procedure is illustrated on data referring to the daily returns of the Athens stock exchange (ASE) index. Let $y_t = \ln(P_t/P_{t-1})$ denote the continuously compound rate of return from time $t-1$ to t, where P_t is the ASE closing price at time t. The data set covers the period from August 30th, 1993 to November 4th, 1996, a total of 800 trading days. Table I presents the descriptive statistics. For an estimated kurtosis equal to 7.25 and an estimated skewness equal to 0.08, the distribution of returns is flat (platykurtic) and has a long right tail relative to the normal distribution. The Jarque Bera (JB) statistic [36] is used to test whether the series is normally distributed. The test statistic measures the difference of the skewness and kurtosis of the series from those of the normal distribution. The JB statistic is computed as

$$\mathrm{JB} = T(S^2 + ((K-3)^2/4))/6 \tag{21}$$

where T is the number of observations, S is the skewness and K is the kurtosis. Under the null hypothesis of a normal distribution, the JB statistic is χ^2 distributed with two degrees of freedom.

From Table I, the value of the JB statistic obtained is 602.38 with a very low p-value (practically zero). So, the null hypothesis of normality is rejected. In order to determine whether $\{y_t\}$ is a stationary process, the Augmented Dickey Fuller test (ADF) [37] and the non-parametric Phillips Perron (PP) test [38, 39] are conducted.

The ADF test examines the null hypothesis, $H_0{:}\gamma = 0$, versus the alternative, $H_1{:}\gamma < 0$, in the following regression:

$$\Delta y_t = c + \gamma y_{t-1} + \sum_{i=1}^{K} \varphi_i \Delta y_{t-i} + \varepsilon_t \tag{22}$$

Table I. Descriptive statistics of the daily returns of the ASE index (30th August 1993–4th November 1996 (800 observations)).

Observations	800
Mean	5.72E-05
Median	−0.00018
Standard deviation	0.012
Skewness	0.08
Kurtosis	7.25
Jarque Bera (JB)	602.38
probability	<0.000001
Augmented Dickey Fuller (ADF)	−12.67
1% critical value	−3.44
Phillips Perron (PP)	−24.57
1% critical value	−3.44

The skewness of a symmetric distribution, as the normal distribution, is zero. Positive skewness implies that the distribution has a long right tail. Negative skewness implies a long left tail distribution. The kurtosis of the normal distribution is 3. If the kurtosis exceeds 3, the distribution is peaked (leptokurtic) relative to the normal. If the kurtosis is less than 3, the distribution is flat (platykurtic) relative to the normal. Under the null hypothesis of a normal distribution, the JB statistic is χ^2 distributed with two degrees of freedom. The reported probability is the probability that the JB statistic exceeds, in absolute value, the observed value under the null hypothesis. ADF: The null hypothesis of non-stationarity is rejected if the ADF value is less than the critical value (four lagged differences). PP: The null hypothesis of non-stationarity is rejected if the PP value is less than the critical value (four truncation lags).

where Δ denotes the difference operator. According to the ADF test, the null hypothesis of non-stationarity is rejected at the 1% level of significance for any lag order up to $\kappa = 12$. The test regression for the PP test is the AR(1) process:

$$\Delta y_t = c + \gamma y_{t-1} + \varepsilon_t \tag{23}$$

While the ADF test corrects for higher order serial correlation by adding lagged differenced terms on the right-hand side, the PP test makes a correction to the t statistic of the γ coefficient from the AR(1) regression to account for the serial correlation in ε_t. The correction is non-parametric since an estimate of the spectrum of ε_t at frequency zero, that is robust to heteroscedasticity and autocorrelation of unknown form, is used. According to the PP test, the null hypothesis is also rejected at the 1% level of significance.

The most commonly used test for examining the null hypothesis of homoscedasticity against the alternative hypothesis of heteroscedasticity is Engle's [1] Lagrange multiplier (LM) test. The ARCH LM test statistic is computed from an auxiliary test regression. To test the null hypothesis of no ARCH effects up to order q in the residuals, the regression model

$$\varepsilon_t^2 = \beta_0 + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}^2 + u_t \tag{24}$$

with $\varepsilon_t = y_t - c$ is run. Engle's test statistic is computed as the product of the number of observations times the value of the coefficient of variation R^2 of the auxiliary test regression. From Table II, the values of the LM test statistic for $q = 1, \ldots, 8$ are highly significant at any reasonable level.

As, according to the results of the above tests, the assumptions of stationarity and ARCH effects seem to be plausible for the process $\{y_t\}$ of daily returns, several ARCH models are

Table II. Lagrange multiplier (LM) test.

$$\varepsilon_t^2 = \beta_0 + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}^2 + u_t$$

$$\varepsilon_t = y_t - c$$

Q	LM statistic	p-value
1	108.203	0.00
2	113.315	0.00
3	127.947	0.00
4	128.577	0.00
5	130.691	0.00
6	133.467	0.00
7	131.573	0.00
8	129.496	0.00

The LM statistic is computed as the number of observations times the R^2 from the auxiliary test regression. It converges in distribution to a χ_q^2. Test the null hypothesis of no ARCH effects in the residuals up to order q.

considered in the sequel. It is assumed, specifically, that the conditional mean is considered as a κth-order autoregressive process as defined in (2) and the conditional variance σ_t^2 is assumed to be related to lagged values of ε_t and σ_t according to a GARCH(p,q) model, an EGARCH(p,q) model or a TARCH(p,q) model as defined by (4), (6) and (7), respectively. Thus, the AR(κ)GARCH(p,q), AR(κ)EGARCH(p,q) and AR(κ)TARCH(p,q) models[¶] are applied, for $\kappa = 0, \ldots, 4$, $p = 0, 1, 2$ and $q = 1, 2$, yielding a total of 90 cases.

Since, in estimating non-linear models, no closed-form expressions are obtainable for the parameter estimators, an iterative method has to be employed. The value of the parameter vector θ that maximizes $l_t(\theta)$, the log likelihood contribution for each observation t, is to be found. Iterative optimization algorithms work by starting with an initial set of values for the parameter vector θ, say $\theta^{(0)}$, and obtaining a set of parameter values $\theta^{(1)}$, which corresponds to a higher value of $l_t(\theta)$. This process is repeated until the objective function $l_t(\theta)$ no longer improves between iterations. In the sequel, the Marquardt algorithm [40] is used. This algorithm modifies the Berndt, Hall, Hall and Hausman, or BHHH, algorithm [41] by adding a correction matrix to the Hessian approximation (i.e. to the sum of the outer product of the gradient vectors for each observation's contribution to the objective function). The Marquardt updating algorithm is computed as

$$\theta^{(i+1)} = \theta^{(i)} + \left(\sum_{t=1}^{T} \frac{\partial l_t^{(i)}}{\partial \theta} \frac{\partial l_t^{(i)}}{\partial \theta'} - aI \right)^{-1} \sum_{t=1}^{T} \frac{\partial l_t^{(i)}}{\partial \theta} \tag{25}$$

where I is the identity matrix and a is a positive number chosen by the algorithm. The effect of this modification is to push the parameter estimates in the direction of the gradient vector. The idea is that when we are far from the maximum, the local quadratic approximation to the function may be a poor guide to its overall shape, so it may be better off to simply follow

¶ Glosten *et al.* 's [4] TARCH model is applied with $\delta = 2$.

the gradient. The correction may provide a better performance at locations far from the optimum, and allows for computation of the direction vector in cases where the Hessian is near singular.

The quasi-maximum likelihood estimator (QMLE) is used, as according to Bollerslev and Wooldridge [42], it is generally consistent, has a limiting normal distribution and provides asymptotic standard errors that are valid under non-normality.

