Predictability and Model Selection in the Context of ARCH Models

Degiannakis, Stavros and Xekalaki, Evdokia

Department of Statistics, Athens University of Economics and Business

2005

Predictability and Model Selection in the Context of ARCH Models

Stavros Degiannakis and Evdokia Xekalaki

Department of Statistics, Athens University of Economics and Business

Abstract

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.

Keywords and Phrases: 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 analyzing 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 (1982), the conditional variance is postulated to be a linear function of the past q squared innovations. Bollerslev (1986) 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 (1991) 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. (1993), the threshold GARCH, or TARCH, model, introduced by Zakoian (1990) and the quadratic ARCH, or QGARCH, model, introduced by Sentana (1995). ARCH models go by such exotic names as AARCH, NARCH, PARCH, PNP-ARCH and STARCH among others. For a comprehensive review of the literature on such models, the interested reader is referred to Degiannakis and Xekalaki (2004).

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 (2003) 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 $\{y_t(\theta)\}_{t\geq1}$ 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, $\{\varepsilon_t(\theta)\}_{t\geq1}$, 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, $\{\varepsilon_t(\theta)\}_{t\geq1}$, can be presented as:

$$
\begin{aligned}
y_t(\theta) &= x_t'\beta + \varepsilon_t(\theta) \\
\varepsilon_t(\theta) &= z_t\sigma_t(\theta) \\
z_t &\overset{i.i.d.}{\sim} f[E(z_t)=0, V(z_t)=1] \\
\sigma_t^2(\theta) &= g(\sigma_{t-1}(\theta), \sigma_{t-2}(\theta),...; \varepsilon_{t-1}(\theta), \varepsilon_{t-2}(\theta),...; \upsilon_{t-1}, \upsilon_{t-2},...),
\end{aligned}
\tag{2.1}
$$

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 $t-1$, υ_t is a vector of predetermined variables included in I_t, and $g(.)$ is a linear or nonlinear 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 (1987) proposed using the student t distribution with an estimated kurtosis regulated by the degrees of freedom parameter. Nelson (1991) proposed the use of the generalized error distribution (Harvey (1981), Box and Tiao (1973)), which is also referred to as the exponential power distribution. Other distributions, that have been employed, include the generalized t distribution (Bollerslev et al. (1994)), the normal Poisson mixture distribution (Jorion (1988)), the normal lognormal mixture (Hsieh (1989)), and a serially dependent mixture of normally distributed variables (Cai (1994)) or student t distributed variables (Hamilton and Susmel (1994)). 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, $\mu_t = E(y_t \mid I_{t-1})$, can be adequately described by a κth order autoregressive $[AR(\kappa)]$ model:

$$y_t = c_0 + \sum_{i=1}^{\kappa}(c_i y_{t-i}) + \varepsilon_t. \qquad (2.2)$$

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 (Scholes and Williams (1977), Lo and MacKinlay (1988)). According to Campbell et al. (1997), "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 1st 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 (1982) introduced the original form of $\sigma_t^2 = g(.)$ as a linear function of the past q squared innovations:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q}(a_i \varepsilon_{t-i}^2). \qquad (2.3)$$

For the conditional variance to be positive, the parameters must satisfy $\alpha_0 > 0$, $a_i \geq 0$, for $i = 1,...,q$. 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 (1986) proposed the generalized ARCH, or GARCH(p,q), model:

4

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q}\left(a_i\varepsilon_{t-i}^2\right) + \sum_{i=1}^{p}\left(b_i\sigma_{t-i}^2\right), \tag{2.4}$$

where $\alpha_0 > 0$, $a_i \geq 0$, for $i = 1,...,q$, and $b_i \geq 0$, for $i = 1,...,p$. 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:

$$E_t\left(\varepsilon_t^2\right) - E_{t-1}\left(\varepsilon_t^2\right) = \varepsilon_t^2 - \sigma_t^2 \equiv v_t. \tag{2.5}$$

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 (1963): "… 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 behavior of stock market prices where a *leverage effect* may be present. The term *leverage effect*, first noted by Black (1976), 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 (1991) exponential GARCH, or EGARCH(p,q), model:

$$\ln\left(\sigma_t^2\right) = a_0 + \sum_{i=1}^{q}\left(a_i\left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right| + \gamma_i\left(\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right)\right) + \sum_{i=1}^{p}\left(b_i\ln\left(\sigma_{t-i}^2\right)\right). \tag{2.6}$$

