ELSEVIER Journal of Empirical Finance 4 (1997) 17–46

Journal of EMPIRICAL FINANCE

An artificial neural network-GARCH model for international stock return volatility

R. Glen Donaldson [a], Mark Kamstra [b,*]

[a] University of British Columbia, Vancouver, BC, Canada

[b] Department of Economics, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

Accepted 15 July 1996

Abstract

We construct a seminonparametric nonlinear GARCH model, based on the Artificial Neural Network (ANN) literature, and evaluate its ability to forecast stock return volatility in London, New York, Tokyo and Toronto. In-sample and out-of-sample comparisons reveal that our ANN model captures volatility effects overlooked by GARCH, EGARCH and GJR models and produces out-of-sample volatility forecasts which encompass those from other models. We also document important differences between volatility in international markets, such as the substantial persistence of volatility effects in Japan relative to North American and European markets.

JEL classification: G12; C32; C52; C53

Keywords: Volatility; Stock prices; ARCH; Artificial neural networks

1. Introduction

The preponderance of empirical evidence presented to date in the financial econometric literature suggests that stock return volatility is not only time varying but that future volatility is asymmetrically related to past return innovations, with negative unexpected returns affecting future volatility more than positive unex-

* Corresponding author. Tel.: (+1) 604-291.4514; fax: (+1) 604-291.5944; e-mail: kamstra@kamstra.econ.sfu.ca.

0927-5398/97/$17.00 Copyright © 1997 Elsevier Science B.V. All rights reserved.
PII S0927-5398(96)00011-4

pected returns.[1] Failure to capture these features of returns behavior may cause incorrect inference in tests of asset pricing relationships and can lead to the suboptimal formation of derivative strategies.[2] For these reasons, interest in asymmetric conditional volatility models – such as Glosten, Jagannathan and Runkle's (1993) Sign-GARCH (hereafter GJR) – has increased of late, as has the use of statistical procedures – such as the Engle and Ng (1993) sign bias tests – which are specifically designed to capture asymmetric volatility effects.

An important task in applied research is to decide which of the many possible volatility models one should employ in any given situation. Recent papers by Pagan and Schwert (1990) and Engle and Ng (1993) attempt to provide some guidance in this respect by evaluating the modelling ability of various statistical procedures. Unfortunately, their results suggest that none of the popularly employed models are able to adequately capture asymmetric effects, at least for the particular stock market indices and time periods studied (the S&P500 Index 1835–1937 in Pagan and Schwert (1990) and the Japanese TOPIX Index 1980–1988 in Engle and Ng (1993)) although some models, such as GJR, do come closer to capturing asymmetric effects in-sample than others, such as EGARCH. However, it is not clear from existing work how well models such as GJR would perform forecasting volatility out-of-sample and in markets other than Japan.

Given lessons learned from the existing literature, the purpose of our paper is twofold. Our first objective is to develop a new model for conditional stock volatility which can capture important asymmetric effects that existing models do not capture. To this end we develop a parsimonious seminonparametric GARCH-type model, inspired by recent work in Artificial Neural Networks (ANNs), that has the functional flexibility to capture the nonlinear relationship between past return innovations and future volatility. Recent work by Donaldson and Kamstra (1996a,b), Hutchinson et al. (1994) and others suggests that nonlinearities in financial data are well approximated by the ANN structures and logistic transforms we employ and, indeed, evidence presented below confirms their usefulness in modelling the conditional volatility of stock returns.

Our second objective is to undertake a more thorough evaluation of various volatility models – including GARCH, EGARCH, GJR and our ANN – to determine which model works best in which situation. In doing so, we depart from existing work in two respects. First, while previous comparative studies have focused on only one returns series at a time, we evaluate our models using daily stock returns data 1970–1990 from four different international markets: London,

[1] Recent papers which discuss volatility asymmetry include Campbell and Hentschel (1992), Engle and Ng (1993), Glosten et al. (1993), Nelson (1991) and Schwert (1989). See Bollerslev et al. (1992) for a review of the conditional volatility literature in general.

[2] See Diebold et al. (1993), MacKinlay and Richardson (1991), Bollerslev et al. (1988), Ferson and Harvey (1991), Engle et al. (1992), Baillie and Myers (1991) and Kroner and Sultan (1993) for evidence on the significant costs of ignoring changing volatility in empirical applications.

New York, Tokyo and Toronto. Not only does this allow us to document interesting differences between volatility in various markets, it also helps us differentiate between a model's ability to fit one particular series from its ability to explain returns behavior more generally. Second, unlike previous work, we investigate the out-of-sample performance of our ANN and various traditional volatility models using parameter estimates that are updated each day, in a rolling fashion, so as to produce for each day a new set of one-step-ahead out-of-sample forecasts on which to base our evaluations. This allows us to better assess model performance in the conditional forecasting environments in which volatility models are intended to operate.

Our ANN model for stock volatility and the data used in our investigation are presented in Sections 2 and 3 respectively. Sections 4 and 5 report model specifications and parameter estimates and present in-sample diagnostics for our various models and indices. The ability of our ANN and other traditional models to forecast stock return volatility out of sample is then studied in Section 6. Results produced reveal that our ANN model generally outperforms popular alternatives. Section 7 concludes.

2. Volatility models

Let R_t be a stock return with conditional forecast $E(R_t|I_{t-1})$, as in Eq. (1):

$$R_t = E(R_t|I_{t-1}) + \epsilon_t \tag{1}$$

where I_{t-1} is the conditioning information set on which forecasts are based and the additive forecast error ϵ_t has zero mean and conditional variance

$$E(\epsilon_t^2|I_{t-1}) = \sigma_t^2. \tag{2}$$

A well recognized problem with stock return data is that the ϵ_t appear to be drawn from a time-dependent heteroskedastic distribution.[3] The central purpose of conditional volatility models is to capture this feature of the data so as to produce a forecasted variance $\hat{\sigma}_t^2$, along with a return forecast error $\hat{\epsilon}_t$, such that the standardized residuals, $\hat{\epsilon}_t/\hat{\sigma}_t$, are homoskedastic and independent.

A great many statistical models have been proposed in an effort to capture the time-dependent heteroskedasticity of stock returns, the most popular of which are members of the ARCH family fathered by Engle (1982). As a symmetric-model benchmark against which to measure results from our ANN model below, we therefore consider the Bollerslev (1986) GARCH model,

$$\sigma_t^2 = \alpha + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j}^2 \tag{3}$$

[3] See, for example, Bollerslev et al. (1992).

which treats positive and negative return shocks symmetrically through lagged terms in ϵ^2 and which nests as special cases a variety of other symmetric volatility models, including the Engle (1982) ARCH ($p=0$), the Engle and Bollerslev (1986) IGARCH ($\Sigma_{i=1}^{p}\beta_i + \Sigma_{j=1}^{q}\gamma_j = 1$), and the Moving Average Variance Model (MAV) used by authors such as French et al. (1987) ($p=0$, $\gamma_j = 1/q\ \forall j$). Due to its extensive use in the literature, we also consider as an asymmetric benchmark the Nelson (1991) Exponential GARCH model,

$$\ln\sigma_t^2 = \alpha + \sum_{i=1}^{p}\beta_i \ln\sigma_{t-i}^2 + \sum_{j=1}^{q}\gamma_j\left[\delta(\epsilon/\sigma)_{t-j} + |(\epsilon/\sigma)_{t-j}| - \sqrt{2/\pi}\right] \tag{4}$$

which uses a logarithmic specification and the term in square brackets to account for asymmetric effects in lagged ϵ. [4]

In addition to the familiar models mentioned above, several new volatility models have been proposed of late, including Multiplicative ARCH, Piecewise-Nonlinear ARCH, Flexible Fourier Forms, Hamilton-style regime switching models, kernel estimators, and the like. Many of these have already been investigated by authors such as Pagan and Schwert (1990) and Engle and Ng (1993). Their results reveal that, while some improvement over more basic ARCH models is sometimes observed, none of these alternative models consistently or substantially outperform simple GARCH models. [5] Results from the Engle and Ng (1993) analysis of Japanese stock returns does suggest, however, that Glosten, Jagannathan and Runkle's (1993) sign-ARCH model – referred to as GJR – shows the most potential. As an additional benchmark against which to compare our new model below, we therefore also consider the GJR model in Eqs. (5) and (6).

$$\sigma_t^2 = \alpha + \sum_{i=1}^{p}\beta_i\sigma_{t-i}^2 + \sum_{j=1}^{q}\gamma_j\epsilon_{t-j}^2 + \sum_{k=1}^{r}\phi_k D_{t-k}\epsilon_{t-k}^2 \tag{5}$$

$$D_{t-k} = \begin{cases} 1 & \text{if} \quad \epsilon_{t-k} < 0, \\ 0 & \text{if} \quad \epsilon_{t-k} \geq 0 \end{cases} \tag{6}$$

[4] The definition of EGARCH(p, q) used by Nelson differs from that subsequently employed by authors such as Pagan and Schwert (1990). Here we follow the Pagan and Schwert definition, where p and q denote the number of lagged variances and normalized residuals, respectively.

[5] In a previous draft of this paper we also investigated a variety of additional models, including restrictions of (3) necessary to produce ARCH, IGARCH and MAV, as well as existing seminonparametric volatility models such as the Flexible Fourier Forms (FFF) found in Pagan and Schwert (1990). However, like Pagan and Schwert, we found that these models performed worse than even simple GARCH. This is especially true for FFF, which suffered from serious overfitting problems in the out-of-sample tests conducted below. Due to space constraints we do not report results for these models here, though they are available from the authors on request.

As can be seen from Eqs. (5) and (6), GJR is basically an augmentation of GARCH that allows past negative unexpected returns to affect volatility differently than positive unexpected returns.

We now introduce a new asymmetric model which builds on the relative success of GJR over GARCH, as documented in the tests below. This model essentially adds seminonparametric nonlinear terms to GJR in an attempt to account for nonlinear effects uncaptured by more basic ARCH models. While the approach of adding nonlinear terms to an ARCH model is by no means unique to this paper (see, for example, Engle and Ng, 1993), the manner in which these terms are added is new and, as we shall soon demonstrate, rather effective. Our particular approach is inspired by the literature on Artificial Neural Networks.

An Artificial Neural Network (ANN) is a collection of transfer functions which relate an output variable of interest, Y, to some input variables, X, which may themselves be functions of even deeper explanatory variables, e.g. $X = g(Z)$. While a wide range of transfer functions has been employed in the ANN literature, the nonlinear logistic function $Y = a + (1 + \exp[-(c + bX)])^{-1}$ is perhaps the most popular. Fully linear functions, such as $Y = a + bX$, can be used to augment the nonlinear functions if desired. [6] Fig. 1 provides a simple example, with input X measured on the horizontal axis, output Y on the vertical axis, and the ANN relating X to Y defined as $Y = \Sigma_{i=1}^{4} y_i$ with $y_1 = 1 + 0.1X$, $y_2 = 0.5 - (1 + \exp[-(75 + 40X)])^{-1}$, $y_3 = 0.5 - (1 + \exp[-(0 + 2X)])^{-1}$, and $y_4 = 1 - (1 + \exp[-(170 - 85X)])^{-1}$ being the four 'information nodes' of the network.