In order to compute the sum of squared standardized one step ahead prediction errors, a rolling sample of constant size equal to 500 is used, or $T = 500$, so 300 one step ahead daily forecasts are estimated. The out-of-sample data set is split into five subperiods and the SPEC model selection algorithm is applied in each subperiod separately. Thus, the model selection is revised every 60 trading days and the information set includes daily continuously compound returns of the two most recently years, or 500 trading days. The choice of a 60-day length for each subperiod is arbitrary. The sum of the squared one step ahead prediction errors, $\sum_{t=T+1}^{T+s}(\hat{z}^2_{t|t-1})$, is estimated for each model and presented in Table AI, in Appendix A. The models selected for each subperiod and their sums of the squared standardized one step ahead prediction errors are:

Subperiod	Model Selected	$\min\left(\sum_{t=T+1}^{T+s}(\hat{z}^2_{t\|t-1})\right)$
1. 25 August 1995–16 November 1995	AR(2) EGARCH(0,1)	21.961
2. 17 November 1995–13 February 1996	AR(0) EGARCH(0,1)	76.315
3. 14 February 1996–14 May 1996	AR(0) EGARCH(0,1)	42.176
4. 15 May 1996–8 August 1996	AR(3) EGARCH(0,1)	27.308
5. 9 August 1996–4 November 1996	AR(1) EGARCH(0,1)	43.920

According to the SPEC selection method, the exponential GARCH(0,1) model describes best the conditional variance for the total examined period of 300 trading days. It is selected by the SPEC selection method in each subperiod. Figure 1 shows the daily value of the ASE index and the one step ahead conditional standard deviation of its returns.

Despite the fact that an asymmetric model is selected by the SPEC algorithm, there are no asymmetries in the ASE index volatility. According to Figure 1, the major episodes of high volatility are not associated with market changes of the same sign. Figure 2 presents the values of the parameters a_1 and γ_1 of the 300 estimated EGARCH(0,1) models, while Figure 3 depicts the relevant standard errors for the parameters a_1 and γ_1. Obviously, the γ_1 parameter, which allows for the asymmetric effect, is positive but statistically insignificant. Therefore, the asymmetric relation between returns and changes in volatility does not characterize the examined period.

An interesting point is that the higher order of the conditional mean autoregressive process is chosen as adequate to produce more accurate predictions for the first and the fourth subperiods. As concerns the first subperiod, the AR(2)EGARCH(0,1) model

$$\begin{aligned} y_t &= c_0 + c_1 y_{t-1} + c_2 y_{t-2} + \varepsilon_t \\ \ln(\sigma_t^2) &= a_0 + a_1\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma_1\left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right) \end{aligned} \tag{26}$$

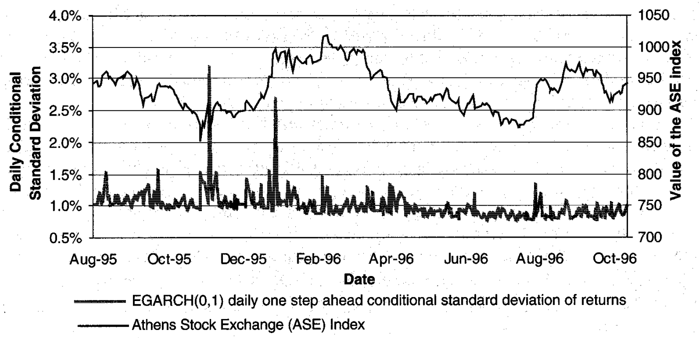


Figure 1. The ASE index and the one step ahead conditional standard deviation of its returns estimated by the EGARCH(0,1) models.

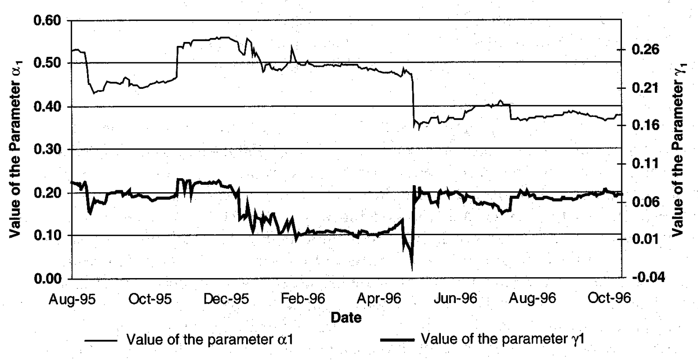


Figure 2. The parameters of the estimated EGARCH(0,1) models.

is the one with the lowest value of $\sum_{t=501}^{560}(\hat{z}_{t|t-1}^{2})$ equal to 21.961. The hypothesis:

H_0: The model AR(2)EGARCH(0,1) has equivalent predictive ability to model X is tested versus.

H_1: The model AR(2)EGARCH(0,1) produces 'better' predictions than model X, with X denoting any one of the remainder models.

Note that the correlation between the standardized one step ahead prediction errors is greater than 0.9 in each case. If $Z_{60}^{\mathrm{AR(2)EGARCH(0,1)},X} \equiv (21.96)^{-1}\sum_{t=501}^{560}\hat{z}_{t|t-1}^{(X)2} > \mathrm{CGR}(k=30, \rho>0.9, a)$, the null hypothesis of equivalent predictive ability of the models is rejected at $100a\%$ level of

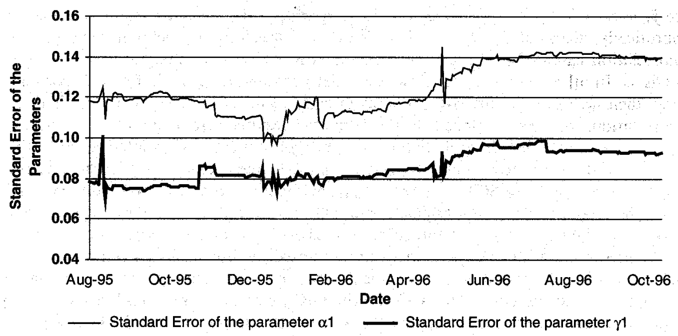


Figure 3. The standard error for the parameters of the estimated EGARCH(0,1) models.

significance and the AR(2)EGARCH(0,1) model is regarded as 'better' than model X. Table AII, in Appendix A, summarizes the results of the hypothesis tests, for each subperiod.

Figure A1, in Appendix A, depicts the one step ahead 95% prediction intervals for the models with the lowest $\sum_{t=T+1}^{T+s} (\hat{z}^2_{t|t-1})$ in each subperiod. The prediction intervals are constructed as the expected rate of return plus\minus 1.96 times the conditional standard deviation, both measurable to $t-1$ information set: $\hat{\mu}_{t|t-1} \pm 1.96\hat{\sigma}_{t|t-1}$. So, each time next day's prediction interval is plotted, only information available at current day is used. Remark that around November 1995, a volatile period, the prediction interval in Figure A1 tracked the movement of the returns quite closely (seven outliers, or 2.33%, were observed).