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:

$$\sigma_t^\delta = a_0 + \sum_{i=1}^{q}\left(a_i\left|\varepsilon_{t-i}\right|^\delta\right) + \gamma\left|\varepsilon_{t-1}\right|^\delta d(\varepsilon_{t-1} \le 0) + \sum_{i=1}^{p}\left(b_i\sigma_{t-i}^\delta\right), \qquad (2.7)$$

where $d(\varepsilon_t \le 0) = 1$ if $\varepsilon_t \le 0$, and $d(\varepsilon_t \le 0) = 0$ otherwise. Zakoian's (1990) model is a special case of the TARCH model with $\delta = 1$, while Glosten et al. (1993) 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 (1993), Bollerslev et al. (1992), Bollerslev et al. (1994), Degiannakis and Xekalaki (2004), Gourieroux (1997) and Hamilton (1994).

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) (Akaike (1973)) and the Schwarz Bayesian Criterion (SBC) (Schwarz (1978)) 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:

$$AIC = l_T(\hat{\theta}) - \breve{\theta} \qquad (3.1)$$

$$SBC = l_T(\hat{\theta}) - 2^{-1}\breve{\theta}\ln(T). \qquad (3.2)$$

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):

$$MSE = T^{-1}\sum_{t=1}^{T}\varepsilon_t^2. \qquad (3.3)$$

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. (1994) present four types of loss functions:

$$L_1 = \sum_{t=1}^{T} \left(\varepsilon_t^2 - \sigma_t^2 \right)^2, \tag{3.4}$$

$$L_2 = \sum_{t=1}^{T} \ln \left(\frac{\varepsilon_t^2}{\sigma_t^2} \right)^2, \tag{3.5}$$

$$L_3 = \sum_{t=1}^{T} \frac{\left(\varepsilon_t^2 - \sigma_t^2 \right)^2}{\sigma_t^4}, \tag{3.6}$$

$$L_4 = \sum_{t=1}^{T} \left(\frac{\varepsilon_t^2}{\sigma_t^2} + \ln \left(\sigma_t^2 \right) \right). \tag{3.7}$$

Pagan and Schwert (1990) used the first two of the loss functions to compare alternative estimators with in-sample and out-of-sample data sets. Andersen et al. (1999), Heynen and Kat (1994), Hol and Koopman (2000), 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. (1993) developed such a criterion based on the portfolio decisions of a risk averse investor. Engle et al. (1993) assumed that the objective was to price options and developed a loss function from the profitability of a particular trading strategy.

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 $\varepsilon_t \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right)$. 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 $\hat{\beta}_{t-1} = \left(\mathbf{X}_{t-1}' \mathbf{X}_{t-1} \right)^{-1} \left(\mathbf{X}_{t-1}' \mathbf{Y}_{t-1} \right)$ is the least square estimator of β at time $t-1$, $\mathbf{Y}_t$ is the $\left(l_t \times 1 \right)$ vector of l_t observations on the dependent variable y_t, and $\mathbf{X}_t$ is the $\left(l_t \times k \right)$ 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 $|\mathbf{X}_t'\mathbf{X}_t| \neq 0$, $t = 0,1,...$. 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 (2.1), is assumed to be that of

the normal distribution and $\hat{z}_{t|t-1} \equiv \hat{\varepsilon}_{t|t-1}\hat{\sigma}_{t|t-1}^{-1}$ denotes the standardized one step ahead

prediction errors[1]. The most commonly used way to model the conditional variance is the

GARCH(p,q) process in (2.4). The GARCH(p,q) process may be rewritten as[2]:

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

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

$\omega' = (b_1,..., b_p)$.

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. (2003) suggested measuring the

predictive behavior 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:

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

---

[1] Consider the case of the AR(1)GARCH(1,1) model as defined by equations (2.2) and (2.4), 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 $\hat{\sigma}_{t|t-1}^2 = \hat{a}_{0,t-1} + \hat{a}_{1,t-1}\hat{\varepsilon}_{t-1|t-1}^2 + \hat{b}_{1,t-1}\hat{\sigma}_{t-1|t-1}^2$, respectively. The estimated parameters are indexed by the subscript t to indicate that they may vary with time.
[2] 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.