Notice from Fig. 1 that, when $X < -2$, $y_2 \approx 0.5$, $y_3 \approx 0.5$ and $y_4 \approx 0$, so Y's behavior is determined largely by the linear node $y_1 = 1 + 0.1X$. As X rises past about -2, y_2 rapidly decreases to its minimum of $y_2 = -0.5$ so that, by the time X reaches roughly -1.6, $Y = \Sigma y_i$ falls from 1.8 down to 0.8. Then, as X continues towards zero, the y_3 node begins to activate – although its response is less immediate given y_3's smaller gain (i.e., $2X$ instead of $40X$) – so that, by the time $X = 1.5$, $Y \approx 0.2$. Finally, as X rises past 2, the y_4 node activates to rise from 0 to 1 so that Y rises from 0.2 up to 1.2. The effect on Y of any further increases in X are then determined largely by the linear node y_1. The final shape produced by Fig. 1's ANN is thus a complicated dip-pattern in which different segments have different slopes. By extension one can see that, with many nodes activating and deactivating with different slopes and intercepts over various ranges of X, one could produce as an output response Y almost any desired function of the input X (or, with higher dimensionality, any desired function of a group of inputs $X_1, \ldots, X_n$, which could themselves be cross functions of each other and/or subordinate functions even deeper variables $X_i = g(Z)$). Indeed, as

[6] See Hertz et al. (1991) for a basic introduction to ANNs and Kuan and White (1994) for a review of the econometric issues involved and some discussion of ANNs' many economic and statistical applications.

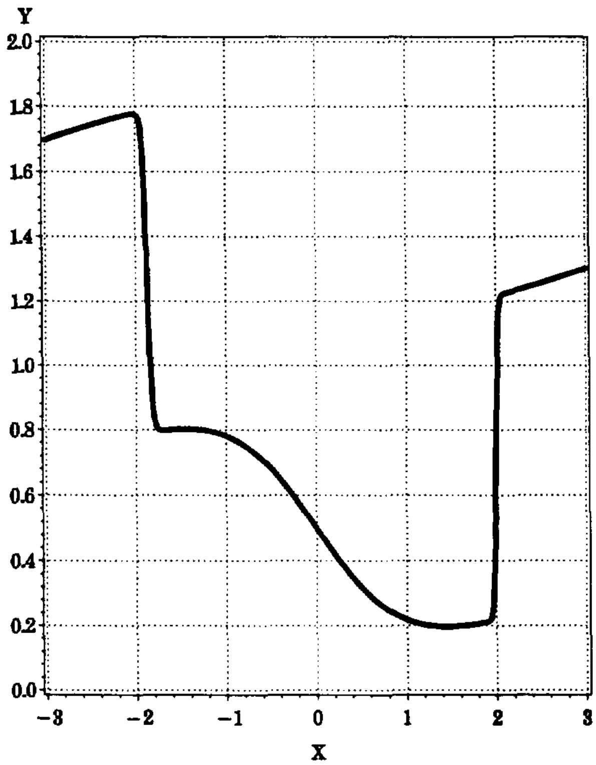

Fig. 1. Example of an ANN transfer function.

demonstrated by authors such as Hornik et al. (1989, 1990), ANNs have the ability to approximate arbitrarily well a large class of functions while requiring only a small number of parameters to be estimated relative to sample size. [7]

An ANN can be as simple as a single node with a single input. In principle, an ANN can also consist of many initial nodes which filter raw input data to produce intermediate outputs, with these intermediate outputs used as inputs to a second layer of nodes, with the second layer's outputs perhaps used as inputs to a third layers of nodes, and so on, until the ultimate output is finally produced. In practice, however, ANN researchers have found that, provided a sufficient number of nodes are placed on the first hidden layer of the ANN, higher layers are not usually needed to establish a satisfactory connection between the initial raw inputs and the final output. For example, the network whose output is graphed in Fig. 1 is a single hidden-layer ANN with four nodes: y_1, y_2, y_3, y_4. In this paper, we employ a single hidden-layer ANN where the number of nodes is selected optimally by reference to the data in a manner to be described below.

[7] For further details see Hertz et al. (1991), Hornik (1991), Stinchcombe and White (1994) and White (1989, 1990).

Use of ANNs in financial econometrics is expanding rapidly. For example, ANNs are used by Dutta and Shekhar (1988) to rate bonds, by Kimijo and Tanigawa (1990) to search for patterns in stock prices, by Tam and Kiang (1992) to predict bank failures, by Hutchinson et al. (1994) to estimate option prices, by Donaldson and Kamstra (1996a) to forecast dividends, and by Donaldson and Kamstra (1996b) to combine financial forecasts. In the current investigation of conditional stock volatility, we employ as our ANN-ARCH model the logistic augmentation of GJR specified in Eqs. (7)–(11).

$$\sigma_t^2 = \alpha + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j}^2 + \sum_{k=1}^{r} \phi_k D_{t-k} \epsilon_{t-k}^2 + \sum_{h=1}^{s} \xi_h \Psi(z_t \lambda_h) \quad (7)$$

$$D_{t-k} = \begin{cases} 1 & \text{if} \quad \epsilon_{t-k} < 0, \\ 0 & \text{if} \quad \epsilon_{t-k} \geq 0 \end{cases}, \quad (8)$$

$$\Psi(z_t \lambda_h) = \left[1 + \exp\left(\lambda_{h,0,0} + \sum_{d=1}^{v} \left[\sum_{w=1}^{m} (\lambda_{h,d,w} z_{t-d}^w) \right] \right) \right]^{-1}, \quad (9)$$

$$z_{t-d} = [\epsilon_{t-d} - E(\epsilon)] / \sqrt{E(\epsilon^2)}, \quad (10)$$

$$\frac{1}{2} \lambda_{h,d,w} \sim \text{uniform}[-1, +1]. \quad (11)$$

Eq. (8) is the GJR dummy variable, while Eq. (9) specifies the logistic ANN nodes. Eq. (10) provides a normalization of ϵ necessary to prepare the lagged unexpected returns as inputs into the nodes. All the data are transformed using the in-sample mean and variance to conform to Eq. (10)'s restriction. To achieve identification of the ξ parameters in Eq. (7) it is necessary to select values for the λ scaling factors in Eq. (9) (this is why ANN is a seminonparametric model as opposed to fully parametric). We follow the computationally simple approach of first choosing the $\lambda_{h,d,w}$ with a uniform random number generator, so the transformed lambdas lie between -1 and $+1$ as specified in Eq. (11), and then estimating the α, β, γ, ϕ, ξ parameters in Eq. (7) with maximum likelihood. Work by Stinchcombe and White (1994) yields a universal approximation result for ANNs with such an arbitrary choice of λ and, in practice, little difference is made in model performance with deviations from this convention. Further explanation is provided in Section 4 below.

3. Data and model specification

Our investigation is based on the daily closing values of the Standard and Poor's 500 Index (S&P500), the Toronto Stock Exchange Composite Index (TSEC), the Japanese Nikkei Index (NIKKEI) and London's Financial Times

Stock Exchange Index (FTSE) from January 1, 1969 to December 31, 1990. In each case the index return, R_t, is the first difference of log prices without dividends. [8]

All the volatility models we study are intended to capture the conditional variance of the unforecastable component of returns. As such we begin by investigating various specifications for conditional returns, including specifications in which returns are modelled as a constant, as a simple AR(1) process, and as a more complicated seasonal process with dummy variables for months of the year, days of the week, and each gap of i days between trading days, plus MA and AR terms as necessary to completely whiten the data in sample. In an attempt to maintain a level playing field between models and across indices, we select from among these choices 'optimal' model specifications with a single common criterion; the Schwarz Criterion. [9] Interestingly, for each index the return model ranked highest by Schwarz is the simple AR(1): $R_t = a_0 + a_1 R_{t-1} + \epsilon_t$. We therefore follow authors such as Akgiray (1989) and model the first difference of log prices, R_t in (1), as an AR(1) process. [10]

Since an important objective of this paper is to compare out-of-sample performance between models and across indices, we begin our model selection procedure by breaking our data sample in half so that model specifications can be chosen on the 1969–1979 subsample and volatility forecasted into 1980–1990. Previous comparative studies conducted on a single index (e.g., Pagan and Schwert, 1990; Engle and Ng, 1993) have exogenously set model specifications to match those traditionally employed in the literature. However, our desires to maintain a level playing field among models and across indices, and to uncover whatever differences may exist between the S&P500, TSEC, NIKKEI and FTSE, leads us to conduct a data-driven search for optimal specifications. In doing so, we use the selection algorithm described below to examine GARCH and EGARCH

[8] Dividends are not included because they are not available for all the indices. However, to investigate the robustness of our US results, we did try including dividends in the S&P500. Not surprisingly, our results were extremely similar to those based on the no-dividend S&P500 as reported in the paper. We also investigated replacing the S&P500 with the frequently employed CRSP index, but again found virtually no difference in our results. We report results for the S&P500 in our paper because the S&P500 represents roughly the same fraction of the relevant market as do the other indices we study, with a similar preponderance of higher capitalization stocks. The S&P500 is therefore more easily comparable to the other indices studied.

[9] For robustness we did investigate the use of other selection criteria, such as AIC and adjusted R^2, but found little difference in volatility results obtained. We settled on Schwarz because it chooses the most parsimonious model specifications.

[10] For robustness we did investigate other specifications for R_t in Eq. (1), including the full seasonal dummy ARMA version, but found no substantive difference in results for the volatility models. If anything, the more complex specifications for Eq. (1) added noise to our out-of-sample forecasts, suggesting the possibility that more complex seasonal ARMA models may overfit the data in some cases.

models on the grid $p,q \in [0,5]$ and GJR on the grid $p,q,r \in [0,5]$. For ANN we first obtain five different randomly produced sets of λ and then, for each of these five random λ sets, we estimate ANN on the grid $p,q,r,s,v,m \in [0,5]$. For each specification in the grid, the random λs that fits the in-sample data 'best' (to be defined below) is chosen as the candidate ANN model for that point on the specification grid. [11] Since the [0, 5] grid more than encompasses the usual range of models employed in the literature (for example, GARCH specifications of (1, 1) are most common), and since for each model and index the data always selects a specification well within the interior of the search grid, we believe that our search grid does not unduly constrain the parameter space.