6. AN ALTERNATIVE APPROACH

In this section an in-sample analysis is performed in order to select the appropriate models describing the data. Then, the selected models are used to estimate the one step ahead forecasts. Having assumed that the conditional mean of the returns follows a κth order autoregressive process, as in (2), Richardson and Smith [43] developed a test for autocorrelation. It is a robust version of the standard Box Pierce [44] procedure. For p_i denoting the estimated autocorrelation between the returns at time t and $t-i$, the test is formulated as

$$\mathrm{RS}(r) = T \sum_{i=1}^{r} \frac{p_i^2}{1+c_i} \tag{27}$$

where T is the sample size and c_i is the adjustment factor for heteroscedasticity, which is calculated as

$$c_i = \frac{\mathrm{Cov}(\bar{y}_t^2, \bar{y}_{t-i}^2)}{\mathrm{Var}(y_t)^2} \tag{28}$$

where $\tilde{y}_t = y_t - T^{-1}\sum_{t=1}^{T} y_t$. Under the null hypothesis of no autocorrelation, the statistic is asymptotically distributed as χ^2 with r degrees of freedom. If the null hypothesis of no autocorrelation cannot be rejected, then the returns' process is equal to a constant plus the residuals, ε_t. In other words, $\{y_t\}$ follows the AR(0) process. If the null of no autocorrelation is rejected, then $\{y_t\}$ follows the AR(1) process. In order to test for the existence of a higher order autocorrelation, the test is applied on the estimated residuals from the AR(1) model. In this case, the statistic, under the null hypothesis, is asymptotically distributed as χ^2 with $r-1$ degrees of freedom. The test is calculated on seven autocorrelations ($r = 7$) for 800 observations yielding a value equal to $\mathrm{RS}(7) = 14,86 > \chi^2_{7,0.05}$. As the null hypothesis of no autocorrelation is rejected the test is run on the estimated residuals from the AR(1) model that gives $\mathrm{RS}(6) = 12,33 < \chi^2_{6,0.05}$. Thus, a first-order autocorrelation is detected for the returns' process. Note that the AR(1) form allows for the autocorrelation imposed by discontinuous trading.

Having defined the conditional mean equation, the next step is the estimation of the conditional variance function. The AIC and the SBC criteria are used to select the appropriate conditional variance equation. Note that the AIC mainly chooses as best the less parsimonious model. Also, under certain regularity conditions, the SBC is consistent, in the sense that for large samples it leads to the correct model choice, assuming the 'true' model does belong to the set of models examined. Thus, the SBC may be preferable to use. As concerns the specific data set, both the AIC and SBC select the GARCH(1,1) model as the most appropriate function to describe the conditional variance. So, performing an in-sample analysis the AR(1)GARCH(1,1) model is regarded as the most suitable, which is the model applied in most researches. Figure A2, in Appendix A, presents the in-sample 95% confidence interval for the AR(1)GARCH(1,1) model. There are 14 observations, or 4.66%, outside the confidence interval.

In order to compare the model selection methods, the choice of the models should be conducted at the same time points. Thus, the Richardson Smith test for autocorrelation detection and the information criteria for model selection are used in each subperiod separately. The models selected for in each subperiod are:

Subperiod	Richardson Smith model selection	SBC model selection	AIC model selection
1.	AR(3)	GARCH(1,1)	EGARCH(1,2)
2.	AR(2)	GARCH(2,1)	GARCH(2,1)
3.	AR(0)	GARCH(1,1)	GARCH(1,1)
4.	AR(0)	GARCH(1,1)	GARCH(1,1)
5.	AR(0)	GARCH(1,1)	TARCH(1,1)

Based on Table AII, the hypothesis that the model selected by the in-sample analysis is equivalent to the model with minimum value of $\sum_{t=T+1}^{T+s}(\hat{z}^2_{t|t-1})$ is rejected in the majority of the cases.

Proceeding as in the previous section, the one step ahead prediction intervals, for the models selected in each subperiod, are created. As in Section 5, next day's prediction is based only on information available at current day. Figures A3 and A4 in Appendix A, present the one step ahead 95% prediction intervals for the models selected by the SBC and AIC, respectively. There are 13 observations, or 4.33%, outside the prediction interval for the models selected by the SBC, whereas there are 14 outliers, or 4.66%, for the models selected by the AIC. Therefore, the importance of selecting a conditional variance model based on its ability to forecast and not on fitting the data gains a lead over. Of course, the construction of the prediction intervals is a naïve way to examine the accuracy of our method's predictability.

7. DISCUSSION

An alternative model selection approach, based on the CGR distribution, was introduced. Instead of being based on evaluating the ability of the models to describe the data (Akaike information and Schwarz Bayesian criteria), the proposed approach is based on evaluating the ability of the models to predict the conditional variance. The method was applied to 800 daily returns of the ASE index, a data set covers the period from August 30th, 1993 to November 4th, 1996. The first T observations were used to estimate the one step ahead prediction of the conditional mean and variance at $T+1$. For $T=500$, a total of 300 one step ahead predictions of the conditional mean and variance were obtained. The out-of-sample data set was split into subsets, one for each of five subperiods and the SPEC model selection algorithm was applied in each subperiod separately. Thus, the model selection was revised every 60 trading days.

The idea of 'jumping' from one model to another, as stock market behaviour alters, is introduced. The transition from one model to another is done according to the SPEC model selection algorithm. Each time the model selection method is applied, the model is used to predict the conditional variance is revised. Of course, the idea of switching from one regime to another has been already applied to the class of switch regime ARCH models introduced by Cai [15] and Hamilton and Susmel [16] and extended by several authors such as Dueker [45] and Hansen [46]. However, these models allow the parameters of a specific ARCH model to come from one of several different regimes, with transitions between regimes governed by an unobserved Markov chain.

Using an alternative approach, based on evaluating the ability of fitting the data, the conditional mean is first modelled and subsequently, an appropriate form for the conditional variance is chosen. Applying the SPEC model selection algorithm, the null hypothesis, that the model selected by the in-sample analysis is equivalent to the model with minimum value of $\sum_{t=T+1}^{T+s}(\hat{z}^2_{t|t-1})$, is rejected in the plurality of the cases at less than 5% level of significance. The in-sample model selection methods and the predictability-based method do not coincide in the sifting of the appropriate conditional variance model. Moreover, 2.33 and 4.33% of the data were outside the $\hat{\mu}_{t|t-1} \pm 1.96\hat{\sigma}_{t|t-1}$ prediction interval constructed based on the SPEC and the SBC model selection methods, respectively.

The predictive ability of the SPEC model selection algorithm has to be further investigated. Among the financial applications where this method could have a potential use are in the fields of portfolio analysis, risk management and trading option derivatives.

APPENDIX A

The sum of the squared one step ahead prediction errors, $\sum_{t=T+1}^{T+s}(\hat{z}^2_{t|t-1})$, is estimated for each model and presented in Table AI.

Table AII summarizes the results of the hypothesis tests, for each subperiod.

The percentage points of the CGR distribution is presented in Table AIII.

Table AI. Sum of squared standardized one step ahead prediction errors for each subperiod.

$$\text{AR}(k) \quad y_t = c_0 + \sum_{i=1}^{k}(c_i y_{t-i}) + \varepsilon_t$$

$$\text{GARCH}(p,q) \quad \sigma_t^2 = a_0 + \sum_{i=1}^{q}(a_i\varepsilon_{t-i}^2) + \sum_{i=1}^{p}(b_i\sigma_{t-i}^2)$$

$$\text{EGARCH}(p,q) \quad \ln(\sigma_t^2) = a_0 + \sum_{i=1}^{q}\left(a_i\left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right| + \gamma_i\left(\left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right|\right)\right) + \sum_{i=1}^{p}(b_i\ln(\sigma_{t-i}^2))$$

$$\text{TARCH}(p,q) \quad \sigma_t^2 = a_0 + \sum_{i=1}^{q}(a_i\varepsilon_{t-i}^2) + \gamma\varepsilon_{t-i}^2 d(\varepsilon_{t-i}\leqslant 0) + \sum_{i=1}^{p}(b_i\sigma_{t-i}^2)$$