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

$$R_T = \frac{\displaystyle\sum_{t=1}^{T} \hat{z}_{t|t-1}^2}{T} \,. \tag{4.1}$$

$\hat{z}_{t|t-1} \equiv \hat{\varepsilon}_{t|t-1}\hat{\sigma}_{t|t-1}^{-1}$ 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, $V\left(y_t \mid I_{t-1}\right) \equiv \sigma_t^2$.

Let $\left(\theta_t\right)$ denote the vector of unknown parameters to be estimated at time t. Under the assumption of constancy of parameters over time, $\left(\theta_1\right) = \left(\theta_2\right) = ... = \left(\theta_T\right) = \left(\theta\right)$, the estimated standardized one step ahead prediction errors $\hat{z}_{t|t-1}, \hat{z}_{t+1|t},..., \hat{z}_{T|T-1}$ are asymptotically independently standard normally distributed. Symbolically,

$$\hat{z}_{t|t-1} \equiv \left(y_t - \hat{y}_{t|t-1}\right)\hat{\sigma}_{t|t-1}^{-1} \sim N\left(0,1\right), \; t = 1,2,...,T. \tag{4.2}$$

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. $\hat{\sigma}_{t|t-1}^2 = \left(u'_t, \eta'_t, w'_t\right)\left(\hat{v}_{t-1}, \hat{\zeta}_{t-1}, \hat{\omega}_{t-1}\right)$. Note that the elements of the vector $\left(u'_t, \eta'_t, w'_t\right)$ belong to the I_{t-1}, so are considered as known values. The $\hat{z}_{t|t-1}$ can be written as:

$$\hat{z}_{t|t-1} = \frac{\left(y_t - \hat{y}_{t|t-1}\right)}{\sqrt{\hat{\sigma}^2_{t|t-1}}} =$$

$$= \frac{\left(x_t'\beta + \varepsilon_t - x_t'\hat{\beta}_{t-1}\right)}{\sqrt{\hat{\sigma}^2_{t|t-1}}} =$$

$$= \frac{\varepsilon_t}{\sqrt{\hat{\sigma}^2_{t|t-1}}} + \frac{\left(x_t'\left(\beta - \hat{\beta}_{t-1}\right)\right)}{\sqrt{\hat{\sigma}^2_{t|t-1}}} =$$

$$= \frac{z_t\sqrt{\sigma^2_t}}{\sqrt{\hat{\sigma}^2_{t|t-1}}} + \frac{\left(x_t'\left(\beta - \hat{\beta}_{t-1}\right)\right)}{\sqrt{\hat{\sigma}^2_{t|t-1}}} =$$

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

We assume that a sample of T observations has been used to estimate the vector of unknown parameters. According to Bollerslev (1986), 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. Greene (1997, 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

$$p\lim(\hat{z}_{t|t-1}) =$$

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

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

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

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

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

$$= z_t.$$

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

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

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

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

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

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

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

Hence, (4.2) has been established.

The result of formula (4.2) 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 $u_t' = (1, |\varepsilon_{t-1}/\sigma_{t-1}|, ..., |\varepsilon_{t-q}/\sigma_{t-q}|)$, $\eta_t' = ([\varepsilon_{t-1}/\sigma_{t-1}]..., [\varepsilon_{t-q}/\sigma_{t-q}])$,

$w_t' = (\ln \sigma_{t-1}^2, ..., \ln \sigma_{t-p}^2)$, $v' = (a_0, a_1, ..., a_q)$, $\zeta' = (\gamma_1, ..., \gamma_q)$, $\omega' = (b_1, ..., b_p)$.