To select our 'best' specification for each model and time series, we begin by jointly estimating an AR(1) specification for the mean Eq. (1) with each volatility model, in turn, using maximum likelihood on the 1969–79 data, with data from 1969 used for pre-sample conditioning. We begin with low order models (e.g., GARCH(1, 1)) and work upward into the [0, 5] grid as required to fit the data. [12] To restrict ourselves to reasonable specifications, models which produced negative variance forecasts in-sample are discarded. We also discard models which produced $\hat{\sigma}$s which lead to in-sample rejection of a Box–Pierce specification test for uncaptured autocorrelation at 24 lags in the squared standardized residuals $(\hat{\epsilon}_t/\hat{\sigma}_t)^2$. For each particular model and data series, the undiscarded specifications are then ranked according to the Schwarz Criterion. The best specification according to Schwarz is chosen as the representative for that particular modelling technique and data series. [13]

Table 1 reports the specifications chosen. As one would expect, GARCH specifications are of relatively low order in all indices. However, our model selection procedure selects the traditionally employed EGARCH(1, 1) specification for only the FTSE. As in Pagan and Schwert (1990), we find that two lags of the squared error are required for the S&P500 EGARCH. For NIKKEI, the more complex EGARCH(3, 2) is chosen by Schwarz. Indeed, the NIKKEI in general chooses more complex models than the other data series, suggesting that volatility effects in the NIKKEI may be more complex than in other markets. Formal tests of this conjecture are presented below.

[11] See Stinchcombe and White (1994) for some theoretical justification of such a random procedure.

[12] For example, consistent with results from the existing literature, GJR specifications beyond $r = 1$ were never required to remove asymmetric effects and were therefore not entertained. Conversely, up to three lagged σ were required for EGARCH in some series.

[13] Other criteria, such as AIC etc., could have been used here instead. We employ Schwarz because it delivers relatively parsimonious specifications and because it is widely used in the literature (e.g., Nelson (1991) uses Schwarz to select EGARCH models). The Schwarz Criterion has been shown to provide consistent estimation of the order of linear ARMA models by Hannan (1980). As noted by Nelson (1991), the asymptotic properties of this criterion are unknown in the context of selecting ARCH models.

Table 1
Model specifications employed [a]

Index name	GARCH p, q	EGARCH p, q	GJR p, q, r	ANN p, q, r, s, v, m
S&P500	1, 1	1, 2	1, 1, 1	1, 1, 1, 1, 1, 4
NIKKEI	2, 1	3, 2	2, 1, 1	2, 1, 1, 3, 2, 2
FTSE	1, 1	1, 1	1, 1, 1	1, 1, 0, 1, 1, 2
TSEC	2, 1	2, 1	2, 1, 1	2, 1, 1, 1, 1, 1

[a] The table reports optimal model specifications chosen with the algorithm outlined in the text for New York's S&P500, Tokyo's NIKKEI, London's FTSE and Toronto's TSEC indices on daily data 1969–1979 for the models listed below. These specifications are employed throughout the article in both in- and out-of-sample tests on 1969–1990 data.

Returns: $R_t = a_0 + a_1 R_{t-1} + \epsilon_t;\ \epsilon_t \sim (0, \sigma_t^2)$;

GARCH: $\sigma_t^2 = \alpha + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j}^2$;

EGARCH: $\ln \sigma_t^2 = \alpha + \sum_{i=1}^{p} \beta_i \ln \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j [\delta(\epsilon/\sigma)_{t-j} + |(\epsilon/\sigma)_{t-j}| - \sqrt{2/\pi}]$;

GJR: $\sigma_t^2 = \alpha + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j}^2 + \sum_{k=1}^{r} \phi_k D_{t-k} \epsilon_{t-k}^2$:

$$D_{t-k} = \begin{cases} 1 & \text{if} \quad \epsilon_{t-k} < 0, \\ 0 & \text{if} \quad \epsilon_{t-k} \geq 0; \end{cases}$$

ANN: $\sigma_t^2 = \alpha + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j}^2 + \sum_{k=1}^{r} \phi_k D_{t-k} \epsilon_{t-k}^2 + \sum_{h=1}^{s} \xi_h \Psi(z_t \lambda_h)$;

$$D_{t-k} = \begin{cases} 1 & \text{if} \quad \epsilon_{t-k} < 0, \\ 0 & \text{if} \quad \epsilon_{t-k} \geq 0; \end{cases}$$

$\Psi(z_t \lambda_h) = [1 + \exp(\lambda_{h,0,0} + \sum_{d=1}^{v} [\sum_{w=1}^{m} (\lambda_{h,d,w} z_{t-d}^{w})])]^{-1}$:

$z_{t-d} = (\epsilon_{t-d} - E(\epsilon))/E(\sqrt{\epsilon^2})$;

$\frac{1}{2} \lambda_{h,d,w} \sim \text{uniform}[-1, +1]$.

In terms of nonlinear effects in the ANN model, it is interesting to note that, for ANN, the S&P500, TSEC and NIKKEI require only one GJR term ($r = 1$) while, in the FTSE, the ANN terms capture enough nonlinearity so that no GJR terms are required ($r = 0$). This suggests that ANN should dominate GJR (and thus GARCH) in-sample, although there is of course no guarantee that ANN should dominate out-of-sample. Indeed, if ANN overfits the in-sample data, as do many seminonparametric forms, then out-of-sample performance will suffer. As shall be demonstrate in Section 6 below, an attractive feature of ANN is that it does not tend to overfit, unlike Flexible Fourier Forms and other seminonparametric structures (see Pagan and Schwert, 1990). [14]

4. Parameter estimation and in-sample diagnostics

Parameter estimation is conducted jointly on an AR(1) specification for Eq. (1) – i.e., $R_t = a_0 + a_1 R_{t-1} + \epsilon_t;\ \epsilon_t \sim (0, \sigma_t^2)$ – and the appropriate model for σ_t^2.

[14] For completeness we did estimate FFF models for our data but, like Pagan and Schwert (1990), found them to perform very badly both in- and out-of-sample in spite of careful model selection procedures. We do not report FFF results here due to space constraints.

Table 2
GARCH estimation results and in-sample diagnostics, January 1, 1969 to December 31, 1979 [a]

Panel A: Parameter estimates (Bollerslev–Wooldridge robust standard errors)

Parameter	S&P500	NIKKEI	FTSE	TSEC
a_0	1.985E−04	4.245E−04	3.050E−05	2.608E−04
	(1.366E−04)	(2.052E−04)	(2.424E−04)	(1.180E−04)
a_1	2.436E−01	1.363E−01	7.742E−02	2.883E−01
	(1.915E−02)	(2.703E−02)	(2.069E−02)	(2.476E−02)
α	7.367E−07	6.645E−06	2.749E−06	5.344E−07
	(2.511E−07)	(1.655E−06)	(9.572E−07)	(2.313E−07)
β_1	9.232E−01	5.734E−01	9.082E−01	9.411E−01
	(1.456E−02)	(6.733E−02)	(1.490E−02)	(1.579E−02)
γ_1	6.584E−02	1.967E−01	7.924E−02	2.539E−01
	(1.425E−02)	(5.920E−02)	(1.256E−02)	(5.499E−02)
γ_2	–	1.950E−01	–	−2.055E−01
	–	(1.055E−01)	–	(5.335E−02)

Panel B: Diagnostics (p-values with the exception of the log likelihood)

Statistic	S&P500	NIKKEI	FTSE	TSEC
Log likelihood value	11099.829	12821.015	9790.554	11528.891
ARCH test	0.929	1.000	0.945	0.904
Sign bias test	0.269	0.152	0.013	0.290
Neg. sign bias test	0.001	0.188	0.028	0.031
Pos. sign bias test	0.045	0.076	0.359	0.033
Joint test	0.019	0.501	0.268	0.054

[a] The table reports parameter estimates and standard diagnostics for New York's S&P500, Tokyo's NIKKEI, London's FTSE and Toronto's TSEC indices on daily data 1969–1979 for the model listed below.
Returns: $R_t = a_0 + a_1 R_{t-1} + \epsilon_t$; $\epsilon_t \sim (0, \sigma_t^2)$;
GARCH: $\sigma_t^2 = \alpha + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j}^2$

Results for the 1969–1979 subperiod are reported in Tables 2–5, which present parameter estimates (along with Bollerslev and Wooldridge (1992) standard errors) and in-sample diagnostics on the standardized residuals, $\hat{\epsilon}_t / \hat{\sigma}_t$, for Table 1's specifications. Standard diagnostics include the model log likelihood and the p-value from a traditional Ljung and Box (1978) test for symmetric ARCH at 24 lags. Since in this paper we are particularly concerned with asymmetric volatility, and our search for optimal model specifications was conducted over a grid with up to five lags – though, as Table 1 reveals, no model optimally selected anywhere near this many lags – we also report the Engle and Ng (1993) Sign Bias Test, Negative Sign Bias Test, Positive Sign Bias Test and Joint Sign Bias Test, all at five lags, for the presence of asymmetric ARCH effects.

Table 2 reports results for GARCH. The parameters a_0 and a_1 are the intercept and AR(1) coefficient, respectively, for the return Eq. (1). The remaining parameters are from the GARCH volatility model in Eq. (3). For all indices, parameter

Table 3
EGARCH estimation results and in-sample diagnostics, January 1, 1969 to December 31, 1979 [a]

Panel A: Parameter estimates (Bollerslev–Wooldridge robust standard errors)

Parameter	S&P500	NIKKEI	FTSE	TSEC
a_0	1.584E−05	1.434E−04	−1.784E−04	4.154E−04
	(1.373E−04)	(1.149E−04)	(2.413E−04)	(1.160E−04)
a_1	2.384E−01	1.738E−01	7.447E−02	2.975E−01
	(1.973E−02)	(3.089E−02)	(2.120E−02)	(2.336E−02)
α	−4.226E−02	−8.077E−02	−1.368E−01	−1.874E−01
	(5.584E−01)	(4.602E−01)	(4.205E−01)	(4.101E−01)
β_1	1.807E+00	1.471E+00	9.840E−01	9.805E−01
	(8.243E−01)	(2.659E−01)	(5.607E−02)	(4.449E−02)
β_2	−8.112E−01	−4.795E−01	–	–
	(8.050E−01)	(2.331E−01)	–	–
δ	−5.937E−01	−3.499E−01	−1.918E−01	4.875E−02
	(1.451E+00)	(1.338E−01)	(1.906E−01)	(1.037E−01)
γ_1	2.824E−02	3.320E−01	1.841E−01	4.365E−01
	(5.211E−02)	(1.665E−01)	(9.419E−02)	(1.738E−01)
γ_2	–	−5.976E−02	–	−2.896E−01
	–	(8.689E−02)	–	(1.414E−01)
γ_3	–	−2.050E−01	–	–
	–	(7.468E−02)	–	–

Panel B: Diagnostics (p-values with the exception of the log likelihood)

Statistic	S&P500	NIKKEI	FTSE	TSEC
Log likelihood value	11123.780	12896.733	9787.618	11536.110
ARCH test	0.901	1.000	0.988	0.972
Sign bias test	0.663	0.750	0.148	0.554
Neg. sign bias test	0.031	0.695	0.390	0.459
Pos. sign bias test	0.346	0.646	0.584	0.367
Joint test	0.198	0.965	0.653	0.357

[a] The table reports parameter estimates and standard diagnostics for New York's S&P500, Tokyo's NIKKEI, London's FTSE and Toronto's TSEC indices on daily data 1969–1979 for the model listed below.
Returns: $R_t = a_0 + a_1 R_{t-1} + \epsilon_t$; $\epsilon_t \sim (0, \sigma_t^2)$;
EGARCH: $\ln \sigma_t^2 = \alpha + \Sigma_{i=1}^{p} \beta_i \ln \sigma_{t-i}^2 + \Sigma_{j=1}^{q} \gamma_j [\delta(\epsilon/\sigma)_{t-j} + |(\epsilon/\sigma)_{t-j}| - \sqrt{2/\pi}]$.

estimates are consistent with those generally reported in the literature. In particular, volatility appears nearly integrated (though in all cases IGARCH restrictions produced models inferior to the unrestricted GARCH). Not surprisingly, the Ljung–Box ARCH Test results show GARCH removing symmetric ARCH effects in-sample for all data series. However, the four sign bias tests reveal significant uncaptured asymmetric ARCH in-sample for 1969–1979, with several p-values below 0.05 in the various indices.

