

Selecting from amongst non–nested conditional variance models: information criteria and portfolio determination

Article

Accepted Version

Brooks, C. ORCID: <https://orcid.org/0000-0002-2668-1153> and Burke, S. (2002) Selecting from amongst non–nested conditional variance models: information criteria and portfolio determination. The Manchester School, 70 (6). pp. 747-767. ISSN 1467-9957 doi: 10.1111/1467-9957.00323 Available at <https://centaur.reading.ac.uk/24161/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1111/1467-9957.00323>

Publisher: Blackwell

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

This is the authors' accepted manuscript of an article published in *The Manchester School*. The definitive version is available at www3.interscience.wiley.com

Selecting From Amongst Non-nested Conditional Variance Models: Information Criteria and Portfolio Determination*

Chris Brooks and Simon P. Burke⁺

November 2001

Abstract

This paper considers the finite sample properties of model selection by information criteria in conditionally heteroscedastic models. Recent theoretical results show that certain popular criteria are consistent in that they will select asymptotically the true model with probability one. To examine the empirical relevance of this property, Monte Carlo simulations are conducted for a set of non-nested data generating processes (DGPs) with the set of candidate models consisting of all types of model used as DGPs. In addition, not only is the best model considered but also those with similar values of the information criterion, called close competitors, thus forming a portfolio of eligible models. To supplement the simulations, the criteria are applied to a set of economic and financial series. In the simulations, the criteria are largely ineffective at identifying the correct model, either as best or a close competitor, the parsimonious GARCH(1,1) model being preferred for most DGPs. In contrast, asymmetric models are generally selected to represent actual data. This leads to the conjecture that the properties of parameterisations of processes commonly used to model heteroscedastic data are more similar than may be imagined and that more attention needs to be paid to the behaviour of the standardised disturbances of such models, both in simulation exercises and empirical modelling.

JEL Classifications: C22, C51, C52

Keywords: Information criteria, conditional heteroscedasticity, posterior odds, portfolio of models, evidential strength, non-nested models.

* The authors would like to thank two anonymous referees and participants of the 1999 Royal Economics Society conference for useful comments on a previous version of this paper. We also wish to thank Ólan Henry for helpful discussions on the material of this paper. The usual disclaimer applies.

⁺ The authors' affiliations are ISMA Centre, Department of Economics, and Department of Economics respectively, both at the University of Reading, P.O. Box 218, Whiteknights, Reading, Berks. RG6 6AA.

1. Introduction

Since their introduction by Engle (1982), the autoregressive conditionally heteroscedastic (ARCH) class of models has spawned a large number of applications. In contrast to traditional time series models, ARCH models allow evolution of both the conditional mean and the conditional variance.

Consider a process x_t , $t=0,1,2,\dots$, having a conditional mean μ_t depending on past values of x and current and past values of other variables, such that

$$x_t = \mu_t + \varepsilon_t, \quad (1)$$

where the disturbances ε_t are conditionally normally distributed with zero mean and time varying variance σ_t^2 , i.e. $\varepsilon_t \sim N(0, \sigma_t^2)$. When σ_t^2 is allowed to depend on its own past values, past values of the disturbances and current and past values of other explanatory variables (including those which enter the conditional mean), the process is said to be conditionally heteroscedastic and σ_t^2 is called the conditional variance.

A model of this type thus consists of equations for both conditional mean and variance, the specification of each of which is an empirical issue, of importance in, for example, forecasting volatility (see, for example, Akgiray, 1989 or Brooks, 1998), forecast confidence intervals (Baillie and Bollerslev, 1992), and pricing financial options (e.g. Day and Lewis, 1992). Many extensions and variants have been proposed to allow for differences in the smoothness of the conditional variance, how quickly it responds to shocks, whether such responses are asymmetric, or whether it is dependent on exogenous factors (see below and Bollerslev, Chou and Kroner, 1992, for references to many examples). The standard GARCH(1,1) model is, however, still the most pervasive in the literature, and most of the models estimated correspond to the weak GARCH concept of Drost and Nijman (1993). Hentschel (1995) proposes a new model that nests many of the popular GARCH formulations.

This paper focuses largely on the conditional variance equation, imposing throughout the simplest conditional mean consistent with the model concerned. Attempts at capturing different stylised features of the data, together with the problem of restricting the (estimated) conditional variance to be non-negative, has led to a wide range of models being suggested, many of which are non-nested with respect to one another (although see Ding, Granger and Engle, 1993, for an over-arching nesting structure). Table 1 provides a summary of the models considered here and their distinguishing characteristics. Although this list is by no means exhaustive, it includes most of the GARCH-type models that are in common usage by empirical researchers in economics and finance.

A practical matter of some importance is the selection from amongst this set of models that which best characterizes an observed set of data. In the case where one structure is nested within another (see table 1), it is possible to construct tests of the implied restrictions, often by means of the Lagrange multiplier principle (Bollerslev and Wooldridge, 1992, Bollerslev, Engle and Nelson, 1995, Wooldridge, 1990). Such an approach is not available when the models are non-nested. In particular, if the models in the choice set are not sequentially nested (so that the smallest model is a restricted version of the next largest and so on up to the largest model under consideration), pairwise comparisons employed by statistical tests do not generate an exhaustive search path through the models.

The use of an information criterion to discriminate between models consists of a ranking, the favoured model being that having the smallest value of the criterion. A potentially valuable extension of this narrow procedure is provided for ARMA models by Poskitt and Tremayne (1983, 1987). The idea is that models with larger but similar values of the information criterion should not be disregarded, but retained in a portfolio of models. This approach is in the spirit of Jeffrey's (1961) 'grades of evidence' based on the posterior odds of competing models.

A situation which provides a large portfolio is one in which the data are uninformative as to structure if the models within the portfolio necessarily have different

implications for the underlying data generating process (DGP). In such circumstances, it would be careless to assume that the model generating the minimum value of the criterion was manifestly the best. In such situations, it may be possible to improve the accuracy of conditional variance forecasts by combining either the models in a more general specification, or by combining the forecasts themselves. For example, Granger and Ramanathan (1984) show how forecasts can be combined with optimal weights being the parameter estimates of an ordinary least squares regression of the ex post volatility measures on all of the forecasts. Such combined forecasts could be employed for the construction of forecast confidence intervals or for use as inputs to option pricing models.

Conversely, if the portfolio contains only models with similar characteristics (e.g. asymmetries), then it is clear that this characteristic of the data is important. Finally, if the portfolio contains only one member, namely that with the smallest value of the criterion, then it is evident that this model provides a relatively sharp characterization of the data in a manner that others cannot. Any feature unique to such a model is likely to be of substantial relevance.

The remainder of the paper is organized as follows. Section 2 describes the information criteria used and provides more detail about portfolio formation. Section 3 discusses an application of the criteria to a number of actual economic and financial time series. Section 4 describes the structure of a set of Monte Carlo experiments designed to explore the performance of the procedure and the results. Section 5 concludes.

2. Information Criteria and Portfolio Selection

Information Criteria

The information criteria are based on those due to Akaike (1974), Hannan (1980) and Hannan and Quinn (1979), and Schwarz (1978), denoted AIC, HQ, and SIC respectively. They are adjusted to allow for the defining heteroscedasticity of the models as well as the additional number of parameters introduced through the conditional variance equation. The

conditions for consistent model selection presented by Sin and White (1996) are satisfied by HQ and SIC, but not by AIC. Sin and White also discuss concepts relating to the model selected when the DGP is not included in the choice set. Additional criteria for use in small samples, based on those of Hurvich and Tsai (1989) and Hurvich, Shumway and Tsai (1990) may also be defined but are not considered here.