According to Nelson (1991), 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 (1993) TARCH(p,q) process, the conditional variance can be written as:

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

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

As pointed out by Glosten et al. (1993), 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 $g\left(\hat{z}_{t|t-1}\right) = \sum_{t=1}^{T}\left(\hat{z}_{t|t-1}^2\right)$, which is a continuous function, then $p\lim \sum_{t=1}^{T}\left(\hat{z}_{t|t-1}^2\right) = \sum_{t=1}^{T}\left(z_t^2\right)$. As convergence in probability implies convergence in distribution, $\sum_{t=1}^{T}\left(\hat{z}_{t|t-1}^2\right) \xrightarrow{d} \sum_{t=1}^{T}\left(z_t^2\right) \sim \chi_T^2$. 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.,

$$TR_T \xrightarrow{d} \chi_T^2. \tag{4.3}$$

According to Kibble (1941), if, for $t = 1,2,...,T$, $\hat{z}_{t|t-1}^{(A)}$ and $\hat{z}_{t|t-1}^{(B)}$ are standard normally distributed variables, following jointly the bivariate standard normal distribution, then the joint distribution of $\left(\dfrac{T}{2}R_T^{(A)}, \dfrac{T}{2}R_T^{(B)}\right)$ is the bivariate gamma distribution with probability density function (p.d.f) given by:

$$f_{\frac{T}{2}R_T^{(A)}, \frac{T}{2}R_T^{(B)}}(x,y) = \frac{\exp\left(-\dfrac{x+y}{1-\rho^2}\right)}{\Gamma(T/2)\left(1-\rho^2\right)^{T/2}} \sum_{i=0}^{\infty}\left(\frac{\left(\rho/\left(1-\rho^2\right)\right)^{2i}}{\Gamma(i+1)\Gamma\left(i+(T/2)\right)}(xy)^{(T/2)-1-i}\right), x,y>0, \tag{4.4}$$

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. $\rho \equiv Cor\left(\hat{z}_{t|t-1}^{(A)}, \hat{z}_{t|t-1}^{(B)}\right)$. Xekalaki et al. (2003) showed that, when the joint distribution of $\left(\dfrac{T}{2}R_T^{(A)}, \dfrac{T}{2}R_T^{(B)}\right)$ 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.:

$$f_{Z_T^{(A,B)}}(z) = \frac{\left(1-\rho^2\right)^{T/2}}{B\left(T/2, T/2\right)} z^{T/2-1}(1+z)^{-T}\left[1-\left(\frac{2\rho}{z+1}\right)^2 z\right]^{-\frac{T+1}{2}}, z>0, \tag{4.5}$$

where $B\left(\dfrac{T}{2}, \dfrac{T}{2}\right) = \Gamma\left(\dfrac{T}{2}\right)^2 \Big/ \Gamma(T)$. Symbolically,

$$Z_T^{(A,B)} \equiv \sum_{t=1}^{T} \hat{z}_{t|t-1}^{2(B)} \Bigg/ \sum_{t=1}^{T} \hat{z}_{t|t-1}^{2(A)} \sim CGR(k,\rho), \qquad (4.6)$$

where $k = T/2$. Xekalaki et al. (2003) referred to the distribution in (4.5) as the Correlated gamma ratio (CGR) distribution. In the Appendix, Figure 8 depicts its probability density function for $z \geq 0$, $0 \leq \rho < 1$ and $k = 30$ and Table 5 presents a sample of the 95th percentile of the CGR distribution. Full tables of the CGR percentage points and of graphs depicting its probability density function can be found in Degiannakis and Xekalaki (1999).

As pointed out by Xekalaki et al. (2003), $R_T^{(A)}$ 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 Xekalaki et al.'s (2003) sense to be defined implicitly through their mean squared prediction errors. Following Xekalaki et al.'s (2003) 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 $Z_T^{(A,B)} > CGR(k,\rho,a)$, where $CGR(k,\rho,a)$ 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 1 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 (Jarque and Bera (1980)) 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:

$$JB = T\left(S^2 + \left((K-3)^2/4\right)\right)/6, \tag{5.1}$$

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 2 degrees of freedom.

Table (1). Descriptive Statistics of the daily returns of the ASE index	
(30th August 1993 to 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 2 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. (4 lagged differences).

PP: The null hypothesis of non-stationarity is rejected if the PP value is less than the critical value. (4 truncation lags).

From Table 1, 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) (Dickey and Fuller (1979)) and the nonparametric Phillips Perron (PP) test (Phillips (1987), Phillips and Perron (1988)) are conducted.

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

$$\Delta y_t = c + \gamma y_{t-1} + \sum_{i=1}^{\kappa} \varphi_i \Delta y_{t-i} + \varepsilon_t, \qquad (5.2)$$

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:

$$\Delta y_t = c + \gamma y_{t-1} + \varepsilon_t. \qquad (5.3)$$

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 nonparametric 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.