Table 4
GJR estimation results and in-sample diagnostics, January 1, 1969 to December 31, 1979 [a]

Panel A: Parameter estimates (Bollerslev–Wooldridge robust standard errors)

Parameter	S&P500	NIKKEI	FTSE	TSEC
a_0	1.455E−06	2.256E−04	−1.600E−04	2.565E−04
	(1.360E−04)	(1.563E−04)	(2.411E−04)	(1.141E−04)
a_1	2.391E−01	1.734E−01	7.301E−02	2.869E−01
	(1.897E−02)	(2.347E−02)	(2.076E−02)	(2.480E−02)
α	4.519E−07	7.832E−06	2.350E−06	5.301E−07
	(1.493E−07)	(2.345E−06)	(8.782E−07)	(2.271E−07)
β_1	9.500E−01	5.492E−01	9.168E−01	9.419E−01
	(8.943E−03)	(8.717E−02)	(1.438E−02)	(1.564E−02)
γ_1	2.402E−03	4.684E−03	4.863E−02	2.532E−01
	(7.468E−03)	(2.200E−02)	(1.265E−02)	(6.074E−02)
γ_2	–	2.338E−01	–	−2.077E−01
	–	(1.266E−01)	–	(5.405E−02)
ϕ_1	8.471E−02	3.041E−01	4.949E−02	3.848E−03
	(1.670E−02)	(9.282E−02)	(1.801E−02)	(2.012E−02)

Panel B: Diagnostics (p-values with the exception of the log likelihood)

Statistic	S&P500	NIKKEI	FTSE	TSEC
Log likelihood value	11124.268	12849.449	9797.037	11528.968
ARCH test	0.691	1.000	0.975	0.907
Sign bias test	0.517	0.753	0.064	0.177
Neg. sign bias test	0.056	0.850	0.198	0.171
Pos. sign bias test	0.279	0.772	0.416	0.616
Joint test	0.265	0.987	0.507	0.536

[a] The table reports parameter estimates and standard diagnostics for New York's S&P500, Tokyo's NIKKEI, London's FTSE and Toronto's TSEC indices on daily data 1969–1979 for the model listed below.

Returns: $R_t = a_0 + a_1 R_{t-1} + \epsilon_t$; $\epsilon_t \sim (0, \sigma_t^2)$;

GJR: $\sigma_t^2 = \alpha + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j}^2 + \sum_{k=1}^{r} \phi_k D_{t-k} \epsilon_{t-k}^2$;

$$D_{t-k} = \begin{cases} 1 & \text{if} \quad \epsilon_{t-k} < 0, \\ 0 & \text{if} \quad \epsilon_{t-k} \geq 0. \end{cases}$$

Table 3 reports results for EGARCH from Eq. (4), with a_0 and a_1 again being the intercept and AR(1) coefficient for the return Eq. (1). In-sample 1969–1979, EGARCH passes all the ARCH and sign tests for uncaptured volatility at the 5% level, except the Negative Sign Bias Test for the S&P500. As in Engle and Ng (1993), we find the asymmetry parameter δ to be significantly negative in the NIKKEI. However, in all other indices, δ is not significantly different from zero, suggesting that EGARCH may not be the preferred specification to model asymmetric effects in the S&P500, FTSE and TSEC.

Table 4 reports results for GJR from Eqs. (5) and (6). Most noteworthy is the observation that in the S&P500, NIKKEI and FTSE the asymmetry parameter ϕ is significantly positive, thereby confirming that negative return innovations lead to more volatility than positive return innovations in these countries. The GJR parameter is not statistically significant in the TSEC, but the Engle–Ng tests nevertheless reveal that the addition of the GJR term removes all significant traces

Table 5
ANN estimation results and in-sample diagnostics. January 1, 1969 to December 31, 1979 [a]

Panel A: Parameter estimates (Bollerslev–Wooldridge robust standard errors)

Parameter	S&P500	NIKKEI	FTSE	TSEC
a_0	1.331E−05	2.332E−04	−1.457E−04	2.157E−04
	(1.351E−04)	(1.263E−04)	(2.374E−04)	(1.178E−04)
a_1	2.433E−01	2.003E−01	7.786E−02	3.114E−01
	(1.900E−02)	(2.048E−02)	(2.043E−02)	(2.429E−02)
α	5.687E−06	−8.685E−05	−3.685E−06	3.568E−06
	(2.284E−06)	(3.923E−05)	(1.684E−06)	(8.644E−07)
β_1	9.510E−01	7.157E−01	9.136E−01	9.481E−01
	(8.400E−03)	(4.027E−02)	(1.486E−02)	(1.525E−02)
γ_1	−1.856E−03	−5.223E−02	9.552E−02	2.566E−01
	(7.515E−03)	(2.992E−02)	(1.702E−02)	(5.850E−02)
γ_2	–	−3.775E−02	–	−2.059E−01
	–	(2.796E−02)	–	(5.248E−02)
ϕ_1	8.967E−02	3.084E−01	–	−3.982E−02
	(1.667E−02)	(7.049E−02)	–	(2.251E−02)
ξ_1	−1.353E−01	−2.576E−04	−2.678E−05	1.508E−01
	(5.770E−02)	(1.221E−04)	(8.727E−06)	(3.829E−02)
ξ_2	–	9.884E−05	–	–
	–	(6.474E−05)	–	–
ξ_3	–	1.059E−04	–	–
	–	(5.113E−05)	–	–

Panel B: Diagnostics (p-values with the exception of the log likelihood)

Statistic	S&P500	NIKKEI	FTSE	TSE
Log likelihood value	11127.538	12896.835	9798.147	11542.716
ARCH test	0.714	1.000	0.953	0.826
Sign bias test	0.435	0.262	0.187	0.268
Neg. sign bias test	0.059	0.491	0.295	0.134
Pos. sign bias test	0.276	0.699	0.602	0.760
Joint test	0.203	0.869	0.828	0.228

[a] The table reports parameter estimates and standard diagnostics for New York's S&P500, Tokyo's NIKKEI, London's FTSE and Toronto's TSEC indices on daily data 1969–1979 for the model listed below.

Returns: $R_t = a_0 + a_1 R_{t-1} + \epsilon_t$; $\epsilon_t \sim (0, \sigma_t^2)$;

ANN: $\sigma_t^2 = \alpha + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j}^2 + \sum_{k=1}^{r} \phi_k D_{t-k} \epsilon_{t-k}^2 + \sum_{h=1}^{s} \xi_h \Psi(z_t \lambda_h)$;

$$D_{t-k} = \begin{cases} 1 & \text{if} \quad \epsilon_{t-k} < 0, \\ 0 & \text{if} \quad \epsilon_{t-k} \geq 0; \end{cases}$$

$\Psi(z_t \lambda_h) = [1 + \exp(\lambda_{h,0,0} + \sum_{d=1}^{v} [\sum_{w=1}^{m} (\lambda_{h,d,w} z_{t-d}^{w})])]^{-1}$;

$z_{t-d} = (\epsilon_{t-d} - \mathrm{E}(\epsilon))/\mathrm{E}(\sqrt{\epsilon^2})$; $\frac{1}{2}\lambda_{h,d,w} \sim$ uniform$[-1, +1]$.

of asymmetric ARCH as intended. Indeed, GJR is the only traditional model to remove all evidence of asymmetric volatility, at the 5% significance level, in every index over the 1969–1979 sample period on which model specifications were selected. In this respect, GJR seems preferred to our EGARCH benchmark as a model for asymmetric volatility.

Table 5 reports results for ANN from Eqs. (7)–(11), joint with Eq. (1). First note that in the S&P500, NIKKEI and TSEC indices where ϕ GJR terms are