	(a) 25 August 1995–16 November 1995 ($s=[501,560]$)					(b) 17 November 1995–13 February 1996 ($s=[561,620]$)					(c) 14 February 1996–14 May 1996 ($s=[621,680]$)				
	$k=0$*	$k=1$	$k=2$	$k=3$	$k=4$	$k=0$*	$k=1$	$k=2$	$k=3$	$k=4$	$k=0$*	$k=1$	$k=2$	$k=3$	$k=4$
GARCH(p,q)															
$p=0,\ q=1$	26.371	25.465	24.843	25.173	26.570	81.183	79.657	79.913	83.204	89.584	45.970	46.740	46.793	47.855	47.882
$p=0,\ q=2$	30.150	29.493	28.940	29.109	30.835	88.007	85.947	88.135	89.575	95.825	46.138	46.323	46.039	47.496	47.382
$p=1,\ q=1$	39.076	38.848	38.289	38.496	38.466	79.571	84.410	85.070	85.671	86.749	50.273	50.205	49.959	50.363	49.320
$p=1,\ q=2$	39.129	38.709	38.159	38.533	38.456	80.684	85.214	85.554	87.046	89.907	50.429	50.097	49.814	50.223	49.330
$p=2,\ q=1$	39.183	38.304	37.882	37.829	37.889	79.703	83.700	86.917	84.920	87.420	50.650	50.334	49.547	49.917	49.843
$p=2,\ q=2$	39.511	38.742	38.336	39.223	38.377	81.230	84.534	85.143	82.863	88.940	50.811	50.126	50.051	50.330	48.975
TARCH(p,q)															
$p=0,\ q=1$	26.795	25.892	25.270	25.683	27.300	81.505	80.810	81.158	84.704	90.674	45.947	46.731	46.749	47.769	47.806
$p=0,\ q=2$	31.151	30.981	30.442	30.619	32.125	88.977	88.465	91.004	92.734	98.915	46.114	46.311	46.001	47.422	47.263
$p=1,\ q=1$	39.070	38.624	38.146	38.506	38.550	81.296	85.321	86.339	87.601	88.412	50.461	50.262	50.006	50.396	49.368
$p=1,\ q=2$	39.016	38.667	38.185	38.660	38.482	86.517	87.338	88.246	92.729	98.976	50.677	50.145	49.830	50.229	49.512
$p=2,\ q=1$	39.279	37.836	37.422	38.005	38.290	81.609	86.085	85.458	84.975	90.097	50.769	49.491	48.737	50.231	49.613
$p=2,\ q=2$	40.975	38.732	38.180	38.755	38.398	89.614	86.608	87.364	91.126	98.289	51.664	49.794	50.262	50.548	50.133
EGARCH(p,q)															
$p=0,\ q=1$	23.770	22.644	21.961	22.047	22.722	76.315	78.689	78.342	78.551	84.422	42.176	42.724	42.688	43.561	43.383
$p=0,\ q=2$	27.289	27.340	26.731	26.896	28.312	87.867	91.361	92.862	93.526	101.216	43.712	44.279	44.178	45.395	44.838
$p=1,\ q=1$	44.281	43.555	43.131	43.321	41.934	88.246	96.778	98.579	99.805	99.650	49.382	48.836	48.837	49.369	48.644
$p=1,\ q=2$	43.754	42.427	41.360	42.235	41.231	98.796	103.714	105.834	107.774	108.783	49.140	48.716	48.592	49.065	48.608
$p=2,\ q=1$	44.620	43.216	43.138	43.142	42.077	90.043	98.056	99.570	101.509	101.531	49.422	48.384	48.301	48.452	48.380
$p=2,\ q=2$	43.926	42.915	42.231	42.645	41.138	93.750	102.953	112.441	105.882	–†	51.970	49.555	–†	48.992	–†

Table AI. *Continued.*

	(d) 15 May 1996–8 August 1996 ($s=[681,740]$)					(e) 9 August 1996–4 November 1996 ($s=[741,800]$)				
	$k=0^*$	$k=1$	$k=2$	$k=3$	$k=4$	$k=0^*$	$k=1$	$k=2$	$k=3$	$k=4$
GARCH(p,q)										
$p=0,\ q=1$	30.568	30.619	29.473	29.346	29.534	48.288	47.469	47.437	49.749	50.771
$p=0,\ q=2$	31.557	32.105	30.967	30.861	30.813	50.795	49.575	49.484	51.426	52.236
$p=1,\ q=1$	36.016	36.440	35.335	35.175	35.013	55.915	54.344	54.572	54.967	55.281
$p=1,\ q=2$	36.098	36.951	35.846	35.706	35.431	56.099	54.631	54.872	55.163	55.399
$p=2,\ q=1$	35.732	37.374	36.069	36.020	35.628	55.807	55.420	55.335	56.306	56.075
$p=2,\ q=2$	35.859	36.647	36.252	35.446	35.437	56.102	54.814	55.145	55.137	55.359
TARCH(p,q)										
$p=0,\ q=1$	30.747	30.605	29.419	29.352	29.593	47.179	47.143	47.101	49.494	50.529
$p=0,\ q=2$	31.821	31.978	30.804	30.785	30.811	49.483	49.131	49.030	51.031	51.935
$p=1,\ q=1$	36.029	36.326	35.157	35.147	35.075	53.866	53.341	53.616	53.897	54.272
$p=1,\ q=2$	36.117	36.636	35.489	35.482	35.298	54.065	53.684	53.835	54.075	54.327
$p=2,\ q=1$	36.279	37.214	35.789	36.224	35.946	53.925	54.199	53.999	54.245	56.211
$p=2,\ q=2$	35.945	37.646	35.776	36.005	36.030	54.181	54.482	54.725	55.039	54.846
EGARCH(p,q)										
$p=0,\ q=1$	29.252	28.733	27.428	27.308	27.330	44.260	43.920	44.047	45.908	46.528
$p=0,\ q=2$	30.310	30.109	28.772	28.644	28.563	46.453	45.986	46.035	47.513	47.990
$p=1,\ q=1$	35.972	36.142	34.806	34.716	34.754	52.752	53.271	53.285	53.801	53.944
$p=1,\ q=2$	36.251	36.923	35.548	35.477	35.460	53.233	54.767	54.191	54.450	54.617
$p=2,\ q=1$	35.706	37.371	36.176	36.190	36.266	53.922	55.703	55.410	55.596	55.726
$p=2,\ q=2$	35.562	35.109	34.329	34.210	34.777	52.438	54.052	53.963	–†	54.716

The AR(k)GARCH(p,q), AR(k)EGARCH(p,q) and AR(k)TARCH(p,q) models are applied, for $k=0,\ldots,4$, $p=0,1,2$ and $q=1,2$.
*Regress the depedent variable on a constant.
†Model fails to converge at least once.

Table AII. Testing the null hypothesis that the model with the lowest sum of the squared standardized one step ahead prediction errors has equivalent predictive ability to model X, with X denoting any of the remaining models.

(a) 25 August 1995–16 November 1995 (1st subperiod)

H_0: The model AR(2)-EGARCH(0,1) is equivalent to model X versus H_1: The model AR(2)-EGARCH(0,1) is "better" than model X.