Table (2). Lagrange multiplier (LM) test. Test the null hypothesis of no ARCH effects in the residuals up to order q.		
$$\varepsilon_t^2 = \beta_0 + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}^2 + u_t$$ $$\varepsilon_t = y_t - c$$		
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.		

The most commonly used test, for examining the null hypothesis of homoscedasticity against the alternative hypothesis of heteroscedasticity, is Engle's (1982) 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

$$\varepsilon_t^2 = \beta_0 + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}^2 + u_t,$$
(5.4)

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 2, the values of the LM test statistic for $q = 1,...,8$ 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 considered in the sequel. It is assumed, specifically, that the conditional mean is considered as a κth order autoregressive process as defined in (2.2) 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 (2.4), (2.6) and (2.7), respectively. Thus, the AR(κ)GARCH(p,q), AR(κ)EGARCH(p,q) and AR(κ)TARCH(p,q) models[3] are applied, for $\kappa = 0,...,4$, $p = 0, 1, 2$ 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 (Marquardt (1963)) is used. This algorithm modifies the Berndt, Hall,

---

[3] Glosten's et al. (1993) TARCH model is applied with $\delta = 2$.

Hall and Hausman, or BHHH, algorithm (Berndt et al. (1974)) 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:

$$\theta^{(i+1)} = \theta^{(i)} + \left(\sum_{t=1}^{T} \frac{\partial l_t^{(i)}}{\partial \theta} \frac{\partial l_t^{(i)}}{\partial \theta'} - aI \right)^{-1} \sum_{t=1}^{T} \frac{\partial l_t^{(i)}}{\partial \theta}, \qquad (5.5)$$

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 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 (1992), 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 5 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, $\sum_{t=T+1}^{T+s} \left(\hat{z}_{t|t-1}^2 \right)$, is estimated for each model and presented in Table 3, in the Appendix. The models selected for each subperiod and their sums of the squared standardized one step ahead prediction errors are:

Subperiod	Model Selected	$\min\left(\sum_{t=T+1}^{T+s} \left(\hat{z}_{t|t-1}^2 \right) \right)$
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.

Figure 1. The ASE index and the one step ahead conditional standard deviation of its returns estimated by the EGARCH(0,1) models
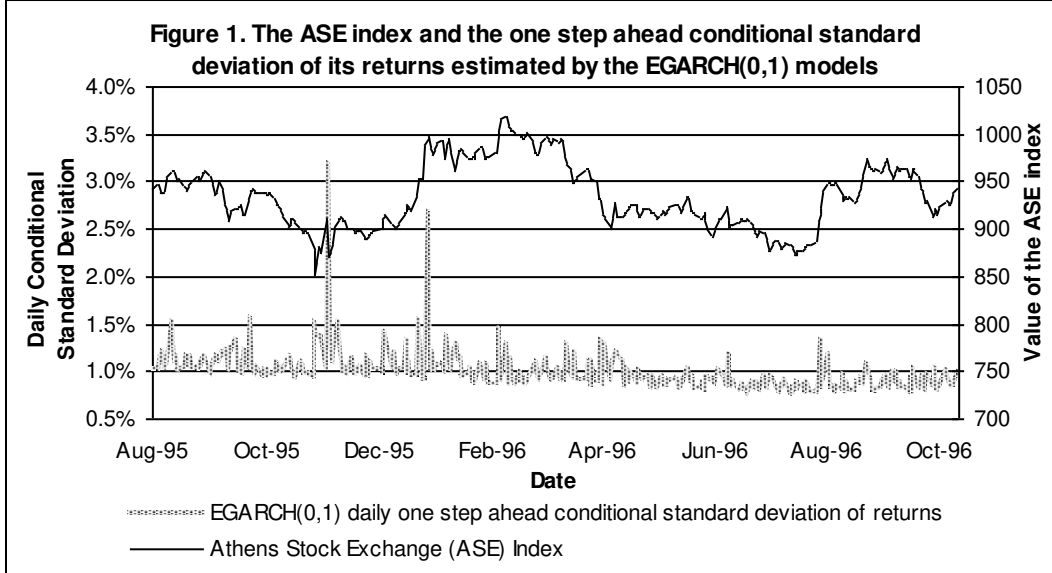
Despite the fact that an asymmetric model is selected by the SPEC algorithm, there are no asymmetries in the ASE index volatility. According to Figure 1, the major episodes of high volatility are not associated with market changes of the same sign. Figure 2 presents the values of the parameters a_1 and 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