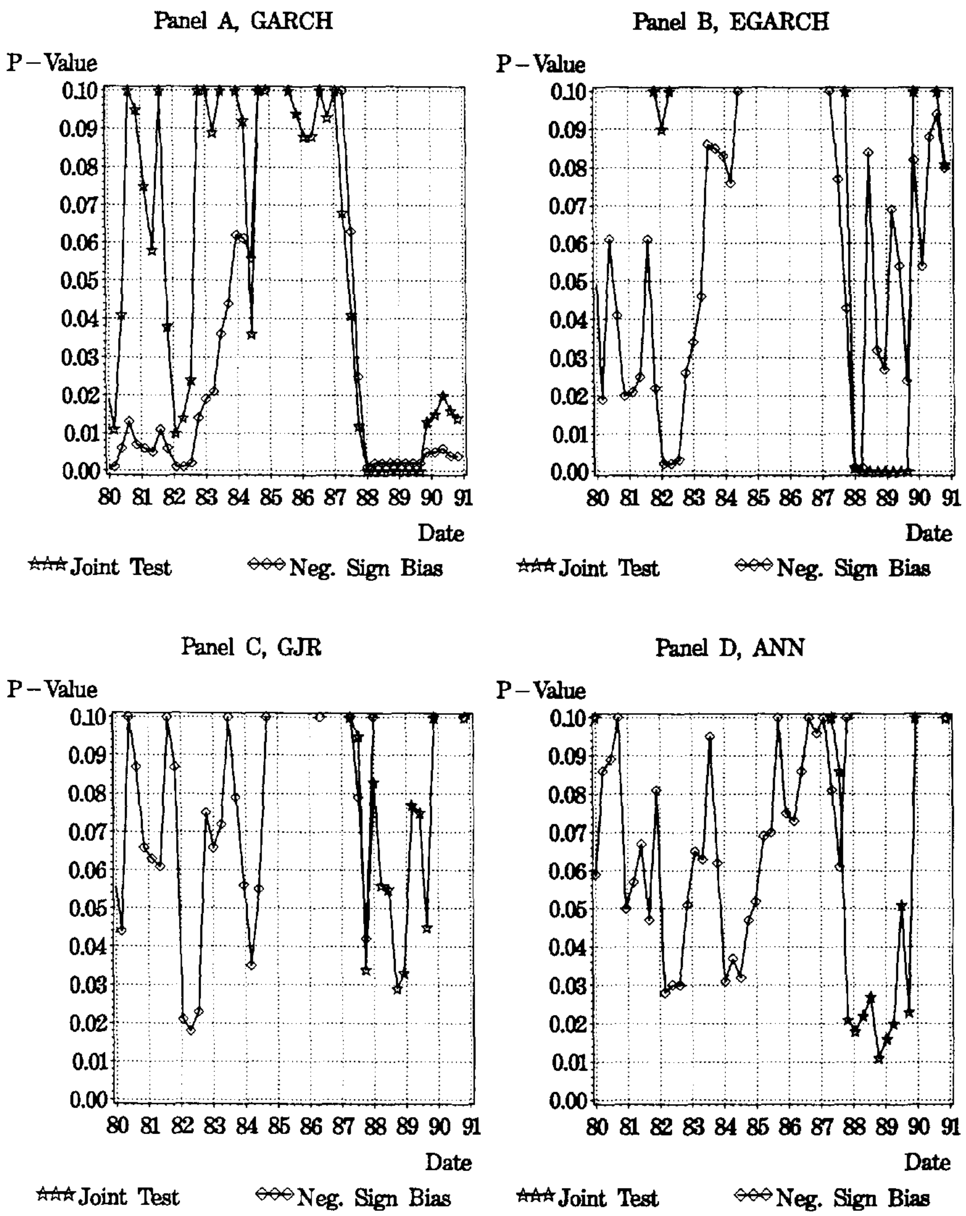

Fig. 2. S&P500 Engle–Ng sign bias test results.

required, the ϕ terms retain their former significance and assume values roughly equal to those from the base GJR model in Table 4. This suggests that, in these indices, the ANN terms are capturing asymmetric volatility effects in addition to – not instead of – effects captured by GJR. Conversely, in the FTSE, the inclusion of ANN terms removes the necessity for GJR terms. (Indeed, in the FTSE the only difference between the model specifications in Tables 4 and 5 is that a GJR term augments a simple GARCH(1, 1) in Table 4, while a single ANN term augments

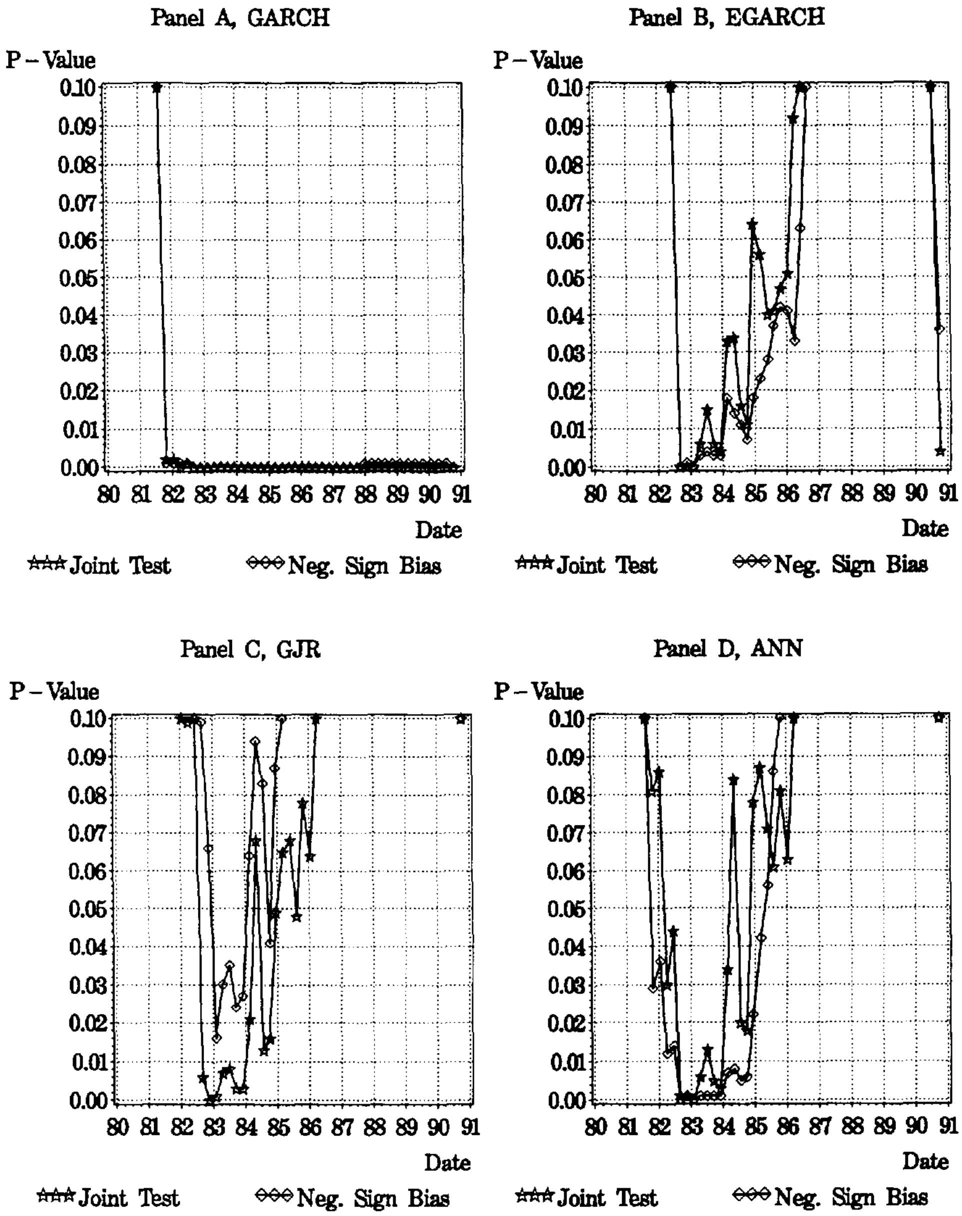

Fig. 3. NIKKEI Engle–Ng sign bias test results.

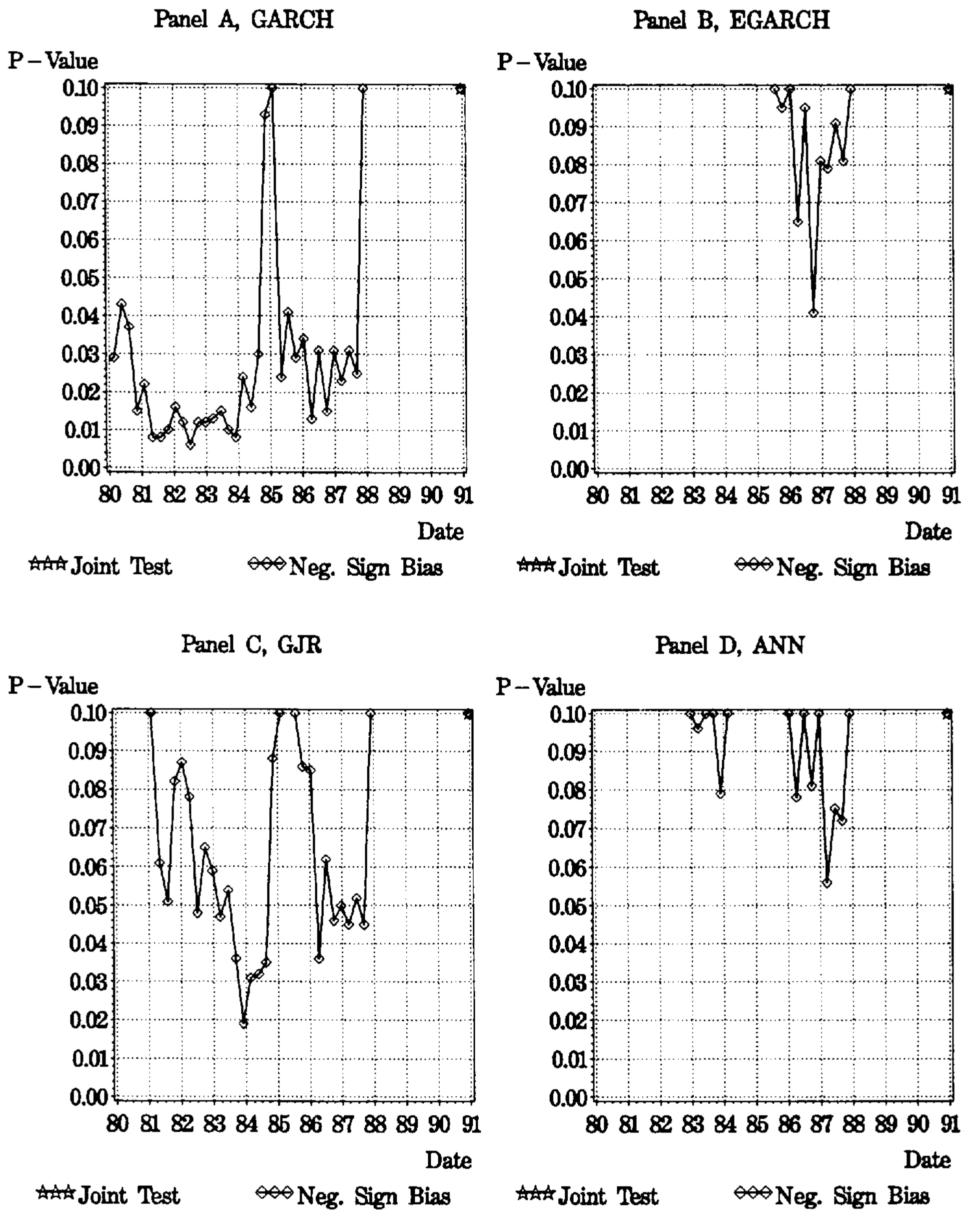

Fig. 4. FTSE Engle–Ng sign bias test results.

the same simple GARCH(1, 1) in Table 5, a distinction whose importance will become evident in Figs. 2–5 below.) The log likelihood value for every index's ANN model is also higher than the log likelihood value for every alternative model, including GJR. This suggests to us that, at least on the 1969–1979 data, ANN is generally preferred to GARCH, EGARCH and GJR for its ability to capture both asymmetric and symmetric ARCH effects in sample.