Moldel for conditional variance	Model for conditional mean				
	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
GARCH(0,1)	1.201	1.160	1.131	1.146	1.210
p-value	<0.10	<0.10	<0.25	<0.25	<0.05
GARCH(0,2)	1.373	1.343	1.318	1.326	1.404
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
GARCH(1,1)	1.779	1.769	1.744	1.753	1.752
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
GARCH(1,2)	1.782	1.763	1.738	1.755	1.751
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
GARCH(2,1)	1.784	1.744	1.725	1.723	1.725
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
GARCH(2,2)	1.799	1.764	1.746	1.786	1.748
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
TARCH(0,1)	1.220	1.179	1.151	1.170	1.243
p-value	<0.05	<0.10	<0.25	<0.10	<0.05
TARCH(0,2)	1.418	1.411	1.386	1.394	1.463
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
TARCH(1,1)	1.779	1.759	1.737	1.753	1.755
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
TARCH(1,2)	1.777	1.761	1.739	1.760	1.752
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
TARCH(2,1)	1.789	1.723	1.704	1.7231	1.744
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
TARCH(2,2)	1.866	1.764	1.739	1.765	1.748
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
E-GARCH(0,1)	1.082	1.031	*	1.004	1.035
p-value	<0.25	>0.25		>0.25	>0.25
E-GARCH(0,2)	1.243	1.245	1.217	1.225	1.289
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
E-GARCH(1,1)	2.016	1.983	1.964	1.973	1.909
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
E-GARCH(1,2)	1.992	1.932	1.883	1.923	1.878
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
E-GARCH(2,1)	2.032	1.968	1.964	1.965	1.916
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
E-GARCH(2,2)	2.000	1.954	1.923	1.942	1.873
p-value	<0.01	<0.01	<0.01	<0.01	<0.01

(b) 17 November 1995–13 February 1996 (2nd subperiod)

H_0: The model AR(0)-EGARCH(0,1) is equivalent to model X versus H_1: The model AR(0)-EGARCH(0,1) is "better" than model X.

Moldel for conditional variance	Model for conditional mean				
	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
GARCH(0,1)	1.064	1.044	1.047	1.090	1.174
p-value	>0.25	>0.25	>0.25	<0.25	<0.1
GARCH(0,2)	1.153	1.126	1.155	1.174	1.256
p-value	<0.25	<0.25	<0.25	<0.1	<0.05
GARCH(1,1)	1.043	1.106	1.115	1.123	1.137
p-value	>0.25	<0.25	<0.25	<0.25	<0.25
GARCH(1,2)	1.057	1.117	1.121	1.141	1.178
p-value	>0.25	<0.25	<0.25	<0.25	<0.1
GARCH(2,1)	1.044	1.097	1.139	1.113	1.146
p-value	>0.25	<0.25	<0.25	<0.25	<0.25
GARCH(2,2)	1.064	1.108	1.166	1.086	1.165
p-value	>0.25	<0.25	<0.25	<0.25	<0.1
TARCH(0,1)	1.068	1.059	1.063	1.110	1.188
p-value	>0.25	>0.25	>0.25	<0.25	<0.1
TARCH(0,2)	1.166	1.159	1.192	1.215	1.296
p-value	<0.1	<0.1	<0.1	<0.05	<0.05
TARCH(1,1)	1.065	1.118	1.31	1.148	1.159
p-value	>0.25	<0.25	<0.25	<0.25	<0.1
TARCH(1,2)	1.134	1.144	1.120	1.215	1.297
p-value	<0.25	<0.25	<0.25	<0.05	<0.05
TARCH(2,1)	1.069	1.128	1.120	1.113	1.181
p-value	>0.25	<0.25	<0.25	<0.25	<0.1
TARCH(2,2)	1.74	1.135	1.145	1.194	1.288
p-value	<0.1	<0.25	<0.25	<0.1	<0.05
E-GARCH(0,1)	*	1.031	1.027	1.029	1.106
p-value		>0.25	>0.25	>0.25	<0.25
E-GARCH(0,2)	1.151	1.197	1.217	1.226	1.326
p-value	<0.25	<0.1	<0.05	<0.05	<0.01
E-GARCH(1,1)	1.156	1.268	1.292	1.308	1.306
p-value	<0.25	<0.05	<0.05	<0.05	<0.05
E-GARCH(1,2)	1.295	1.359	1.387	1.412	1.425
p-value	<0.05	<0.01	<0.01	<0.01	<0.01
E-GARCH(2,1)	1.180	1.285	1.305	1.330	1.330
p-value	<0.1	<0.05	<0.05	<0.01	<0.01
E-GARCH(2,2)	1.228	1.349	1.473	1.387	*
p-value	<0.05	<0.01	<0.01	<0.01	

Table AII. *Continued.*

(c) 14 February 1996–14 May 1996 (3rd subperiod)
H_0: The model AR(2)-EGARCH(0,1) is equivalent to model X versus H_1: The model AR(2)-EGARCH(0,1) is "better" than model X

	Model for conditional mean				
	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
GARCH(0,1)	1.090	1.108	1.109	1.135	1.135
p-value	<0.25	<0.25	<0.25	<0.25	<0.25
GARCH(0,2)	1.094	1.098	1.092	1.126	1.123
p-value	<0.25	<0.25	<0.25	<0.25	<0.25
GARCH(1,1)	1.192	1.190	1.185	1.194	1.169
p-value	<0.1	<0.1	<0.1	<0.1	<0.1
GARCH(1,2)	1.196	1.188	1.181	1.191	1.170
p-value	<0.1	<0.1	<0.1	<0.1	<0.1
GARCH(2,1)	1.201	1.193	1.175	1.184	1.182
p-value	<0.1	<0.1	<0.1	<0.1	<0.1
GARCH(2,2)	1.205	1.188	1.187	1.193	1.161
p-value	<0.1	<0.1	<0.1	<0.1	<0.1
TARCH(0,1)	1.089	1.108	1.108	1.133	1.133
p-value	<0.25	<0.25	<0.25	<0.25	<0.25
TARCH(0,2)	1.093	1.098	1.091	1.124	1.121
p-value	<0.25	<0.25	<0.25	<0.25	<0.25
TARCH(1,1)	1.196	1.192	1.186	1.195	1.711
p-value	<0.1	<0.1	<0.1	<0.1	<0.1
TARCH(1,2)	1.202	1.189	1.181	1.191	1.174
p-value	<0.1	<0.1	<0.1	<0.1	<0.1
TARCH(2,1)	1.204	1.173	1.156	1.191	1.176
p-value	<0.1	<0.1	<0.25	<0.1	<0.1
TARCH(2,2)	1.225	1.181	1.192	1.199	1.189
p-value	<0.05	<0.1	<0.1	<0.1	<0.1
E-GARCH(0,1)	*	1.013	1.012	1.033	1.029
p-value		>0.25	>0.25	>0.25	>0.25
E-GARCH(0,2)	1.036	1.050	1.047	1.076	1.063
p-value	>0.25	>0.25	>0.25	>0.25	>0.25
E-GARCH(1,1)	1.171	1.158	1.158	1.171	1.153
p-value	<0.1	<0.1	<0.1	<0.1	<0.25
E-GARCH(1,2)	1.165	1.155	1.152	1.163	1.153
p-value	<0.1	<0.25	<0.25	<0.1	<0.25
E-GARCH(2,1)	1.172	1.147	1.145	1.149	1.147
p-value	<0.1	<0.25	<0.25	<0.25	<0.25
E-GARCH(2,2)	1.232	1.175	*	1.162	*
p-value	<0.05	<0.1		<0.1	

(d) 15 May 1996–8 August 1996 (4th subperiod)
H_1: The model AR(3)-EGARCH(0,1) is equivalent to model X versus H_1: The model AR(3)-EGARCH(0,1) is "better" than model X