Each of the criteria is of the form

$$IC = \sum_{t=1}^T \log(\tilde{\sigma}_t^2) + \sum_{t=1}^T \frac{\tilde{\varepsilon}_t^2}{\tilde{\sigma}_t^2} + f(T, g), \quad (2)$$

where $f(T, g)$ is the dimensionality penalty term, T being the sample size, g the total number of estimated parameters and $\tilde{\sigma}_t^2$ the maximum likelihood estimator of σ_t^2 , the conditional variance at time t . The three criteria are:

$$AIC = \sum_{t=1}^T \log(\tilde{\sigma}_t^2) + \sum_{t=1}^T \frac{\tilde{\varepsilon}_t^2}{\tilde{\sigma}_t^2} + 2g \quad (3)$$

$$HQ = \sum_{t=1}^T \log(\tilde{\sigma}_t^2) + \sum_{t=1}^T \frac{\tilde{\varepsilon}_t^2}{\tilde{\sigma}_t^2} + 2g \log(\log(T)) \quad (4)$$

$$SIC = \sum_{t=1}^T \log(\tilde{\sigma}_t^2) + \sum_{t=1}^T \frac{\tilde{\varepsilon}_t^2}{\tilde{\sigma}_t^2} + g \log(T) \quad (5)$$

The dimensionality penalty term does not grow quickly enough with T for consistent model selection by AIC, but does for SIC. Strictly speaking, consistent selection by HQ requires the substitution of 2 by a constant, $c > 2$, in the dimensionality term, but (4) represents the conventional formulation of this criterion.¹

¹ For homoscedastic models, Lutkepohl (1991, proposition 4.3) shows that, for $T > 16$, the size of the dimensionality penalty terms is such that the model selected by SIC minimization cannot have more parameters than that selected by minimizing HQ, which in turn cannot be bigger than that obtained using the AIC. It is not clear that this goes through in the heteroscedastic case as the unconditional (constant) disturbance variance has been replaced by a time varying variance.

Posterior Odds

The ratio of the posterior probabilities of two models, the posterior odds ratio, provides a measure of the extent to which one model is preferred to another. Let $p(M_i|X_T)$ denote the posterior probability of model i (M_i), given the data X_T . The posterior odds ratio of two models i and j , $R(i, j)$, is given by

$$R(i, j) = \frac{p(M_i|X_T)}{p(M_j|X_T)}. \quad (6)$$

Model i is preferred to model j if $R(i, j) > 1$.

Poskitt and Tremayne (1983) develop the relationship between the values of information criteria and the posterior odds ratio for a wide class of linear and non-linear models. This is applied to ARMA model selection by Poskitt and Tremayne (1987). The class of models is defined by

$$h_t(\theta) = u_t$$

where $u_t \sim NIID(0, \sigma^2)$, h_t is non-linear function of observations, and θ is the parameter vector.² The conditional variance models considered in this paper clearly fall into this class since, in general, these models are of the form,

$$x_t = \mu_t + \varepsilon_t$$

$$\varepsilon_t \sim NID(0, \sigma_t^2).$$

This can be rewritten

$$\frac{x_t - \mu_t}{\sigma_t} = \frac{\varepsilon_t}{\sigma_t} \sim NIID(0, 1),$$

which satisfies the required conditions with $\sigma = 1$, and $h_t(\theta) = \frac{x_t - \mu_t}{\sigma_t}$. The fixing of the

variance at 1 simplifies the analysis presented in Poskitt and Tremayne (1983) since it is no

² Poskitt and Tremayne (1983) deal with the case where the function is vector valued.

longer necessary to integrate out the variance parameter. Bayes theorem indicates the calculations necessary to obtain the posterior odds ratio, since it states

$$p(M|X_T) = p(X_T|M)p(M)/p(X_T)$$

where $p(M)$ is the prior probability of the model. The term $p(X_T)$ is the same for all models so will cancel in the formation of the posterior odds ratio. The prior probability of the model is implicit in the choice of information criterion, as described by Poskitt and Tremayne (1983), and will vary with the model. The remaining term, $p(X_T|M)$, is the marginal likelihood of the model. This is obtained by integrating out θ from the likelihood

$$L(X_T|\theta, M) \propto \left(\prod_{t=1}^T \sigma_t \right)^{-1} \exp \left\{ -1/2 \sum_{t=1}^T \varepsilon_t^2 / \sigma_t^2 \right\},$$

where $\left(\prod_{t=1}^T \sigma_t \right)^{-1}$ can be regarded as the Jacobian of the transformation to the standard normal case. Following Poskitt and Tremayne (1983) and substituting into $p(X_T|M)p(M)$ gives

$$p(M|X_T) \propto p(M) \left(\prod_{t=1}^T \tilde{\sigma}_t \right)^{-1} \exp \left\{ -1/2 \sum_{t=1}^T \tilde{\varepsilon}_t^2 / \tilde{\sigma}_t^2 \right\} T^{-g/2} (2\pi)^{g/2},$$

where $\tilde{\sigma}_t$ and $\tilde{\varepsilon}_t$ are maximum likelihood estimates. Using subscripts as the model index, the posterior odds ratio can be written

$$\begin{aligned} R(i, j) &= \frac{p(M_i|X_T)}{p(M_j|X_T)} \\ &= \frac{p(M_i)}{p(M_j)} \frac{\prod \tilde{\sigma}_{j,t}}{\prod \tilde{\sigma}_{i,t}} \exp \left\{ -1/2 \sum \left[\tilde{\varepsilon}_{i,t}^2 / \tilde{\sigma}_{i,t}^2 - \tilde{\varepsilon}_{j,t}^2 / \tilde{\sigma}_{j,t}^2 \right] \right\} (2\pi/T)^{(g_i - g_j)/2} \end{aligned}$$

With the appropriate choice of model priors, $p(M_i)$, $p(M_j)$, this can be written in terms of the difference between the values of the information criteria of the two models. For example, with

$$p(M) \propto \{T/2\pi\}^{-g/2} \exp\{-g\},$$

and using equation (3) as the definition of the AIC, the posterior odds ratio may be written

$$R(i, j) = \exp\left[-\frac{1}{2}(AIC_i - AIC_j)\right].$$

In general, but with differing model priors, the posterior odds ratio may be written,

$$R(i, j) = \exp\left[-\frac{1}{2}(IC_i - IC_j)\right],^3 \quad (7)$$

where IC_i and IC_j are values of the information criteria of the same type for models i and j (equations 2-5 above). See Poskitt and Tremayne, 1983, table 1, for the implicit model priors in the case of other information criteria, all of which go through in the GARCH case. Economic theory of accumulated evidence could also be used to tilt the prior odds in favour of particular models.