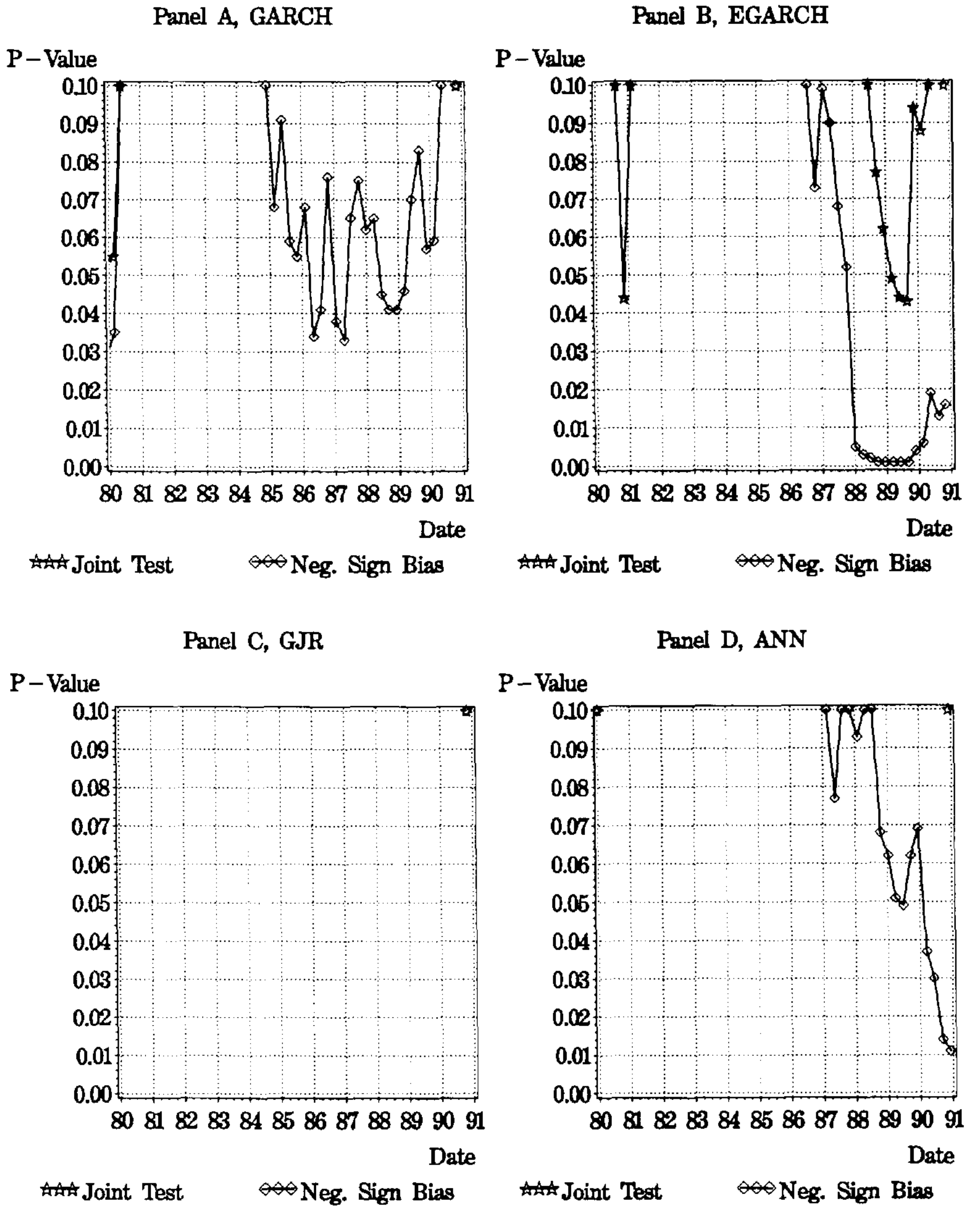

Fig. 5. TSEC Engle–Ng sign bias test results.

Table 6 contains summary statistics on the in-sample conditional variances, $\hat{\sigma}_t^2$, and standardized returns, $\hat{\epsilon}_t/\hat{\sigma}_t$, from our four models and four data series. If these models were producing conditional variances that yielded standard normal standardized returns then, for the standardized returns, skewness would be zero and kurtosis would be 3. As can be seen from Table 6, all the models produce skewed and leptokurtic distributions, though ANN generally comes closest to normality (e.g., in three of the four indices ANN produces the lowest standardized

Table 6
Summary statistics on in-sample fitted variances and returns, January 1, 1969 to December 31, 1979 [a]

Method	Fitted conditional variance				Fitted standardized returns		
	Mean $\times 10^{-4}$	Std. dev $\times 10^{-4}$	Skew	Kurtosis	Std dev	Skew	Kurtosis
Panel A: S&P500							
Raw data	0.691	1.504	7.305	86.345	1.000	0.352	5.743
GARCH	0.689	0.550	2.599	10.986	1.000	0.084	3.483
EGARCH	0.672	0.513	2.426	9.746	1.000	0.050	3.382
GJR	0.676	0.521	2.233	8.251	1.000	0.048	3.387
ANN	0.676	0.513	2.093	7.737	1.000	0.051	3.326
Panel B: NIKKEI							
Raw data	0.726	2.860	14.888	308.63	1.000	−1.001	16.526
GARCH	0.823	1.593	8.091	87.118	1.000	−1.297	15.184
EGARCH	1.008	13.321	50.503	2640.7	1.000	−1.449	18.222
GJR	0.826	1.788	9.394	116.71	1.000	−1.312	16.700
ANN	0.697	1.039	9.224	126.98	1.000	−1.021	11.602
Panel C: FTSE							
Raw data	2.333	5.287	6.426	61.482	1.000	0.240	6.141
GARCH	2.312	2.333	3.405	17.991	1.000	0.046	3.550
EGARCH	2.261	2.007	2.702	12.074	1.000	0.074	3.620
GJR	2.294	2.199	2.848	12.850	1.000	0.059	3.524
ANN	2.388	2.677	3.660	20.135	1.000	0.054	3.584
Panel D: TSEC							
Raw data	0.545	2.109	13.142	234.86	1.000	−0.569	16.003
GARCH	0.544	0.774	9.033	139.95	1.000	0.160	7.524
EGARCH	0.500	0.672	19.156	628.49	0.999	0.307	8.930
GJR	0.544	0.775	9.077	141.18	1.000	0.159	7.525
ANN	0.504	0.633	8.758	132.88	1.000	0.181	6.956

[a] The table reports summary statistics on the in-sample fitted conditional variances, $\hat{\sigma}_t^2$, and standardized returns, $\hat{\epsilon}_t / \hat{\sigma}_t^2$, for New York's S&P500, Tokyo's NIKKEI, London's FTSE and Toronto's TSEC indices on daily data 1969–1979 (with 1969 used as pre-sample conditioning information) for the models listed at the bottom of Table 1.

return kurtosis). Only EGARCH is ill behaved, occasionally producing conditional variances and/or standardized returns with more skew and/or kurtosis than the raw data. This confirms the Engle and Ng (1993, p. 1171) conjecture that EGARCH may be too extreme in the tails for some data series and supports our favoring of ANN over other models, at least in sample.

5. One-step-ahead recursive in-sample diagnostics

Before proceeding to out-of-sample testing, we first compute and evaluate in-sample the updated models used to obtain the one-step-ahead out-of-sample forecasts on which Section 6's out-of-sample tests will be based. To do this, we first estimate parameters for each model on the January 1, 1969 to December 31, 1979 data and produce an out-of-sample forecast of $\hat{\epsilon}_t/\hat{\sigma}_t$ for the first trading day of 1980. We then re-estimate the model parameters on data up to and including the first trading day of 1980 – but still using the original specifications listed in Table 1 – to obtain a new set of updated parameter values and in-sample diagnostic test results, and to produce a one-step-ahead out-of-sample forecast of $\hat{\epsilon}_t/\hat{\sigma}_t$ for the second trading day of 1980. We then repeat this estimation–diagnostic-forecast process using data up to and including the second trading day of 1980, then the third day, and so forth, until we have obtained recursively updated parameter estimates and one-step-ahead out-of-sample volatility forecasts for each model and stock index for every trading day from January 1, 1980 to December 31, 1990.

We obviously cannot report the recursively updated parameter estimates and in-sample diagnostic results in the same detail as Tables 2–5 above. However, the results of the Engle–Ng Tests are summarized in Figs. 2–5. Along the bottom of each panel we report the last date in the recursively updated sample; e.g., 19*XX*. The vertical axis reports the *p*-values from the tests performed on that sample. A *p*-value of .*XX* therefore signifies rejection of the null hypothesis of no uncaptured volatility at the *XX*% significance level for the particular Engle–Ng test in question. For viewing simplicity, and since test results normally change little from day to day, we only plot test results for the last day of each quarter. To avoid clutter, we also only report results for the Negative Sign Bias Test (diamonds) and the Joint Test (stars) since these tests are failed much more often than the Basic Sign Bias Test and Positive Sign Bias Test.

Fig. 2 reports results for the S&P500, with Panel A giving *p*-values for Engle–Ng tests on GARCH $\hat{\epsilon}_t/\hat{\sigma}_t$ standardized residuals, Panel B for EGARCH, Panel C for GJR, and Panel D for ANN. From Panel A we see that GARCH fails to capture negative asymmetric effects in the S&P500 during 1980–1983 and again from the Crash of 1987 to the end of the sample. EGARCH performs slightly better than GARCH, with some recovery following the '87 crash. GJR does better than either standard model at capturing the conditional volatility of stock returns prior to the 1987 Crash and recovers more quickly following the Crash. In the S&P500, the performance of ANN is roughly similar to GJR, with ANN failing less significantly than GJR in the early 1980s but more significantly during the late 1980s.

Results in Fig. 3 reveal essentially the same ranking for the NIKKEI as was observed in the S&P500: GARCH performs worst, EGARCH next, and GJR and ANN about the same (if we had also plotted the Basic Sign Bias results then ANN would be clearly favored as GJR fails the Basic Sign Bias Test during the early 1980s while others do not). However, while in the S&P500 the quality progression from GARCH to ANN is gradual, in the NIKKEI there is a clear split between the linear GARCH model and nonlinear EGARCH, GJR and ANN models. In particular, the three nonlinear models are easily able to capture

NIKKEI volatility during the 1987 crash, while linear GARCH fails miserably for almost the entire sample. This suggests that nonlinear effects may be particularly important in the NIKKEI.

Fig. 4 reports results for the FTSE. This figure presents the clearest example of the progression of in-sample fit from GARCH through ANN, with ANN being the only model that does not fail at least once at the 5% significance level. Perhaps the most interesting feature of the FTSE results is the observation that the ANN's performance on the Engle–Ng tests is very similar to that of EGARCH, while GJR is closer to GARCH. Thus, while in the S&P500 ANN mimics GJR more closely than any other model with respect to Engle–Ng performance, in the NIKKEI ANN mimics EGARCH more closely than any other model. This finding reveals something of the ANN's flexibility to fit a wide variety of interesting specifications; when GJR is best ANN looks something like GJR, but when EGARCH is best ANN looks more like EGARCH. This is supported by Tables 1 and 5 where it is revealed that, in the FTSE, ANN selects a model specification with no GJR term, while in the other indices a GJR term is included.