	Model for conditional mean				
	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
GARCH(0,1)	1.119	1.121	1.079	1.075	1.081
p-value	<0.25	<0.25	<0.25	>0.25	<0.25
GARCH(0,2)	1.156	1.176	1.134	1.130	1.128
p-value	<0.25	<0.1	<0.25	<0.25	<0.25
GARCH(1,1)	1.319	1.334	1.294	1.288	1.282
p-value	<0.01	<0.01	<0.05	<0.05	<0.05
GARCH(1,2)	1.322	1.353	1.313	1.308	1.297
p-value	<0.01	<0.01	<0.01	<0.05	<0.05
GARCH(2,1)	1.308	1.369	1.321	1.319	1.305
p-value	<0.01	<0.01	<0.01	<0.01	<0.05
GARCH(2,2)	1.313	1.342	1.328	1.298	1.298
p-value	<0.01	<0.01	<0.01	<0.05	<0.05
TARCH(0,1)	1.126	1.121	1.077	1.075	1.084
p-value	<0.25	<0.25	>0.25	>0.25	<0.25
TARCH(0,2)	1.165	1.171	1.128	1.127	1.128
p-value	<0.1	<0.1	<0.25	<0.25	<0.25
TARCH(1,1)	1.319	1.330	1.287	1.287	1.284
p-value	<0.01	<0.01	<0.05	<0.05	<0.05
TARCH(1,2)	1.323	1.342	1.300	1.299	1.293
p-value	<0.01	<0.01	<0.05	<0.05	<0.05
TARCH(2,1)	1.329	1.363	1.311	1.327	1.316
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
TARCH(2,2)	1.316	1.379	1.310	1.318	1.319
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
E-GARCH(0,1)	1.071	1.052	1.004	*	1.001
p-value	>0.25	>0.25	>0.25		>0.25
E-GARCH(0,2)	1.110	1.103	1.054	1.049	1.046
p-value	<0.25	<0.25	>0.25	>0.25	>0.25
E-GARCH(1,1)	1.317	1.323	1.275	1.271	1.273
p-value	<0.01	<0.01	<0.05	<0.05	<0.05
E-GARCH(1,2)	1.327	1.352	1.302	1.299	1.299
p-value	<0.01	<0.01	<0.05	<0.05	<0.05
E-GARCH(2,1)	1.308	1.368	1.325	1.325	1.328
p-value	<0.05	<0.01	<0.01	<0.01	<0.01
E-GARCH(2,2)	1.302	1.286	1.257	1.253	1.274
p-value	<0.05	<0.05	<0.05	<0.05	<0.05

Table AII. *Continued.*

(e) 9 August 1996–4 November 1996 (5th subperiod)
H_0: The model AR(1)-EGARCH(0,1) is equivalent to model X versus H_1: The model AR(1)-EGARCH(0,1) is "better" than model X

	Model for conditional mean				
	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
GARCH(0,1)	1.099	1.081	1.080	1.133	1.156
p-value	<0.25	<0.25	<0.25	<0.25	<0.25
GARCH(0,2)	1.157	1.129	1.127	1.171	1.189
p-value	<0.25	<0.25	<0.25	<0.1	<0.1
GARCH(1,1)	1.273	1.237	1.243	1.252	1.259
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
GARCH(1,2)	1.277	1.244	1.249	1.256	1.261
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
GARCH(2,1)	1.271	1.262	1.260	1.282	1.277
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
GARCH(2,2)	1.277	1.248	1.256	1.255	1.260
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
TARCH(0,1)	1.074	1.073	1.072	1.127	1.150
p-value	>0.25	>0.25	>0.25	<0.25	<0.25
TARCH(0,2)	1.127	1.119	1.116	1.162	1.183
p-value	<0.25	<0.25	<0.25	<0.1	<0.1
TARCH(1,1)	1.226	1.215	1.221	1.227	1.236
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
TARCH(1,2)	1.231	1.222	1.226	1.231	1.237
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
TARCH(2,1)	1.228	1.234	1.230	1.235	1.280
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
TARCH(2,2)	1.234	1.240	1.246	1.253	1.249
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
E-GARCH(0,1)	1.008	*	1.003	1.045	1.059
p-value	>0.25		>0.25	>0.25	>0.25
E-GARCH(0,2)	1.058	1.047	1.048	1.082	1.093
p-value	>0.25	>0.25	>0.25	<0.25	<0.25
E-GARCH(1,1)	1.201	1.213	1.213	1.225	1.228
p-value	<0.1	<0.05	<0.05	<0.05	<0.05
E-GARCH(1,2)	1.212	1.247	1.234	1.240	1.244
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
E-GARCH(2,1)	1.228	1.268	1.262	1.266	1.269
p-value	<0.05	<0.05	<0.05	<0.05	<0.05
E-GARCH(2,2)	1.194	1.231	1.229	*	1.246
p-value	<0.1	<0.05	<0.05		<0.05

*Model fails to converge at least once.

Table AIII. Percentage points of the CGR distribution for $a=0.05$.

$$\Phi(z) = \int_0^z \frac{(1-\rho^2)^k}{B(k,k)} x^{k-1}(1+x)^{-2k}\left[1-\left(\frac{2\rho}{x+1}\right)^2 x\right]^{-(2k+1)/2} \mathrm{d}x = 1-a$$

	ρ									
k	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
1	19.202	19.158	19.02	18.808	18.50	18.109	17.62	17.06	16.40	15.663
2	6.388	6.377	6.342	6.283	6.202	6.097	5.968	5.816	5.64	5.441
3	4.284	4.277	4.257	4.224	4.177	4.117	4.043	3.956	3.855	3.74
4	3.438	3.433	3.419	3.396	3.362	3.32	3.267	3.205	3.133	3.051
5	2.978	2.975	2.963	2.945	2.919	2.885	2.844	2.795	2.739	2.674
6	2.687	2.684	2.674	2.659	2.637	2.609	2.575	2.535	2.487	2.434
7	2.484	2.481	2.473	2.46	2.441	2.417	2.388	2.353	2.312	2.265
8	2.333	2.331	2.324	2.312	2.296	2.275	2.249	2.218	2.182	2.141
9	2.217	2.215	2.209	2.198	2.184	2.164	2.141	2.113	2.081	2.044
10	2.124	2.122	2.117	2.107	2.093	2.076	2.055	2.029	2	1.966
11	2.048	2.046	2.041	2.032	2.019	2.003	1.984	1.96	1.933	1.902
12	1.984	1.982	1.977	1.969	1.957	1.943	1.924	1.902	1.877	1.848
13	1.929	1.928	1.923	1.915	1.905	1.891	1.874	1.853	1.829	1.802
14	1.882	1.881	1.876	1.869	1.859	1.846	1.83	1.81	1.788	1.762
15	1.841	1.84	1.835	1.829	1.819	1.807	1.791	1.773	1.752	1.727
16	1.804	1.803	1.799	1.793	1.784	1.772	1.757	1.74	1.72	1.697
17	1.772	1.771	1.767	1.761	1.752	1.741	1.727	1.711	1.691	1.669
18	1.743	1.742	1.738	1.732	1.724	1.713	1.7	1.684	1.666	1.644
19	1.717	1.716	1.712	1.706	1.698	1.688	1.675	1.66	1.643	1.622
20	1.693	1.692	1.688	1.683	1.675	1.665	1.653	1.638	1.621	1.602
21	1.671	1.67	1.667	1.661	1.654	1.644	1.633	1.619	1.602	1.583
22	1.651	1.65	1.647	1.642	1.635	1.625	1.614	1.6	1.584	1.566
23	1.632	1.631	1.629	1.624	1.617	1.608	1.597	1.584	1.568	1.55
24	1.615	1.614	1.612	1.607	1.6	1.591	1.581	1.568	1.553	1.536
25	1.599	1.599	1.596	1.591	1.585	1.576	1.566	1.553	1.539	1.522
26	1.585	1.584	1.581	1.577	1.57	1.562	1.552	1.54	1.526	1.51
27	1.571	1.57	1.567	1.563	1.557	1.549	1.539	1.527	1.514	1.498
28	1.558	1.557	1.555	1.55	1.544	1.536	1.527	1.515	1.502	1.487
29	1.546	1.545	1.542	1.538	1.532	1.525	1.516	1.504	1.491	1.476
30	1.534	1.534	1.531	1.527	1.521	1.514	1.505	1.494	1.481	1.466
35	1.486	1.485	1.483	1.479	1.474	1.467	1.459	1.449	1.438	1.425
40	1.448	1.447	1.445	1.442	1.437	1.431	1.424	1.415	1.404	1.392
45	1.417	1.416	1.415	1.412	1.407	1.402	1.395	1.386	1.377	1.366
50	1.392	1.391	1.389	1.387	1.383	1.377	1.371	1.363	1.354	1.344
55	1.37	1.37	1.368	1.365	1.362	1.357	1.351	1.343	1.335	1.325
60	1.352	1.351	1.35	1.347	1.344	1.339	1.333	1.327	1.319	1.309
k	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
1	14.835	13.92	12.91	11.83	10.65	9.392	8.041	6.596	5.049	3.368
2	5.217	4.969	4.696	4.397	4.072	3.719	3.336	2.919	2.456	1.923
3	3.611	3.467	3.309	3.135	2.944	2.736	2.507	2.255	1.971	1.633
4	2.959	2.856	2.742	2.616	2.478	2.327	2.159	1.973	1.76	1.503
5	2.601	2.52	2.429	2.33	2.22	2.098	1.964	1.813	1.64	1.428
6	2.373	2.305	2.229	2.145	2.053	1.951	1.837	1.709	1.56	1.377
7	2.213	2.154	2.088	2.016	1.935	1.846	1.747	1.634	1.503	1.34
8	2.094	2.042	1.984	1.919	1.847	1.768	1.679	1.578	1.46	1.312