Equation (7) is intuitively appealing as a means of distinguishing between models, being based on the difference between criterion values. The extent to which $R(i, j)$ exceeds unity is a measure of the extent to which the data support model i rather than model j and may be used to ‘grade the decisiveness of evidence’ (Jeffreys, 1961, appendix B) in favour of model i . Jeffreys (1961) employs powers of $\sqrt{10}$ to define the limits of various classes of decisiveness, such that, if $1 < R(i, j) < \sqrt{10}$, the evidence distinguishing the models is regarded as ‘not worth more than a bare mention’ (Jeffreys, p432).⁴ To use Poskitt and Tremayne’s (1987) term, such models are ‘close competitors’. Intervals delimited by higher integer powers of $\sqrt{10}$, for example $\sqrt{10} \leq R(i, j) < 10$, indicate increasingly decisive support for model i rather than model j , so that the models compete less and less.⁵

³ Equation (2.1) of Poskitt and Tremayne (1987) defines the posterior odds ratio with an additional factor of T in the exponent because their definition of the criteria has an additional divisor of T relative to the definitions of equations (2) to (5) given in this paper.

⁴ The choice of powers of root 10 to define the limits of the intervals is the choice of Jeffreys, p432, and is followed by Poskitt and Tremayne. A researcher is entitled to substitute alternative limits. However, the choice should be distinguished from the choice of prior probabilities.

⁵ Note that Jeffreys (1961) defines the posterior odds ratio as the inverse of $R(i, j)$ and thus works in negative powers of root 10 when defining grades of evidence.

Model Portfolio

An additional role is now given to model i of equation (7): it is the model from the choice set that minimises the information criterion concerned. It is thus always the best model, ensuring $R(i, j) \geq 1$. Let IC^* be the minimised information criterion value and redefine $R(j)$, with a single argument, to be

$$R(j) = \exp\left[-\frac{1}{2}(IC^* - IC_j)\right], \quad (8)$$

for $j=1, 2, \dots, N$. Thus $R(j)$ measures the distance from the best model to model j and can be used to define a portfolio of competing structures: the best model plus those for which $R(j)$ lies below some upper limit, in this case $\sqrt{10}$.

The portfolio of models need not be the same for all criteria as both the best model and the posterior odds may differ. Poskitt and Tremayne (1983) show that the posterior odds ratio of the true model to any other tends to infinity with the sample size for SIC and HQ (if 2 is replaced by $c > 2$), ensuring that the true model will eventually be selected as best and that there will be no close competitors. Sin and White (1996) also show that the information criteria applied here have similar consistency properties. Inconsistency of AIC would be likely to result in over-parameterization, but the impact in the non-nested environment explored here is difficult to predict.

The use of $R(j) < \sqrt{10}$ to define a close competitor implies odds of around 3:1 or better, so that if the best model had a posterior probability of 0.9, a competing specification having a posterior probability of at least 0.35 would be regarded as a close competitor. The Monte Carlo results of Poskitt and Tremayne (1987, table 2) indicate that this definition of a close competitor works satisfactorily for ARMA models, and so it is retained here.⁶ To make assessments of the closeness of any model to the optimal one clearer, $R(j)$ (equation 8) is transformed to

⁶ Some evidence on the sensitivity to the choice of this gradation is provided in tables 5 below.

$$r(j) = \frac{2 \ln(R(j))}{\ln(10)}. \quad (9)$$

Thus a close competitor has $r(j) < 1$, and Jeffreys' gradings in terms of the $r(j)$ are intervals of unit length with integer limits. Root 10 close competitors are thus defined by $(0 \leq r(j) < 1)$ and are denoted by \mathbf{CC}_1 , and any model which falls into the next lowest Jeffreys' class is defined by $(1 \leq r(j) < 2)$ and denoted by \mathbf{CC}_2 .

In these realistic circumstances, there is no guarantee that the choice set includes the true model. It is also possible that a better, though still mis-specified model lies outside the choice set. The selection procedure simply seeks the best approximating model from the candidates offered.⁷

3. Empirical Applications

In order to investigate the discriminatory power and close competitor selection of the criteria, they are applied to 31 economic and financial time series observed at quarterly, monthly, or daily frequency. The model choice set is that of table 1 with all orders set to 1, with the exception of *ARCH*, the candidate model from this family being of order 4. The series under consideration are given in table 2, and are for the UK unless otherwise stated. All the series were obtained from Datastream International, and were transformed to 100 times the first difference of the logs. Before the conditional variance models were estimated, the data were filtered by an estimated autoregressive model of order 10.

Table 3 reports the criteria-minimising models for each series, together with the grade 1 and grade 2 close competitors, as defined above. The main features of the results are:

- i) Model selection is broadly consistent across criteria (the same best model is selected by all three criteria for 20 out of 31 series).
- ii) The GARCH(1,1) model does not dominate (selected 10, 9, and 4 times by SIC, HQ, AIC respectively). When GARCH(1,1) is selected by a particular criterion, it is also selected by a

more parsimonious criterion (one with a larger penalty term), and thus is chosen by all criteria to represent just four of the series investigated.

iii) The most commonly jointly selected model is GJR (7 out of 20), this model also being selected most frequently by each criterion individually.

iv) CC_1 close competitors increase in frequency as the penalty term becomes less severe (2, 4, and 7 for SIC, HQ, AIC respectively), and are rare when the sample size is large (1, 0, 3 for SIC, HQ, AIC respectively, out of 19 series of length 2444). CC_2 close competitors occur with greater frequency than CC_1 's for SIC and HQ, and to a similar extent for each criterion. They too decrease in frequency with sample size, suggesting that the criteria home in on a particular model as the sample size increases.

Two other features of the results are that asymmetric models (EGARCH, GJR, and TGARCH) are favoured as the type of approximation to the DGP⁸, and that the parsimonious criteria (SIC and HQ) do not seem to be discriminating strongly against large models. To illustrate, for M4, EC, FTSE, and DAX, all criteria select the TGARCH model despite the fact that this is one of the larger models available with 5 estimated parameters. The ARCH model features only for one series, RPI, where, oddly, it is chosen by all criteria without any close competitors. Of the 20 cases where all criteria select the same model, 15 are of selection of an asymmetric model.

A typical example of the operation of portfolio selection is provided by the GDP series. In this case, SIC selects GARCH(1,1) with no CC_1 competitors, but with GJR as a CC_2 competitor. HQ, a more profligate criterion, selects GJR outright, with GARCH becoming a CC_1 competitor, and ARCH entering as a CC_2 competitor. Finally, AIC dispenses with GARCH completely, selecting GJR with ARCH as a CC_1 competitor. Similar behaviour can be observed amongst all the shorter series. This evidence suggests that one of the main differences in the operation of the criteria in relatively small samples is the importance given

⁷ The optimal model has $r(j)=0$.

to the parameterization of irregularities in the data. In this exercise such parameterization is through either asymmetries or long ARCH structure.

None of the selected models is being suggested as the best available for the data: the choice set is limited. It is likely that even in the case of the best model there remains further structure to be exploited. The spirit of the exercise is one of initial data exploration and from this it is apparent that almost all these series exhibit features that are better captured by models allowing for asymmetries than by those that do not.

This result is consistent with much recent research in finance which suggests that a negative shock to returns will generate higher volatility than a positive shock of the same magnitude (see Nelson, 1991; Pagan and Schwert, 1990; Engle and Ng, 1993; Henry, 1998). Such asymmetries are also observed in many of the economic series considered here.

4. Monte Carlo Structure and Results

The previous section suggested that the criteria have discriminatory power for selection of an appropriate heteroscedastic model. In order to assess the properties of model selection in a controlled setting, a Monte Carlo study is conducted. There are six candidate models, given in table 1 above (GARCH, ARCH, GARCH-M, EGARCH, GJR, TGARCH). Each of these models is also used as a data generating process. The values of the three information criteria, together with those of $r(j)$, are calculated for each model and for each DGP. The values of the parameters in the DGPs are taken from various models estimated in the literature using real economic or financial data as follows.