Finally, Fig. 5 plots results for the TSEC. In this index GJR is clearly the most successful at passing the Engle–Ng tests, although ANN's failure occurs only in during 1990. This failure is no doubt due in part to the fact that we use the same model specification estimated on 1969–1979 data for the entire 1980–1990 period (though parameter estimates are of course updated each day) and thus the farther away one gets from the specification period the less well the model might be expected to perform. More significant is the observation that EGARCH in particular has trouble capturing the 1987 crash in the TSEC, just as all the volatility models had trouble capturing the 1987 crash in the S&P500. Conversely, none of the nonlinear models have particular trouble capturing the stock market crash in the NIKKEI or FTSE. This finding suggests that there is a fundamental difference in the unexplainable asymmetric volatility effects of the '87 crash in the various markets we study, as seen through the eyes of our various volatility models: New York's S&P500 seems affected most, Toronto's TSEC second, Tokyo's NIKKEI third, and London's FTSE least.

6. Forecasting volatility out of sample

Thus far, all our comparisons and evaluations of the various models under consideration have been based on in-sample diagnostics. However, as noted by authors such as Pagan and Schwert (1990), the true test of a volatility model is its ability to forecast the conditional volatility of stock returns out of sample. This is especially true when evaluating seminonparametric models since there is always the chance that superior in-sample performance might be the result of overfitting the data. Indeed, in their comparison of various volatility models using pre-war US data, Pagan and Schwert (1990) found serious problems associated with substantial

Table 7
Summary statistics on one-step-ahead out-of-sample variance and return forecasts. January 1, 1980 to December 31, 1990 [a]

Method	Forecasted conditional variance				Forecasted standardized returns		
	Mean $\times 10^{-4}$	Std dev $\times 10^{-4}$	Skew	Kurtosis	Std Dev	Skew	Kurtosis
Panel A: S&P500							
Raw data	1.230	9.532	42.679	2042.7	1.000	−2.575	61.484
GARCH	1.143	2.881	13.103	193.41	1.060	−1.042	15.502
EGARCH	0.944	1.159	12.623	241.43	1.105	−1.544	23.425
GJR	1.193	3.607	13.985	220.27	1.059	−0.830	12.097
ANN	1.196	3.688	14.363	232.71	1.059	−0.790	11.626
Panel B: NIKKEI							
Raw data	0.936	6.287	29.380	1034.0	1.000	−0.499	46.154
GARCH	0.929	3.194	15.931	341.95	1.053	−1.024	13.056
EGARCH	0.806	6.127	51.224	2751.6	1.076	−0.761	9.816
GJR	0.912	3.533	18.525	451.06	1.066	−0.728	8.885
ANN	0.850	3.231	20.637	580.83	1.076	−0.735	8.758
Panel C: FTSE							
Raw data	1.200	4.017	22.531	663.96	1.000	−0.901	12.423
GARCH	1.244	1.183	10.147	137.03	0.977	−0.850	11.388
EGARCH	1.211	0.932	9.056	128.04	0.989	−0.789	10.518
GJR	1.257	1.387	11.423	164.79	0.978	−0.811	11.147
ANN	1.270	1.705	14.040	240.56	0.980	−0.737	10.198
Panel D: TSEC							
Raw data	0.762	3.932	22.097	616.25	1.000	−0.418	27.665
GARCH	0.770	1.768	13.489	249.48	1.017	−0.645	8.258
EGARCH	0.687	1.106	12.497	209.68	1.017	−0.602	7.954
GJR	0.772	1.792	13.242	230.53	1.016	−0.647	8.235
ANN	0.765	1.789	13.521	237.71	1.017	−0.666	8.271

The table reports summary statistics on the one-step-ahead out-of-sample forecasted conditional variances, $\hat{\sigma}_t^2$, and standardized returns, $\hat{\epsilon}_t / \hat{\sigma}_t^2$, for New York's S&P500, Tokyo's NIKKEI, London's FTSE and Toronto's TSEC indices on daily data 1980–1990 for the models listed at the bottom of Table 1.

overfitting in the seminonparametric models they investigated. [15] For this reason it is especially interesting to note in the tests below that our ANN model produces superior results without overfitting the data.

[15] As noted above, we also investigated the Flexible Fourier Form model using our four-country data and found substantial overfitting problems in our data, confirming the nature of Pagan and Schwert's results.

Table 7 contains summary statistics on the one-step-ahead out-of-sample conditional volatility forecasts, $\hat{\sigma}_t^2$, and the one-step-ahead out-of-sample forecasted standardized residuals, $\hat{\epsilon}_t/\hat{\sigma}_t$, from January 1, 1980 to December 31, 1990, as obtained with the updating–forecasting procedure described at the beginning of the previous section. Panels A, B, C and D present results for the S&P500, NIKKEI, FTSE and TSEC, respectively (as one would expect, the period surrounding the 1987 Crash is the source of the increased skew and kurtosis in the raw data over that from Table 6.) The most significant finding from Table 7 is the revelation that the ANN models do not overfit the data, as evidenced by the fact that ANN's volatility forecasts are not excessively variable. Indeed, in three of the four indices ANN produces the lowest standardized return kurtosis of any model. Conversely, EGARCH produces a forecasted conditional variance that has kurtosis even greater than the NIKKEI raw data, suggesting that EGARCH may not be a suitable model for the conditional volatility of the NIKKEI.

To evaluate the out-of-sample forecasting performance of each model relative to the other models, we conduct a Chong and Hendry (1986) forecast encompassing test. [16] To formalize the notion of forecast encompassing, note that the forecast error from a correctly specified model has a conditional first moment of 0 given any conditioning information available to the forecaster. Thus, given two different models estimated on the same data, Model j's forecast error should be orthogonal to Model k's forecast provided that Model j accurately fits the data. Provided that Model j does accurately fits the data, conditioning on the forecast from Model k should therefore not help to explain any of Model j's forecast error.

Model j encompasses Model k if Model j can explain what Model k cannot explain, without Model k being able to explain what Model j cannot explain. The Chong and Hendry (1986) encompassing tests are therefore based on a set of OLS regressions of the forecast error from one model on the forecast from the other model. Thus, with $(\hat{\epsilon}_{t,j}^2 - \hat{\sigma}_{t,j}^2)$ being Model j's forecast error and $\hat{\sigma}_{t,k}^2$ being

[16] In addition to the encompassing tests reported below several other tests were also investigated. For example, we conducted comparisons by mean squared forecast error and mean absolute forecast error. MSFE and MAFE comparisons do reveal a slight preference for ANN over other models; however, these results do not allow formal comparisons of significant difference and are considerably less informative than the encompassing tests. MSFE and MAFE results are therefore not reported here due to space constraints, though they are available from the authors. We also investigated use of the Diebold and Mariano (1995) test for predictive accuracy. However, on conducting a Monte Carlo analysis of the test's size and power properties for our particular model comparisons, we found that the Diebold–Mariano test had low power relative to the encompassing tests and a significant size bias. Because of this, and since results from the Diebold–Mariano tests we conducted generally support results from the encompassing tests in any case, we do not report the Diebold–Mariano test results here. Finally, multi-period forecast comparisons were also considered but not employed due to the great difficulty in their implementation with nonlinear models.

Model k's forecast, we test for significance of the β parameter in the regression in Eq. (12):

$$\left(\hat{\epsilon}_{t,j}^{2} - \hat{\sigma}_{t,j}^{2}\right) = \alpha_{j,k} + \beta_{j,k}\sigma_{t,k}^{2} + \nu_{t,j} \tag{12}$$

in which ν is a random error. [17]

We first regress the forecast error from Model j on the forecast from Model k, as in Eq. (12), to obtain the estimated coefficient $\hat{\beta}_{j,k}$. We then regress the forecast error from Model k on the forecast from Model j to obtain $\hat{\beta}_{k,j}$. If $\hat{\beta}_{j,k}$ is not significant at some predetermined level, but $\hat{\beta}_{k,j}$ is significant, then we reject the null hypothesis that neither model encompasses the other in favour of the alternative hypothesis that Model j encompasses Model k. Conversely, if $\hat{\beta}_{k,j}$ is not significant, but $\hat{\beta}_{j,k}$ is significant, then we say that Model k encompasses Model j. If both $\hat{\beta}_{j,k}$ and $\hat{\beta}_{k,j}$ are significant, or if both $\hat{\beta}_{j,k}$ and $\hat{\beta}_{k,j}$ are not significant, then we fail to reject the null hypothesis that neither model encompass the other. Multicollinearity can lead to both estimated coefficients being insignificant, while sufficiently non-overlapping information sets can lead to both estimated coefficients being significant.

Since the encompassing test has an easily derivable distribution when applied to the out-of-sample data, but not when applied to the in-sample data, we present only out-of-sample encompassing results in Table 8. As before, panels A, B, C and D report the S&P500, NIKKEI, FTSE and TSEC, respectively. The name of the dependent variable from (12) is listed down the left side of the table, while the independent variable is listed along the top. The entries in Table 8 are thus robust p-values on β from Eq. (12) with the left variable regressed on the top variable. P-values less than 0.05 therefore reveal that the forecast from the model listed along the top of the table explains, with 5% significance, the forecast error from the model listed down the left side of the table and thus that the model listed down the side cannot encompass the model listed along the top, at the 5% level. To isolate the effects of the 1987 Crash, we report results from the full 1980–1990 sample as well as the pre-Crash subsample.

---

[17] Another way to think of this test is to consider the regression $\hat{\epsilon}_{t,j}^{2} = \alpha_{j,k} + \beta_{j,k}^{0}\hat{\sigma}_{t,k}^{2} + \beta_{j,k}^{1}\hat{\sigma}_{t,j}^{2} + \nu_{t,j}$ and consider testing the null that $\beta_{j,k}^{0} = 0$, $\beta_{j,k}^{1} = 1$. In this framework one could then add forecasts from other models as additional independent variables and test the significance of one against all the others. Unfortunately, multicollinearity problems make such a test impractical in our application. It should be noted that Chong and Hendry (1986) consider only pairwise comparisons and restrict $\alpha_{j,k} = 0$; i.e., $\hat{\epsilon}_{t,j}^{2} = \beta_{j,k}\hat{\sigma}_{t,k}^{2} + (1 - \beta_{j,k}\hat{\sigma}_{t,j}^{2}) + \nu_{t,j}$ and test the null that $\beta_{j,k} = 0$. We found that all our model's out-of-sample forecast errors were somewhat biased from 0 and hence restricting $\alpha_{j,k} = 0$ led invariably to rejections. While finding all forecast errors to have a non-zero mean is of some interest, we decided to present encompassing tests that reveal failures in modelling movements of the dependent variable, not level effects. We therefore do not restrict $\alpha_{j,k} = 0$ in our tests. One could potentially also consider measures for 'the observed outcome' other than $\hat{\epsilon}_{t,j}^{2}$, as in Lopez (1995).