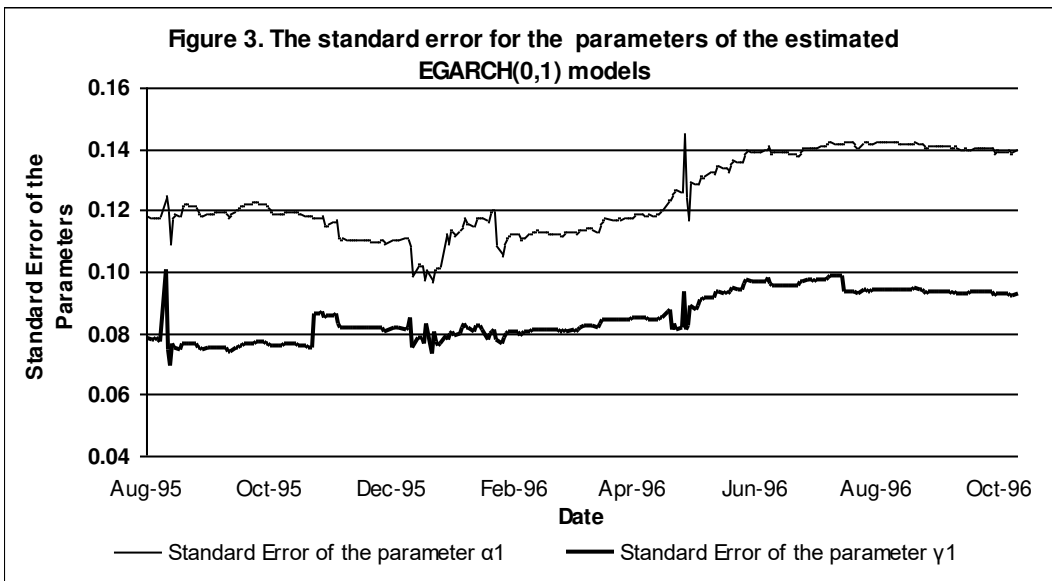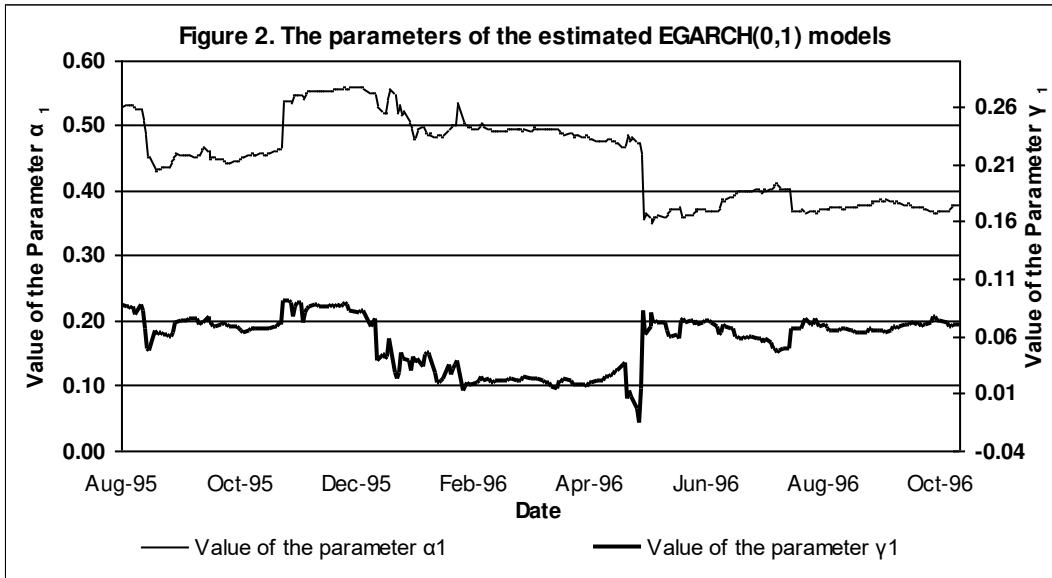$$y_t = c_0 + c_1 y_{t-1} + c_2 y_{t-2} + \varepsilon_t$$

$$\ln\left(\sigma_t^2\right) = a_0 + a_1 \left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma_1 \left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right),$$ (5.6)

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

H₀: The model AR(2)EGARCH(0,1) has equivalent predictive ability to model X

18

is tested versus

H₁: The model AR(2)EGARCH(0,1) produces "better" predictions than model X, with X denoting any one of the remainder models.