ARCH(4): Engle (1982, table 3).

$$r_t = \varepsilon_t \quad , \quad \varepsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = 14 \times 10^{-6} + 0.382\varepsilon_{t-1}^2 + 0.286\varepsilon_{t-2}^2 + 0.191\varepsilon_{t-3}^2 + 0.096\varepsilon_{t-4}^2$$

GARCH(1,1): Bollerslev (1986, equation (3.1)):

⁸ Asymmetric models represent only half the models in the choice set, but are selected approximately two thirds of the time by each criterion individually.

$$r_t = \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = 0.007 + 0.135\varepsilon_{t-1}^2 + 0.829\sigma_{t-1}^2$$

GARCH(1,1)-M: Sentana (1995, table 3, column 2)

$$r_t = 0.243\sigma_t^2 + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = 0.08926 + 0.20752\varepsilon_{t-1}^2 + 0.67928\sigma_{t-1}^2$$

EGARCH: Hsieh (1993, table 5, column 1)

$$r_t = \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2)$$

$$\log(\sigma_t^2) = -0.688 + 0.136\left(\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}}\right) + (2/\pi)^{0.5} + 0.929\log(\sigma_{t-1}^2) - 0.111\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}}$$

GJR: Glosten *et al.* (1993, table 3, model 2, excluding the GARCH-in-mean term)

$$r_t = \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = 0.00074 + 0.257\varepsilon_{t-1}^2 + 0.483\sigma_{t-1}^2 - 0.340\varepsilon_{t-1}^2 I_{t-1}$$

TGARCH: Zakoian (1994, table 4, column 4)

$$r_t = \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t = 0.049 + 0.111\varepsilon_{t-1}^+ - 0.192\varepsilon_{t-1}^- + 0.833\sigma_{t-1}$$

The simulations are for sample sizes 250, 500, 1000 and 2000, and consist of 5000 replications. The results are given in tables 4 and 5. Table 4 presents the percentage of times each model is selected by each of the criteria, and the number of times each model is a close competitor for each DGP. There are four entries per cell, one for each sample size. Even numbered columns (2, 4, 6, 8, 10, 12, 14, labelled ‘best’) give the proportion of times that each of the six models is selected as optimal by the criteria, and odd numbered columns (3, 5, 7, 9, 11, 13, 15, labelled CC) give the proportion of times a model is selected as a close competitor. Thus the numbers on the leading diagonal represent the percentage of times the correct model is either chosen outright or as a close competitor. Defining the portfolio of candidate models to be the set of those that are optimal or close competitors, the proportion of times a model is in the portfolio is given by summing the ‘best’ and ‘CC’ numbers.

The median and ninetieth percentile of the distributions of the $r(j)$ are computed for the SIC for each DGP and presented in table 5. A model that tends to be a close competitor for a given DGP will have the bulk of the distribution to the left of 1. The more skewed above 1 is the distribution, the less satisfactory an approximation that model provides, compared with that which is best.

GARCH Model Selection

Consider first the frequency with which the GARCH(1,1) model is selected across the range of DGPs. In contrast to the results for real data, the GARCH model is frequently selected as the best model for most of the data generating processes. At smaller sample sizes, GARCH is the dominant SIC and HQ selection, almost irrespective of the DGP. As sample size increases, the rate of correct GARCH selection increases, but it is selected less frequently as the best model for other DGPs (table 4a, b). Most notably, it is selected very infrequently by SIC when GJR is the DGP and the sample size 2000 (2.61%). When the DGP is GARCH-M or TGARCH, erroneous selection by SIC of the GARCH model occurs at around 65% for $T=2000$, considerably reduced from the values at smaller sample sizes and indicating an asymptotic increase in correct model selection in these cases. Erroneous GARCH selection rates reduce much more slowly when the DGP is ARCH or EGARCH (still around 92% for SIC, but better for HQ, table 4b).

Table 4c presents the results for AIC. They are similar, but selection rates for GARCH are reduced in all cases (including when it is the DGP). In particular, incorrect GARCH selection occurs for $T=2000$ at a rate much above 50% only when the DGP is ARCH (62.02%). All incorrect GARCH selection rates decrease with sample size, and correct selection improves. These rates of change are still slow in some cases however, indicating that even larger sample sizes are required for the reduction of incorrect GARCH selection to reasonable levels.

Selection of close competitors using the SIC (table 4a) is uncommon, but more common with HQ, and most frequent when AIC is used. GARCH is chosen as a close

competitor most frequently when the DGP is TGARCH (around 10% at all sample sizes). For most DGPs, the frequency with which GARCH is chosen as a close competitor decreases with sample size, including when GARCH is the DGP. When GJR is the DGP, GARCH becomes a close competitor on very few occasions (0.88% for SIC, 0.25% for AIC, $T=2000$).

Portfolio Selection

A model may not be chosen as best, but may still appear in the portfolio as a close competitor. However, the results indicate that even at $T=2000$, portfolio selection of the correct model is infrequent. At this sample size, the GARCH, ARCH, GARCH-M, EGARCH, GJR and TGARCH models are correctly selected by SIC as members of the portfolio with the following frequencies: 97.99, 0.04, 34.06, 0.49, 4.65, 0.40. The corresponding figures for AIC are: 91.11, 0.22, 33.18, 2.07, 2.52, 35.77 (the higher TGARCH selection rate being due to its selection as a close competitor). Clearly, GARCH is correctly identified very successfully (improving with sample size), ARCH, EGARCH and GJR rarely, and GARCH-M relatively well in the context of these experiments. TGARCH is a case where the use of the profligate criterion brings it into consideration as a close competitor, but its selection as the best model remains low even for AIC (4.62%).

The real data results indicated that the GJR model was a favoured representation. In the simulations, the selection of GJR as a portfolio member depends on which criterion is used. The more profligate criteria frequently admit it as a close competitor. It is a member of the AIC portfolio when $T=2000$ at the rates 92.22, 59.22, 65.18, 54.63, 2.52, 43.19 for the DGPs GARCH, ARCH, GARCH-M, EGARCH, GJR, and TGARCH respectively. Of the other models in the choice set, ARCH is virtually never selected for any DGP by any criterion (consistent with real data results). GARCH-M is selected by SIC only when it is the DGP, but is a strong close competitor for many DGPs when AIC is used. While EGARCH is not selected when it is the DGP, it is selected by SIC when the DGP is of an alternative asymmetric form, that is either GJR or TGARCH, both selection frequencies increasing with

sample size. TGARCH is the dominant selection under the GJR DGP for all criteria, once a reasonable sample size is obtained.

These results, and those for the real series, encourage the consideration that appropriate portfolio selection of asymmetric models may characterise the results. This is true to a limited extent, and best demonstrated when the criterion used is profligate, and the sample size large. Using the AIC at $T=2000$, asymmetric models are selected as best for the asymmetric DGPs, EGARCH, GJR, TGARCH, at the rates 35.28, 99.37, 70.46 respectively. The GJR DGP generally leads to selection of an TGARCH (portfolio selection rate 82.87%), or EGARCH (30.2%) model, and TGARCH is generally interpreted as EGARCH (92%). However, the EGARCH DGP still seems to be confused with a GARCH process (73.52%), GJR and TGARCH models being portfolio members less often (54.66% and 39.83%).