Table AIII. *Continued.*

	ρ									
9	2.002	1.954	1.902	1.843	1.779	1.706	1.625	1.533	1.425	1.29
10	1.927	1.884	1.836	1.783	1.723	1.657	1.582	1.497	1.397	1.272
11	1.866	1.826	1.782	1.732	1.677	1.616	1.546	1.467	1.374	1.256
12	1.815	1.778	1.736	1.69	1.638	1.581	1.516	1.442	1.354	1.243
13	1.771	1.736	1.697	1.654	1.605	1.551	1.49	1.42	1.337	1.232
14	1.733	1.7	1.663	1.622	1.577	1.525	1.467	1.401	1.322	1.222
15	1.7	1.669	1.634	1.595	1.551	1.502	1.447	1.384	1.309	1.213
16	1.67	1.641	1.607	1.57	1.529	1.482	1.43	1.369	1.297	1.205
17	1.644	1.616	1.584	1.548	1.509	1.464	1.414	1.356	1.287	1.198
18	1.62	1.593	1.563	1.529	1.491	1.448	1.399	1.344	1.277	1.192
19	1.599	1.573	1.544	1.511	1.474	1.433	1.386	1.333	1.269	1.186
20	1.58	1.554	1.526	1.495	1.459	1.42	1.375	1.323	1.261	1.181
21	1.562	1.538	1.51	1.48	1.446	1.407	1.364	1.313	1.253	1.176
22	1.545	1.522	1.496	1.466	1.433	1.396	1.354	1.305	1.247	1.171
23	1.53	1.508	1.482	1.454	1.421	1.385	1.344	1.297	1.24	1.167
24	1.516	1.494	1.47	1.442	1.411	1.376	1.336	1.29	1.234	1.163
25	1.503	1.482	1.458	1.431	1.401	1.367	1.328	1.283	1.229	1.159
26	1.491	1.47	1.447	1.421	1.391	1.358	1.32	1.276	1.224	1.156
27	1.48	1.46	1.437	1.411	1.382	1.35	1.313	1.27	1.219	1.153
28	1.469	1.449	1.427	1.402	1.374	1.343	1.307	1.265	1.215	1.15
29	1.459	1.44	1.418	1.394	1.366	1.336	1.3	1.26	1.211	1.147
30	1.45	1.431	1.41	1.386	1.359	1.329	1.294	1.255	1.207	1.144
35	1.41	1.393	1.374	1.352	1.328	1.301	1.269	1.233	1.189	1.132
40	1.378	1.363	1.345	1.326	1.303	1.278	1.25	1.216	1.176	1.123
45	1.353	1.339	1.322	1.304	1.283	1.26	1.233	1.202	1.165	1.116
50	1.332	1.318	1.303	1.286	1.267	1.245	1.22	1.191	1.156	1.109
55	1.314	1.301	1.287	1.271	1.253	1.232	1.209	1.181	1.148	1.104
60	1.299	1.287	1.273	1.258	1.241	1.221	1.199	1.173	1.141	1.099

Figure A1 depicts the one step ahead 95% prediction intervals for the models with the lowest $\sum_{t=T+1}^{T+s}(\hat{z}_{t|t-1}^2)$ in each subperiod.

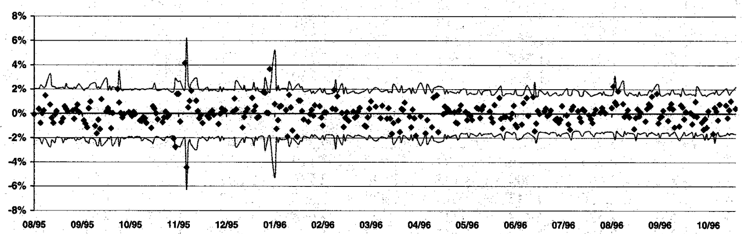


Figure A1. One step ahead 95% forecasted interval for the models with the lowest sum of the squared standardized one step ahead prediction errors.

Figure A2 presents the in-sample 95% confidence interval for the AR(1)GARCH(1,1) model.
Figures A3 and A4 present the one step ahead 95% prediction intervals for the models selected by the SBC and AIC, respectively.
The probability density function of the CGR distribution is presented in Figure A5.

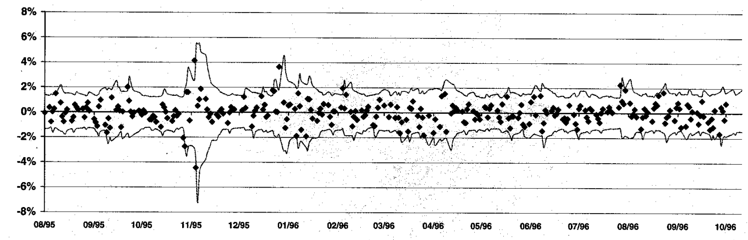


Figure A2. In-sample 95% confidence interval for the AR(1) GARCH(1,1) model.

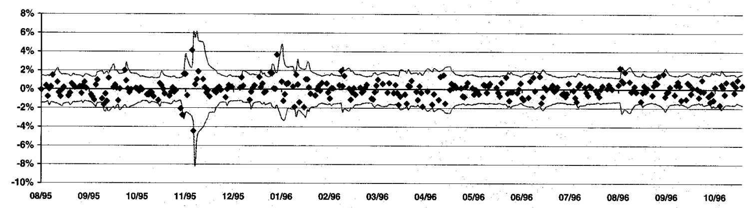


Figure A3. One step ahead 95% forecasted intervals for the models selected by the SBC.

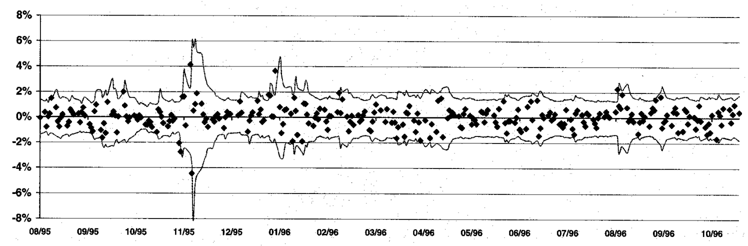
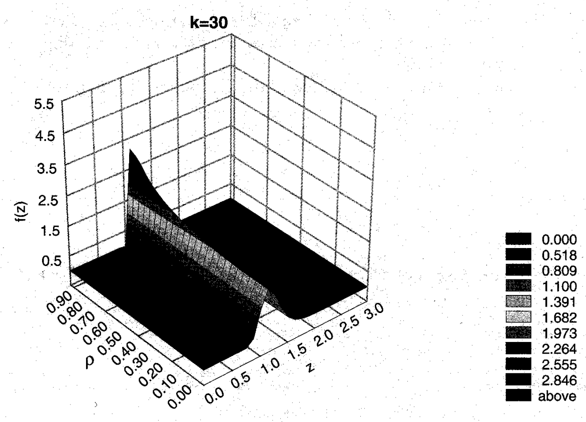


Figure A4. One step ahead 95% forecasted intervals for the models selected by the AIC.