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

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

Note that the correlation between the standardized one step ahead prediction errors is greater than 0.9 in each case. If $Z_{60}^{AR(2)EGARCH(0,1),X} \equiv (21.96)^{-1} \sum_{t=501}^{560} \hat{z}_{t|t-1}^{(X)2}$ $> CGR(k = 30, \rho > 0.9, a)$, the null hypothesis of equivalent predictive ability of the models is rejected at $100a\%$ level of significance and the AR(2)EGARCH(0,1) model is regarded as "better" than model X. Table 4, in the Appendix, summarizes the results of the hypothesis tests, for each subperiod.

Figure 4, in the Appendix, depicts the one step ahead 95 per cent prediction intervals for the models with the lowest $\sum_{t=T+1}^{T+s}\left(\hat{z}_{t|t-1}^2\right)$ 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: $\hat{\mu}_{t|t-1}\pm1.96\hat{\sigma}_{t|t-1}$. 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 4 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 (2.2), Richardson and Smith (1994) developed a test for autocorrelation. It is a robust version of the standard Box Pierce (Box and Pierce (1970)) procedure. For p_i denoting the estimated autocorrelation between the returns at time t and $t-i$, the test is formulated as:

$$RS(r)=T\sum_{i=1}^{r}\frac{p_i^2}{1+c_i},\qquad(6.1)$$

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

$$c_i=\frac{Cov\left(\bar{y}_t^2,\bar{y}_{t-i}^2\right)}{Var\left(y_t\right)^2},\qquad(6.2)$$

where $\bar{y}_t=y_t-T^{-1}\sum_{t=1}^{T}y_t$. 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 7 autocorrelations $(r=7)$ for 800 observations yielding a value equal to $RS(7)=14,86>\chi^2_{7,0.05}$. As the null hypothesis of no autocorrelation is rejected the test is run on the estimated residuals from the AR(1) model that gives $RS(6)=12,33<\chi^2_{6,0.05}$. 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 dataset, 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 5, in the Appendix, presents the in-sample 95 per cent confidence interval for the AR(1)GARCH(1,1) model. There are fourteen 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 4, the hypothesis that the model selected by the in-sample analysis is equivalent to the model with minimum value of $\sum_{t=T+1}^{T+s}\left(\hat{z}^2_{t|t-1}\right)$ 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

21

is based only on information available at current day. Figures 6 and 7, in the Appendix, present the one step ahead 95 per cent prediction intervals for the models selected by the SBC and AIC, respectively. There are thirteen observations, or 4.33%, outside the prediction interval for the models selected by the SBC, whereas there are fourteen 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 dataset 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 5 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 behavior 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 (1994) and Hamilton and Susmel (1994) and extended by several authors such as Dueker (1997) and Hansen (1994). 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 modeled 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 $\sum_{t=T+1}^{T+s}\left(\hat{z}_{t|t-1}^2\right)$, 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.

R e f e r e n c e s

Akaike, H. (1973). Information Theory and the Extension of the Maximum Likelihood Principle. *Proceeding of the Second International Symposium on Information Theory. Budapest*, 267-281

Andersen, T., T. Bollerslev and S. Lange (1999). Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon. *Journal of Empirical Finance*, 6, 457-477

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

Berndt, E. R., B. H. Hall, R. E. Hall, and J. A. Hausman (1974). Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement*, 3, 653-665

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

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, 31, 307-327

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

Bollerslev, T., R. C. Chou and K. Kroner (1992). ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence. *Journal of Econometrics*, 52, 5-59

Bollerslev, T., R. F. Engle and D. Nelson (1994). ARCH Models. Ch.49 in R. F. Engle and D. L. McFadden eds. *Handbook of Econometrics, IV Elsevier*

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

Box, G.E.P and D. A. Pierce (1970). Distribution of residual autocorrelation in ARIMA time series models. *Journal of the American Statistical Association*, 65, 1509-1526