Portfolio Selection and Sample Statistics of $r(j)$

Table 5a provides an alternative measure of how competitive a model is, showing the median of the distributions of the $r(j)$. This is the upper limit on the value of $r(j)$ necessary for 50% of replications to result in model j being a member of the portfolio. The higher the number in the table, the less likely it is that the fitted model concerned will be in the portfolio.⁹ Table 5b presents similar information, but for 90% of replications to generate portfolio membership.

In the case of most DGPs, there is no evidence that model selection is converging by $T=2000$, either to the correct or an incorrect solution. For example, for a fitted GJR model to be in the portfolio on 50% of occasions when it also the DGP, a close competitor definition of $r(j)<1.77$ is required at $T=250$, rising to $r(j)<7.34$ at $T=2000$ (table 5a). In the latter case, this would result in TGARCH also being in the portfolio on over 90% of occasions (the 90th percentile for fitted TGARCH when the DGP is GJR being only 4.37), and EGARCH on over 50% (median 1.06).¹⁰ Exceptions are GARCH (correct selection) and GJR (selection of

⁹ A value of zero indicates that the model concerned is selected as best on at least 50% of occasions.

¹⁰ Jeffrey's (1961) argues that $r(j)>4$ indicates decisive evidence *against* the competing model.

TGARCH). A somewhat pathological case is GARCH-M, where rates of correct selection increase with sample size, but both median and 90th percentile increase.¹¹

Reading the main diagonal blocks of table 5a, GARCH is best correctly identified at all sample sizes, the second best model always being asymmetric, but varying with sample size (GJR at smaller sample sizes, TGARCH at the largest).

5. Conclusions

Conditional variance models have been applied widely in finance and to macroeconomic data (for example Engle, 1982 on inflation; Price, 1995 on investment; Cuthbertson and Gasparro, 1993, on output). In all cases the selection of the model is an important issue, that can, in theory, be addressed by the use of information criteria. While the SIC and HQ have been shown by Sin and White (1996) to be consistent in model selection, this empirical investigation has produced ambiguous evidence. Some of the models are likely to be much more difficult to estimate than the standard GARCH(1,1) formulation, which may go some way to explaining the difficulty in selecting the true data generating process in the simulations study. Moreover, given that the coefficient values employed in the simulations imply a process with a conditional variance close to the non-stationary boundary, larger sample sizes are likely to be required to achieve consistency of the selection procedure than would otherwise be the case.

In applications to real data series, where all models are necessarily mis-specified, the criteria discriminate quite widely, asymmetric models being favoured and the GJR being dominant. When the sample size is relatively small, close competitors exist, but are rare when it is large. Such findings are potentially useful.

¹¹ The sample size is not large enough for the correct selection of GARCH-M to dominate. Although, as sample size increases, the GARCH-M has the lowest criterion value more often, when it is not chosen, the difference in criterion values between it and the best (overwhelmingly likely to be GARCH in this case) becomes greater. Evidence that discrimination is also becoming sharper when GARCH-M is selected is provided by table 5b which shows that the GARCH 90th percentile is increasing (although the median has not shifted).

In contrast, the simulated data results in the frequent selection of the parsimonious GARCH(1,1) model, the exception being when the DGP has certain asymmetric features. It is quite likely however, that the preferred parameterization of the asymmetry will not be that which generated the data.

There are two related implications of this. Firstly, the type of processes often supposed to satisfactorily approximate the DGPs of actual data do not provide sufficient information for the criteria to discriminate between them. Secondly, model selection in practice must be driven by features of the data that do not correspond very closely to the properties of the models considered, because in applications to real series, the criteria make quite tight distinctions, often discriminating against GARCH.

These points reinforce the message that the criteria do not identify the true model in either situation, but select the best approximating structure from those available. This interpretation is consistent with the asymptotic results. It also suggests that conclusions about economic or financial structure relying heavily on the details of the selected model may be mis-leading. This enhances the role of close competitors. Their existence indicates a lack of resolution in model selection, but may offer a guide as to which are important and which are spurious points of distinction.

References

- Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 716-723.
- Akgiray, V. (1989) 'Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts', *Journal of Business* 62(1), 55-80.
- Baillie, R.T. and Bollerslev, T. (1992) Prediction in dynamic models with time dependent conditional variances. *Journal of Econometrics*, 52, 91-113.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.
- Bollerslev, T. and Wooldridge, J.M. (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11, 143-172.
- Bollerslev, T., Chou, R.Y., and Kroner, K.F. (1992) ARCH modelling in finance: review of the theory and empirical evidence. *Journal of Econometrics*, 52, 5-59.
- Bollerslev, T., Engle, R.F., and Nelson, D.B. (1995) ARCH models. In Engle, R.F. and McFadden, D.L. (eds.), *Handbook of Econometrics*, vol. IV. Amsterdam: Elsevier.
- Brooks, C. (1998) Forecasting Stock Return Volatility: Does Volume Help? *Journal of Forecasting* 17, 59-80.
- Cuthbertson, K. and Gasparro, D. (1993) The Determinants of Manufacturing Investments in the UK *Economic Journal* 103, 1479-1492.
- Day, T.E. and Lewis, C.M. (1992) 'Stock Market Volatility and the Information Content of Stock Index Options', *Journal of Econometrics* 52, 267-287.
- Ding, Z., Granger, C.W.J. and Engle, R.F. (1993) A Long Memory Property of Stock Market Returns and a New Model *Journal of Empirical Finance* 1, 83-106
- Drost, F.C. and Nijman, T.E. (1993) Temporal Aggregation of GARCH Processes *Econometrica* 61(4), 909-927
- Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, 50, 987-1007.
- Engle, R.F., Lilien, D.M., and Robins, R.P. (1987) Estimating time varying risk premia in the term structure: the ARCH model. *Econometrica*, 55, 391-407.
- Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993) On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks *The Journal of Finance* 48(5), 1779-1801
- Hannan, E.J. (1980) The estimation of the order of an ARMA process. *Annals of Statistics*, 8, 1071-1081.

- Hannan, E.J. and Quinn, B.G. (1979) The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B*, 41, 190-195.
- Hendry, D.F. and Doornik, J. (1996) Empirical Econometric Modelling Using PcGive for Windows. London: International Thompson Business Press.
- Henry, O.T. (1998) Modelling the asymmetry of stock market volatility *Applied Financial Economics* Vol. 8, pp. 145-153.
- Hentschel, L. (1995) All in the Family: Nesting Symmetric and Asymmetric GARCH Models *Journal of Financial Economics* 39, 71-104
- Hsieh, D.A. (1993) Implications of Nonlinear Dynamics for Financial Risk Management *Journal of Financial and Quantitative Analysis* 28(1), 41-46
- Hurvich, C.M. and Tsai, C.-L. (1989) Regression and time series model selection in small samples. *Biometrika*, 76, 297-307.
- Hurvich, C.M., Shumway, R., and Tsai, C.-L. (1990) Improved estimators of Kullback-Leibler information for autoregressive model selection in small samples. *Biometrika*, 77, 709-719.
- Jeffreys, H. (1961) Theory of Probability. Reprinted (1998) , Oxford Classical Texts in the Physical Sciences. Oxford: OUP.
- Lutkepohl, H. (1991) Introduction to Multiple Time Series. Heidelberg: Springer-Verlag.
- Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach *Econometrica* 59(2), 347-370
- Poskitt, D.S. and Tremayne, A.R. (1987) Determining a portfolio of linear time series models. *Biometrika*, 74, 125-137.
- Poskitt, D.S. and Tremayne, A.R. (1983) On the posterior odds of time series models. *Biometrika*, 70, 157-162.
- Price, S.G. (1995) Aggregate Uncertainty, Capacity Utilisation, and Manufacturing Investment *Applied Economics* 27, 147-154
- 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.
- Sin, C.-Y., and White, H. (1996) Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics*, 71, 207-225.
- Zakoian, J.-M. (1994) Threshold Heteroskedastic Models *Journal of Economic Dynamics and Control* 18, 931-955