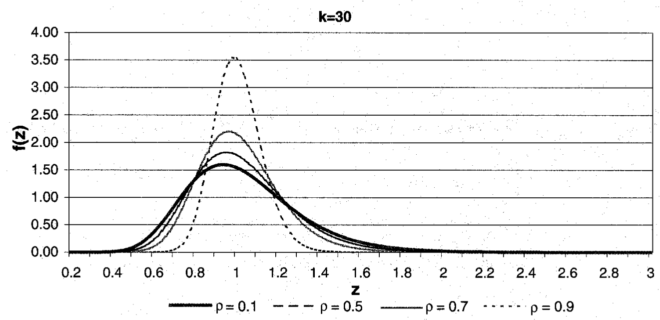


Figure A5. The probability density function of the CGR distribution.

$$f(z) = \frac{(1-\rho^2)^k}{B(k,k)} z^{k-1}(1+z)^{-2k}\left[1-\left(\frac{2\rho}{z+1}\right)^2 z\right]^{-(2k+1)/2} \quad \text{for } z \geqslant 0,\ 0 \leqslant \rho < 1,\ k = 30.$$

ACKNOWLEDGEMENTS

We would like to thank the associate editor and the referee for their constructive comments that helped improve an earlier version of the paper.

REFERENCES

1. Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 1982; **50**(4):987–1007.
2. Bollerslev T. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 1986; **31**:307–327.
3. Nelson DB. Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* 1991; **59**:347–370.
4. Glosten LR, Jagannathan R, Runkle DE. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 1993; **48**:1779–1801.
5. Zakoian JM. Threshold heteroskedasticity models. Manuscript, CREST, INSEE, 1990.
6. Sentana E. Quadratic ARCH models. *Review of Economic Studies* 1995; **62**:639–661.
7. Degiannakis S, Xekalaki E. Autoregressive conditional heteroscedasticity models: a review. *Quality Technology and Quantitative Management* 2004; **1**(2):271–324.
8. Xekalaki E, Panaretos J, Psarakis S. A predictive model evaluation and selection approach—the correlated gamma ratio distribution. In *Stochastic Musings: Perspectives from the Pioneers of the Late 20th Century*, Panaretos J (ed.). Lawrence Erlbaum Associates Publishers: London, 2003; 188–202.
9. Bollerslev T. A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 1987; **69**:542–547.
10. Harvey AC. *The Econometric Analysis of Time Series*. Philip Allan: Oxford, 1981.
11. Box GEP, Tiao GC. *Bayesian Inference in Statistical Analysis*. Addison-Wesley: Reading, MA, 1973.
12. Bollerslev T, Engle RF, Nelson D. ARCH models. In *Handbook of Econometrics*, vol. IV, Chapter 49. Engle RF, McFadden DL (eds). Elsevier: Amsterdam, 1994.
13. Jorion P. On jump processes in the foreign exchange and stock markets. *Review of Financial Studies* 1988; **1**:427–445.
14. Hsieh DA. Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business and Economic Statistics* 1989; **7**:307–317.
15. Cai J. A Markov model of switching-regime ARCH. *Journal of Business and Economic Statistics* 1994; **12**:309–316.
16. Hamilton J, Susmel R. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 1994; **64**:307–333.
17. Scholes M, Williams J. Estimating betas from non-synchronous data. *Journal of Financial Economics* 1977; **5**: 309–327.
18. Lo A, MacKinlay C. Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies* 1988; **1**:41–66.
19. Campbell J, Lo A, MacKinlay AC. *The Econometrics of Financial Markets*. Princeton University Press: New Jersey, 1997.
20. Mandelbrot B. The variation of certain speculative prices. *Journal of Business* 1963; **36**:394–419.
21. Black F. Studies of stock market volatility changes. *Proceedings of the American Statistical Association*, Business and Economic Statistics Section, 1976; 177–181.
22. Bera AK, Higgins ML. ARCH models: properties, estimation and testing. *Journal of Economic Surveys* 1993; **7**: 305–366.
23. Bollerslev T, Chou RC, Kroner K. ARCH modelling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* 1992; **52**:5–59.
24. Gourieroux C. *ARCH Models and Financial Applications*. Springer Series. Springer: Berlin, 1997.
25. Hamilton J. *Time Series Analysis*. Princeton University Press: New Jersey, 1994.
26. Akaike H. Information theory and the extension of the maximum likelihood principle. *Proceeding of the Second International Symposium on Information Theory*, Budapest, 1973; 267–281.
27. Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978; **6**:461–464.
28. Pagan AR, Schwert GW. Alternative models for conditional stock volatility. *Journal of Econometrics* 1990; **45**: 267–290.
29. Andersen T, Bollerslev T, Lange S. Forecasting financial market volatility: sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance* 1999; **6**:457–477.
30. Heynen R, Kat H. Volatility prediction: a comparison of the stochastic volatility, GARCH(1,1), and EGARCH(1,1) models. *Journal of Derivatives* 1994; **94**:50–65.
31. Hol E, Koopman S. Forecasting the variability of stock index returns with stochastic volatility models and implied volatility. *Discussion Paper No. 104, 4*, Tinbergen Institute, 2000.
32. West KD, Edison HJ, Cho D. A utility based comparison of some models for exchange rate volatility. *Journal of International Economics* 1993; **35**:23–45.
33. Engle RF, Hong CH, Kane A, Noh J. Arbitrage valuation of variance forecasts with simulated options. *Advances in Futures and Options Research* 1993; **6**:393–415.
34. Greene WH. *Econometric Analysis* (3rd edn). Prentice-Hall: Englewood Cliffs, NJ, 1997.
35. Kibble WF. A two variate gamma type distribution. *Sankhya* 1941; **5**:137–150.
36. Jarque CM, Bera AK. Efficient tests for normality, heteroscedasticity and serial independence of regression residuals. *Economic Letters* 1980; **6**:255–259.

37. Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 1979; **74**:427–431.
38. Phillips PCB. Time series regression with a unit root. *Econometrica* 1987; **55**:277–301.
39. Phillips PCB, Perron P. Testing for a unit root in time series regression. *Biometrika* 1988; **75**:335–346.
40. Marquardt DW. An algorithm for least squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* 1963; **11**:431–441.
41. Berndt ER, Hall BH, Hall RE, Hausman JA. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement* 1974; **3**:653–665.
42. Bollerslev T, Wooldridge JM. Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews* 1992; **11**:143–172.
43. Richardson M, Smith T. A unified approach to testing for serial correlation in stock returns. *Journal of Business* 1994; **67**:371–399.
44. Box GEP, Pierce DA. Distribution of residual autocorrelation in ARIMA time series models. *Journal of the American Statistical Association* 1970; **65**:1509–1526.
45. Dueker MJ. Markov switching in GARCH processes and mean-reverting stock market volatility. *Journal of Business and Economic Statistics* 1997; **15**:26–34.
46. Hansen BE. Autoregressive conditional density estimation. *International Economic Review* 1994; **35**:705–730.