Box, G.E.P. and G.C. Tiao (1973). Bayesian Inference in Statistical Analysis. *Reading, Mass*

Cai, J. (1994). A Markov Model of Switching-Regime ARCH. *Journal of Business and Economic Statistics*, 12, 309-316

Campbell, J., A. Lo, and A. C. MacKinlay (1997). The Econometrics of Financial Markets, *New Jersey: Princeton University Press*

Degiannakis, S. and E. Xekalaki (1999). Predictability and Model Selection in the Context of ARCH Models. *Athens University of Economics and Business, Department of Statistics*, Technical Report, 69, http://stat-athens.aueb.gr/~exek/papers/Xekalaki-TechnRep69(1999)ft.pdf

Degiannakis, S. and E. Xekalaki (2004). Autoregressive Conditional Heteroscedasticity Models: A Review. *Quality Technology and Quantitative Management*, 1(2), 271-324

Dickey, D.A. and W.A. Fuller (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427–431

Dueker, M.J. (1997). Markov Switching in GARCH Processes and Mean-Reverting Stock Market Volatility. *Journal of Business and Economic Statistics*, 15, 26-34

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation. *Econometrica*, 50-4, 987-1007

Engle, R. F., C. H. Hong, A. Kane and J. Noh (1993). Arbitrage Valuation of Variance Forecasts with Simulated Options, *Advances in Futures and Options Research,* 6, 393-415

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

Gourieroux, C., (1997). ARCH models and Financial Applications, *Springer Series*

Greene, W.H. (1997). Econometric Analysis, 3rd edition, *Prentice Hall*

Hamilton, J., (1994). Time Series Analysis, *New Jersey: Princeton University Press*

Hamilton, J. and R. Susmel (1994). Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics*, 64, 307-333

Hansen, B.E. (1994). Autoregressive Conditional Density Estimation. *International Economic Review*, 35, 705-730

Harvey, A.C. (1981). The Econometric Analysis of Time Series, *Oxford*

Heynen, R., and H. Kat, (1994). Volatility Prediction: A Comparison of the Stochastic Volatility, GARCH(1,1), and EGARCH(1,1) Models. Journal of Derivatives, Winter 94, 50-65

Hol, E., and S. Koopman, (2000). Forecasting the Variability of Stock Index Returns with Stochastic Volatility Models and Implied Volatility. *Tinbergen Institute, Discussion Paper* No. 104,4

Hsieh, D.A. (1989). Modeling Heteroscedasticity in Daily Foreign-Exchange Rates. *Journal of Business and Economic Statistics*, 7, 307-317

Jarque, C. M. and A. K. Bera (1980). Efficient Tests for Normality, Heteroscedasticity and Serial Independence of Regression Residuals. *Economic Letters,* 6, 255-259

Jorion, P. (1988). On Jump Processes in the Foreign Exchange and Stock Markets. *Review of Financial Studies*, 1, 427-445

Kibble, W. F. (1941). A Two Variate Gamma Type Distribution. *Sankhya*, 5, 137-150

Lo, A. and C. MacKinlay (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1, 41-66

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

Marquardt, D.W. (1963). An algorithm for Least Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics,* 11, 431-441

Nelson, D.B. (1991). Conditional Heteroscedasticity 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

Phillips, P.C.B. (1987). Time Series Regression with A Unit Root. *Econometrica,* 55, 277-301

Phillips, P.C.B. and P. Perron (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, 75, 335–346

Richardson, M. and T. Smith (1994). A unified approach to testing for serial correlation in stock returns. *Journal of Business*, 67, 371-399

Scholes, M. and J. Williams (1977). Estimating betas from non-synchronous data. *Journal of Financial Economics*, 5, 309-327

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

Sentana, E. (1995). Quadratic ARCH models. *Review of Economic Studies*, 62, 639-661

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

Xekalaki E., J. Panaretos and S. Psarakis (2003). A Predictive Model Evaluation and Selection Approach - The Correlated Gamma Ratio Distribution. *Stochastic Musings: Perspectives from the Pioneers of the Late 20th Century, (J. Panaretos, ed.),* Lawrence Erlbaum Associates Publishers, 188-202

Zakoian, J. M. (1990). Threshold Heteroskedasticity Models, manuscript, *CREST, INSEE*