Table 1 Format: Conditional Variance Models

Model	Main Reference	Conditional Mean	Conditional Variance ¹²	Nested Models	Main Characteristic
1 ARCH(m)	Engle (1982)	$y_t = \mu + \varepsilon_t$	$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2$	None	Simplest model
2 GARCH(m,n)	Bollerslev (1986)	$y_t = \mu + \varepsilon_t$	$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^n \gamma_j \sigma_{t-j}^2$	1	Most commonly applied, standard model
3 GARCH-M	Engle, Lilien & Robins (1987)	$y_t = \mu + \rho \sigma_t^2 + \varepsilon_t$	$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^n \gamma_j \sigma_{t-j}^2$	1, 2	Feedback term into mean eqn.
4 EGARCH	Nelson (1991)	$y_t = \mu + \varepsilon_t$	$\log(\sigma_t^2) = \alpha_0 + \sum_{j=1}^m \alpha_j \left(\left \frac{\varepsilon_{t-j}}{\sqrt{\sigma_{t-j}^2}} \right - (2/\pi) \right) + \sum_{i=1}^n \gamma_i \log(\sigma_{t-i}^2) + \sum_{k=1}^p \delta_k \frac{\varepsilon_{t-k}}{\sqrt{\sigma_{t-k}^2}}$	None	Logarithmic form for cond. var. + asymmetry term
5 GJR ¹³	Glosten, Jaganathan & Runkle (1993)	$y_t = \mu + \varepsilon_t$	$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^n \gamma_j \sigma_{t-j}^2 + \beta \varepsilon_{t-1}^2 I_{t-1}$	1,2	Asymmetry term
6 TGARCH ¹⁴	Zakoian (1994)	$y_t = \mu + \varepsilon_t$	$\sigma_t = \alpha_0 + \sum_{i=1}^m \lambda_{1i} \varepsilon_{t-i}^- + \lambda_{2i} \varepsilon_{t-i}^+ + \sum_{j=1}^n \gamma_j \sigma_{t-j}$	1,2	Asymmetry term

¹² The versions of the models used in this paper all have $n=m=p=1$.

¹³ Where $I_t = 1$ if $\varepsilon_t > 0$, 0 otherwise.

¹⁴ $\varepsilon_t^- = \varepsilon_t$ if $\varepsilon_t < 0$, 0 otherwise; $\varepsilon_t^+ = \varepsilon_t$ if $\varepsilon_t \geq 0$, 0 otherwise.

Table 2: Actual Data Series to Which the Criteria are Applied⁺

Series	Acronym	Series	Acronym
Gross Domestic Product	GDP	French CAC-40 Stock Index	CAC
Retail	RPI	Toronto Composite Stock Index	TOR
Prices Index			
Narrow money	M0	Australian All-Ordinaries Stock Index	AAO
Broad money	M4	Japanese Yen	JPY
Halifax House Prices Index	HHP	US dollar	USD
Industrial Production	IP	German mark	DEM
Numbers of wholly unemployed	U	Italian lira	ITL
Average Earnings	AE	Swiss Franc	CHF
Housing Starts	HS	Austrian shilling	ATS
Energy Consumption	EC	Danish krone	DKK
Daily mid-day FTSE-100 Index	FTSE	Nickel price	NICK
German DAX Share Index	DAX	Tin price	TIN
Dow Jones Industrial Average	DJIA	Zinc price	ZINC
Korean Composite Stock Index	KC	Cocoa price	COCO
Japanese Nikkei Dow Stock Index	NIK	Gold price	GOLD
Hang Seng Stock Index	HANG		

⁺Series are for the UK unless otherwise stated.

**Table 3: Selected Models and Close Competitors for
Economic & Financial Series^{\$}**

Series	No. of Observations (data frequency)	Sample Period	SIC [#]			HQ [#]			AIC [#]		
			Best	CC ₁	CC ₂	Best	CC ₁	CC ₂	Best	CC ₁	CC ₂
GDP	174 (Quarterly)	Q1/55-Q2/98	GAR	-	GJR	GJR	GAR	ARC	GJR	ARC	-
RPI	509 (Monthly)	1/65-5/98	ARC	-	-	ARC	-	-	ARC	-	-
M0	348 (Monthly)	6/69-5/98	GJR	-	-	GJR	-	-	GJR	-	-
M4	192 (Monthly)	6/82-5/98	TGA	-	GAR	TGA	-	-	TGA	-	-
HHP	186 (Monthly)	1/83-6/98	GAR	-	GJR	GAR	-	GJR	GAR	GJR	GAM
IP	331 (Monthly)	1/71-7/98	GAR	-	GJR	GAR	GJR	EGA	GJR	GAR	EGA
U	331 (Monthly)	1/71-7/98	EGA	-	-	EGA	-	-	EGA	-	-
AE	331 (Monthly)	1/71-7/98	GJR	TGA	GAR	TGA	GJR	GAR	TGA	-	GJR
HS	331 (Monthly)	1/71-7/98	GAR	-	EGA	GAR	EGA	GAM	EGA	GAR	GAM
EC	331 (Monthly)	1/71-7/98	TGA	-	-	TGA	-	-	TGA	-	-
FTSE	1440 (Daily)	4/1/93-10/7/98	TGA	-	-	TGA	-	-	TGA	-	-
DAX	1440 (Daily)	4/1/93-10/7/98	TGA	-	GAR	TGA	-	-	TGA	-	-
DJIA	2444 (Daily)	1/6/89-13/10/98	EGA	-	-	EGA	-	-	EGA	-	-
KC	2444 (Daily)	1/6/89-13/10/98	GJR	-	-	GJR	-	-	GJR	TGA	-
NIK	2444 (Daily)	1/6/89-13/10/98	GJR	-	-	TGA	-	-	TGA	-	-
HANG	2444 (Daily)	1/6/89-13/10/98	GJR	-	-	GJR	-	-	GJR	TGA	-
CAC	2444 (Daily)	1/6/89-13/10/98	GJR	-	-	GJR	-	-	GJR	-	-
TOR	2444 (Daily)	1/6/89-13/10/98	EGA	-	-	EGA	-	-	EGA	-	-
AAO	2444 (Daily)	1/6/89-13/10/98	GJR	-	-	GJR	-	-	GJR	-	-
JPY	2444 (Daily)	1/6/89-13/10/98	GAR	-	-	GAR	-	-	GAM	-	-
USD	2444 (Daily)	1/6/89-13/10/98	GAR	-	-	GAR	-	-	GAR	-	-
DEM	2444 (Daily)	1/6/89-13/10/98	GJR	-	-	GJR	-	-	GJR	-	-
ITL	2444 (Daily)	1/6/89-13/10/98	GAR	-	-	GAR	-	-	GJR	-	TGA
CHF	2444 (Daily)	1/6/89-13/10/98	GAR	-	-	GAR	-	-	GAR	-	-
ATS	2444 (Daily)	1/6/89-13/10/98	GJR	-	-	GJR	-	-	EGA	-	-
DKK	2444 (Daily)	1/6/89-13/10/98	GAR	-	-	GAR	-	-	GJR	-	-
NICK	2444 (Daily)	1/6/89-13/10/98	GAR	GAM	-	GAR	-	TGA	GAR	GJR	GAM
TIN	2444 (Daily)	1/6/89-13/10/98	GJR	-	-	GAM	-	GAR	TGA	-	-
ZINC	2444 (Daily)	1/6/89-13/10/98	EGA	-	-	GJR	-	-	GAM	-	-
COCO	2444 (Daily)	1/6/89-13/10/98	GJR	-	-	GJR	-	-	GJR	-	-
GOLD	2444 (Daily)	1/6/89-13/10/98	EGA	-	-	EGA	-	-	EGA	-	-

^{\$} Best: $r(j)=0$; CC₁: $0 \leq r(j)<1$; CC₂: $1 \leq r(j)<2$.

[#]Model acronyms abbreviated to first three letters, aside from GARCH-M, which is shorted to GAM

Table 4a: Simulation Results: SIC Relative Frequencies(%) of Selection as Best Model and Root 10 Close Competitor

DGP	GARCH		ARCH		GARCH-M		EGARCH		GJR		TGARCH	
Fitted Model	Best	CC	Best	CC	Best	CC	Best	CC	Best	CC	Best	CC
GARCH	84.52	5.72	94.26	3.28	85.46	7.11	95.91	2.58	65.43	15.73	77.01	10.60
	90.81	2.76	94.66	1.78	84.77	11.5	96.89	1.74	39.71	16.87	78.98	9.78
	95.29	1.48	93.55	1.82	77.66	4.82	96.42	1.94	13.44	7.74	76.56	9.96
	96.91	1.08	92.21	1.78	66.07	1.56	93.48	4.02	2.61	0.88	66.44	10.76
ARCH	0.00	0.06	0.00	0.02	0.02	0.02	0.00	0.03	0.02	0.02	0.00	0.00
	0.02	0.08	0.00	0.00	0.06	0.08	0.00	0.04	0.0	0.02	0.00	0.02
	0.02	0.00	0.00	0.00	0.00	0.04	0.02	0.02	0.02	0.04	0.00	0.00
	0.00	0.00	0.04	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.02	0.10
GARCH-M	4.85	6.91	2.80	5.73	6.95	7.05	0.19	0.36	0.02	0.45	0.86	1.56
	3.25	4.80	2.66	4.28	11.22	6.00	0.02	0.86	0.04	1.77	0.28	0.46
	2.01	2.81	3.88	4.16	20.17	3.73	0.28	1.50	0.63	2.10	0.06	0.78
	1.14	2.03	5.06	3.90	33.56	0.50	1.94	6.14	0.33	0.50	0.04	2.08
EGARCH	6.83	9.46	0.00	0.00	3.84	6.63	0.18	0.74	17.43	9.53	19.71	19.79
	2.90	4.44	0.00	0.02	1.75	4.05	0.27	0.37	29.26	39.79	18.60	17.45
	0.48	1.20	0.00	0.00	0.51	1.42	0.25	0.27	40.48	7.67	21.16	15.94
	0.08	0.24	0.00	0.00	0.37	0.43	0.29	0.20	39.84	7.59	31.22	11.86
GJR	4.22	8.20	2.96	6.42	3.80	7.63	3.27	6.23	10.72	17.06	2.24	7.39
	3.49	5.78	2.72	4.38	2.36	4.64	2.41	5.31	13.56	13.71	1.95	6.37
	2.51	3.57	2.82	5.02	1.70	2.86	2.50	4.29	9.18	8.09	2.10	6.18
	2.12	1.88	2.74	4.58	1.08	1.70	3.26	6.14	3.09	1.56	2.24	6.02
TGARCH	0.18	0.30	0.00	0.00	0.12	0.06	0.44	1.35	6.39	9.84	0.37	0.71
	0.08	0.10	0.00	0.00	0.04	0.00	0.52	1.29	17.38	12.50	0.24	0.65
	0.00	0.00	0.00	0.00	0.02	0.00	0.60	1.10	36.26	12.80	0.12	0.36
	0.00	0.00	0.00	0.00	0.00	0.00	1.06	1.12	54.14	9.71	0.04	0.36

Note: Four entries per cell are for sample sizes T = 250, 500, 1000, 2000 respectively.

Table 4b: Simulation Results: HQ Relative Frequencies(%) of Selection as Best Model and Root 10 Close Competitor

DGP	GARCH		ARCH		GARCH-M		EGARCH		GJR		TGARCH	
Fitted Model	Best	CC	Best	CC	Best	CC	Best	CC	Best	CC	Best	CC
GARCH	73.32	11.87	85.32	9.38	72.91	13.46	87.66	7.68	40.00	20.83	58.22	19.53
	80.48	9.78	86.83	7.42	72.20	5.93	87.29	8.01	16.00	14.76	58.26	19.04
	87.50	6.82	84.05	8.06	67.82	8.02	84.91	8.74	4.74	3.44	54.90	16.48
	90.13	5.68	82.67	7.88	61.17	3.86	74.20	13.82	0.81	0.63	47.82	13.42
ARCH	0.08	0.18	0.02	0.00	0.06	0.06	0.08	0.30	0.10	0.44	0.02	0.34
	0.10	0.18	0.00	0.00	0.26	0.20	0.18	0.30	0.08	0.50	0.14	0.76
	0.06	0.04	0.06	0.00	0.16	0.14	0.20	0.18	0.04	0.15	0.22	0.66
	0.00	0.00	0.08	0.10	0.04	0.04	0.10	0.22	0.02	0.02	0.30	0.88
GARCH-M	8.25	18.08	6.89	20.69	12.35	11.43	0.17	5.63	0.08	4.74	1.78	4.12
	6.99	14.84	6.70	18.94	17.11	6.85	0.46	7.82	0.42	3.84	0.62	2.94
	4.92	13.41	8.72	18.12	24.68	2.01	2.18	11.01	0.65	1.53	0.06	4.56
	3.69	11.90	10.52	18.08	33.18	0.00	8.98	13.46	0.19	0.23	0.16	8.44
EGARCH	10.92	20.62	0.00	0.09	7.22	18.69	0.71	1.98	21.63	11.44	34.04	34.02
	5.31	11.10	0.00	0.02	4.58	10.01	0.60	1.15	31.03	9.93	35.19	30.44
	1.56	4.24	0.00	0.00	2.04	4.06	0.48	1.00	32.24	9.79	38.90	64.78
	0.42	0.76	0.00	0.00	0.91	1.24	0.66	0.73	27.81	8.44	46.28	24.18
GJR	7.48	22.91	7.78	22.71	7.47	22.71	7.71	22.01	16.00	24.99	4.75	20.84
	7.47	19.21	6.48	21.14	6.08	17.56	7.45	20.37	14.02	14.95	4.44	18.60
	6.24	16.43	7.20	18.82	5.40	13.30	7.81	20.46	6.52	6.64	4.84	17.95
	6.01	15.05	6.72	17.10	4.82	10.30	11.35	20.96	2.27	0.86	4.58	18.76
TGARCH	0.63	1.02	0.02	0.08	0.24	0.58	3.67	7.50	22.30	19.93	1.51	4.99
	0.18	0.50	0.00	0.06	0.04	0.04	4.21	7.04	38.45	18.83	1.45	5.04
	0.02	0.04	0.02	0.28	0.02	0.00	4.45	5.64	55.81	13.15	1.08	4.31
	0.00	0.02	0.06	0.52	0.00	0.00	4.77	6.01	68.91	8.02	0.86	4.22

Note: Four entries per cell are for sample sizes T = 250, 500, 1000, 2000 respectively.

Table 4c: Simulation Results: AIC Relative Frequencies(%) of Selection as Best Model and Root 10 Close Competitor

DGP	GARCH		ARCH		GARCH-M		EGARCH		GJR		TGARCH	
Fitted Model	Best	CC	Best	CC	Best	CC	Best	CC	Best	CC	Best	CC
GARCH	56.19	22.64	69.65	20.66	56.89	22.17	71.90	17.44	21.13	20.62	39.26	26.97
	62.10	22.25	68.95	21.17	58.11	18.04	68.88	18.96	6.70	8.97	38.62	25.42
	69.34	20.08	64.82	21.63	56.42	13.18	63.29	22.11	2.13	2.19	36.68	21.46
	71.32	19.79	62.02	22.01	52.29	9.22	51.20	22.32	0.42	0.25	27.80	21.22
ARCH	0.20	0.52	0.02	0.00	0.14	0.22	0.61	3.23	0.44	4.32	0.32	3.12
	0.38	0.36	0.00	0.02	0.50	0.60	0.78	2.08	0.40	1.55	0.90	4.00
	0.12	0.02	0.06	0.02	0.46	0.52	0.46	1.14	0.17	0.38	1.14	3.66
	0.00	0.00	0.10	0.12	0.12	0.18	0.40	0.60	0.04	0.08	1.30	3.30
GARCH-M	13.85	55.61	14.22	70.30	16.84	17.63	0.75	52.49	0.34	15.17	2.66	35.00
	13.60	61.56	15.06	65.56	20.83	6.99	1.88	44.09	0.66	5.27	0.98	31.30
	12.29	68.88	18.08	57.94	26.03	1.26	6.12	36.38	0.34	1.75	0.20	26.50
	11.38	70.43	21.74	50.52	33.18	0.00	13.29	31.53	0.17	0.27	0.44	23.08
EGARCH	15.72	32.78	0.00	0.17	12.03	29.11	1.37	2.83	21.83	11.93	46.00	44.57
	9.11	18.61	0.00	0.05	8.01	17.88	1.08	2.04	26.95	11.46	48.00	42.71
	3.42	6.84	0.00	0.00	3.90	7.20	0.91	1.27	25.16	10.41	50.00	41.28
	0.90	1.74	0.00	0.00	1.67	2.03	1.00	1.07	21.04	7.68	57.10	35.38
GJR	13.41	63.42	16.09	71.85	13.76	66.10	13.48	61.81	17.07	29.92	7.75	48.84
	14.73	69.84	15.92	66.57	12.79	64.35	14.85	54.99	11.17	14.91	7.35	45.94
	15.03	75.22	16.66	56.16	13.36	58.90	16.88	45.01	4.78	4.94	7.68	43.42
	16.61	75.61	15.08	44.14	12.88	52.30	20.30	34.33	1.97	0.55	8.74	34.45
TGARCH	1.21	19.06	0.06	19.24	0.68	17.39	11.90	25.60	39.18	26.98	4.67	22.68
	0.56	18.48	0.10	18.30	0.10	15.96	12.90	24.71	54.11	19.62	4.31	27.30
	0.10	17.99	0.42	18.01	0.02	15.56	12.54	26.29	67.42	11.06	4.31	28.93
	0.04	18.08	1.10	17.92	0.00	13.84	13.98	25.85	76.36	6.23	4.62	31.15

Note: Four entries per cell are for sample sizes T = 250, 500, 1000, 2000 respectively.

Table 5a: Simulation Results Median of Distribution of $r(j)$ for SIC.

DGP	GARCH	ARCH	GARCH-M	EGARCH	GJR	TGARCH
Fitted Model						
GARCH	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0.48 3.56 10.68	0 0 0 0
ARCH	13.38 19.00 27.26 42.09	45.95 51.45 56.22 56.07	14.16 16.06 18.58 23.88	5.93 8.06 10.02 11.94	4.73 6.23 9.93 18.71	5.31 6.12 7.35 9.34
GARCH-M	2.30 2.57 2.87 3.16	2.23 2.49 2.74 3.01	3.96 8.64 18.06 34.87	2.42 2.76 3.12 3.36	2.62 3.81 6.88 14.36	2.43 2.77 3.15 3.62
EGARCH	2.38 3.49 5.53 9.23	* * * *	2.60 3.69 5.75 10.36	* * * *	53.13 2.46 0.93 1.06	1.37 1.55 1.73 1.49
GJR	2.21 2.48 2.79 3.08	2.18 2.48 2.77 3.17	2.19 2.52 2.88 3.25	2.22 2.49 2.80 2.98	1.77 2.03 3.44 7.34	2.25 2.54 2.79 3.03
TGARCH	4.55 5.17 5.80 6.39	4.57 5.15 5.72 6.37	4.58 5.22 5.89 6.55	4.28 4.89 5.41 5.85	2.67 1.99 0.89 0	4.15 4.56 5.10 5.36

* Indicates value in excess of 100.

Note: Four entries per cell are for sample sizes $T = 250, 500, 1000, 2000$ respectively.

Table 5b: Simulation Results 90th Percentile of Distribution of $r(j)$ for SIC.

DGP	GARCH	ARCH	GARCH-M	EGARCH	GJR	TGARCH
Fitted Model						
GARCH	0.95 0 0 0	0 0 0 0	0.52 0.83 2.95 8.78	0 0 0 0	1.69 3.90 8.11 17.04	1.30 1.18 1.51 2.61
ARCH	29.69 39.04 50.65 68.64	* * * *	28.98 32.80 37.05 44.51	8.20 10.68 13.40 16.07	6.92 9.07 14.89 26.91	7.67 7.81 9.64 12.89
GARCH-M	4.50 3.30 3.11 3.42	2.51 2.91 3.45 4.43	12.17 21.99 41.24 76.19	3.29 4.43 6.49 10.75	4.31 6.76 11.19 20.50	3.77 4.16 5.18 6.82
EGARCH	4.19 6.26 9.56 15.34	* * * *	7.92 15.70 37.27 76.02	* * * *	* * * *	2.39 2.68 2.98 3.23
GJR	2.89 2.70 3.00 3.29	2.45 2.85 3.44 4.37	2.44 2.82 5.25 11.43	2.80 3.38 4.16 5.36	3.40 4.96 8.06 14.45	3.20 3.68 4.68 6.84
TGARCH	4.83 5.40 6.00 6.61	4.82 5.50 6.24 7.28	4.81 5.50 8.27 14.80	4.87 5.66 6.63 7.84	4.69 4.66 4.59 4.37	4.99 5.53 6.75 9.41

* Indicates value in excess of 100.

Note: Four entries per cell are for sample sizes $T = 250, 500, 1000, 2000$ respectively.