Table 8
Tests for out-of-sample forecast encompassing robust p-values on $\beta_{j,k}$ from the regression $(\hat{\epsilon}_{t,j}^2 - \hat{\sigma}_{t,j}^2) = \alpha_{j,k} + \beta_{j,k}\sigma_{t,k}^2 + \nu_{t,j}$ [a]

Forecast error $(\hat{\epsilon}_{t,j}^2 - \hat{\sigma}_{t,j}^2)$ from ↓	January 1, 1980 to December 31, 1990				January 1, 1980 to September 30, 1987			
	Forecast $\hat{\sigma}_{t,k}^2$ from ↓				Forecast $\hat{\sigma}_{t,k}^2$ from ↓			
	GARCH	EGARCH	GJR	ANN	GARCH	EGARCH	GJR	ANN
Panel A: S&P500								
GARCH	–	0.888	0.327	0.251	–	0.966	0.388	0.334
EGARCH	0.011	–	0.002	0.002	0.686	–	0.597	0.596
GJR	0.952	0.722	–	0.722	0.495	0.893	–	0.215
ANN	0.956	0.897	0.777	–	0.312	0.763	0.147	–
Panel B: NIKKEI								
GARCH	–	0.001	0.002	0.001	–	0.009	0.195	0.052
EGARCH	0.127	–	0.042	0.021	0.365	–	0.944	0.589
GJR	0.308	0.037	–	0.076	0.562	0.258	–	0.584
ANN	0.236	0.063	0.187	–	0.617	0.547	0.776	–
Panel C: FTSE								
GARCH	–	0.288	0.351	0.610	–	0.015	0.020	0.022
EGARCH	0.626	–	0.719	0.654	0.006	–	0.003	0.006
GJR	0.270	0.229	–	0.369	0.039	0.017	–	0.025
ANN	0.333	0.295	0.264	–	0.112	0.076	0.072	–
Panel D: TSEC								
GARCH	–	0.413	0.205	0.190	–	0.229	0.084	0.069
EGARCH	0.888	–	0.736	0.651	0.144	–	0.128	0.096
GJR	0.277	0.420	–	0.203	0.098	0.232	–	0.070
ANN	0.553	0.711	0.463	–	0.365	0.610	0.331	–

[a] The table reports robust p-values on $\beta_{j,k}$ from the OLS regression: $(\hat{\epsilon}_{t,j}^2 - \hat{\sigma}_{t,j}^2) = \alpha_{j,k} + \beta_{j,k}\hat{\sigma}_{t,k}^2 + \nu_{t,j}$, where $\hat{\sigma}_{t,k}^2$ is model k's one-step-ahead out-of-sample forecasted variance and $(\hat{\epsilon}_{t,j}^2 - \hat{\sigma}_{t,j}^2)$ is model j's one-step-ahead out-of-sample forecast error for New York's S&P500, Tokyo's NIKKEI, London's FTSE and Toronto's TSEC indices on daily data 1980–1990 for the models listed at the bottom of Table 1.

Consider first Panel A; the S&P500. P-values less than 0.05 along the EGARCH row in the 1980–1990 data reveals that the GARCH, GJR and ANN volatility forecasts all explain part of EGARCH's forecast error from 1980–1990. Conversely, the absence of any p-value less than 0.05 in the EGARCH column reveals that the EGARCH volatility forecast cannot explain the forecast errors from GARCH, GJR or ANN. Thus, we conclude that GARCH, GJR and ANN all encompass EGARCH at the 5% level in the S&P500 from 1980–1990. The absence of any p-values less than 0.05 in the GARCH, GJR and ANN rows from 1980–1990 further reveals that none of these models' forecast errors can be explained by other models' forecasts and thus that GARCH, GJR and ANN are all not encompassed in the 1980–1990 data. In other words, in those situations where GARCH, GJR and ANN fail to forecast S&P500 volatility correctly, this failure cannot be accounted for by the other models under consideration. The additional observation that no model encompasses any other in the 1980:01–1987:09 S&P500 subsample – i.e., there are no p-values less than 0.05 in any position 1980:01–1987:09 in Panel A of Table 8 – suggests that EGARCH's encompassing by the other models in the 1980–1990 S&P500 sample is due largely to EGARCH's failure to adequately capture S&P500 volatility effects after the 1987 Crash.

Next consider results for the NIKKEI 1980–1990, reported on the left side of Panel B. The row of p-values less than 0.05 for GARCH reveals that the GARCH forecast error is explained by EGARCH, GJR and ANN. Conversely, the column of p-values greater than 0.05 for GARCH reveals that the GARCH forecast cannot explain any other models' error. Thus, GARCH is encompassed by all other models. (This is, of course, not terribly surprising given the very poor in-sample performance of GARCH relative to other models post 1985 in Fig. 3.) For the NIKKEI 1980–1990 we also see from the EGARCH row that EGARCH's forecast error is explained by both GJR and ANN and, from the EGARCH column, that EGARCH's forecast explains GJR's error but not the error from ANN, at the 5% significance level.

From the EGARCH–GJR comparison in the NIKKEI we therefore learn that GJR and EGARCH each explain a significant portion of the other's forecast error and thus that neither model encompasses the other. From the ANN comparison to EGARCH, as well as to GARCH and GJR, we also learn that the ANN forecast error is the only error that is not explained by any other model, at 5%. ANN therefore encompasses EGARCH, as it did GARCH. In summary, EGARCH, GJR and ANN all encompass GARCH, but only ANN encompasses EGARCH. GJR and ANN are not encompassed in the 1980–1990 NIKKEI at the 5% significance level, though ANN does encompass GJR at 10%. Thus, ANN encompasses all the other models in the NIKKEI at the 1%, 5% or 10% levels. The generally higher p-values in the 1980:01–1987:09 pre-Crash subsample suggests that the inability of GARCH, EGARCH and GJR to fully account for the Crash of '87 in their out-of-sample forecasts is largely to blame for this outcome.

Panel C of Table 8 reports results for FTSE. The absence of any p-values below 0.05 in the 1980–1990 sample reveals that no forecast from any model explains the error from any other model over the entire decade that includes the Crash. However, in the pre-Crash subsample 1980:01–1987:09 ANN again encompasses all the other models at the 5% level of significance. Indeed, ANN is the only model that encompasses any other and is also the only model that is not encompassed. In other words, ANN's one-step-ahead out-of-sample forecasts can explain variance effects not captured by the other models, while none of the traditional ARCH models can significantly account for ANN's volatility forecasting errors. This finding is consistent with results from Fig. 4, which plots the recursively calculated in-sample Sign Test p-values for the various models applied to FTSE data. Note that, from 1980–1987, ANN never fails any of the Sign Tests, while all the other models do fail – at 5% or less – during this period.

Finally, Panel D of Table 8 reports results for TSEC. Over the entire sample 1980–1990 no model encompasses any other. However, from the right half of Panel D we see that ANN again encompasses every other model at the 10% significance level in the pre-Crash subsample 1980:01–1987:09. Together with previously articulated results from Panels A, B and C, Table 8 therefore reveals that ANN often encompasses traditional models in terms of out-of-sample forecasting ability. Furthermore, ANN is the only model that is never encompassed itself. Results from the forecast encompassing tests therefore lead us to conclude that our new ANN model does significantly better than traditional models at capturing the conditional volatility of stock returns. This result is strongest in the out-of-sample encompassing tests, which are arguably the most important of any we have considered.

7. Summary and conclusions

In this paper we have introduced a new nonlinear seminonparametric model for conditional stock volatility and have compared its in- and out-of-sample performance with that of other popular volatility models in four international stock market indices: S&P500, NIKKEI, FTSE and TSEC. In-sample comparisons reveal that GARCH most often fails to capture the empirical regularity that past negative return innovations lead to more volatility than positive return innovations. This is true for all indices studied, though the results of the Engle–Ng Sign Bias Tests plotted in Figs. 2–5 suggest that asymmetric effects are more prominent in the NIKKEI before 1987, and in the S&P500 after 1987, than in either the FTSE or TSEC. In-sample summary statistics from Table 6 further reveal that EGARCH may not be an appropriate model for the NIKKEI since it sometimes produces volatility estimates that exceed the variance of the squared return innovation. In terms of log likelihood (Tables 2–5), ability to remove excess skew and kurtosis (Table 6) and standard in-sample diagnostics, such as the Engle–Ng Sign Tests whose p-values are reported in Figs. 2–5, the GJR model seems more able to fit the asymmetric heteroskedasticity in the data than either GARCH or EGARCH.

The best performing model of all appears to be our new ANN model. From the ANN parameter values reported in Table 5 we see that, in S&P500, TSEC and NIKKEI, the ANN terms capture asymmetric volatility effects in addition to – not instead of – effects captured by GJR. Conversely, in the FTSE, the inclusion of ANN terms removes the necessity for GJR terms. Indeed, from we see some evidence that the highly flexible ANN behaves more like EGARCH in the FTSE and more like GJR in the S&P500. Results from the out-of-sample tests in Table 8 also show that ANN is the only model whose one-step-ahead out-of-sample forecasts encompass forecasts from other models. The statistics in Table 7 confirm that, unlike many seminonparametric forms, the ANN's superior performance in this respect is not the result of overfitting the data. This suggests to us that, at least on the 1969–1990 data, ANN is preferred to more traditional ARCH-type models for its ability to flexibly capture asymmetric and symmetric volatility in a variety of international stock markets.

Finally, a cross-country comparison of our results reveals that there may be important differences between the processes driving returns volatility in the four countries we study. For example, Figs. 2–5 reveal that shocks to asymmetric volatility following the crash of 1987 are difficult to capture in the New York market, but are less problematic in the Tokyo and London indices. From the number of lagged terms in the model specifications in Table 1, it also appears that ARCH effects may be longer lived (i.e. more persistent) in Tokyo than in London, Toronto or New York. Together, these findings suggest the potential usefulness of incorporating other types of information in the volatility forecasting information set. Such an investigation is a subject for ongoing research.

Acknowledgements

We are grateful to three anonymous referees and the journal editor, Franz Palm, for helpful comments. We also thank Burton Hollifield, Peter Kennedy, Lisa Kramer and Nathalie Moyen for insightful suggestions and the Social Sciences and Humanities Research Council of Canada for financial support. The usual disclaimer applies.

References

Akgiray, V., 1989, Conditional heteroskedasticity in time series of stock returns: Evidence and forecasts, Journal of Business 62, 55–80.

Baillie, R.T. and R.J. Myers, 1991, Bivariate GARCH estimation of the optimal commodity futures hedge, Journal of Applied Econometrics 6, 109–124.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307–327.

Bollerslev, T. and J.M. Wooldridge, 1992, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, Econometric Reviews 11, 143–172.

Bollerslev, T., R.Y. Chou and K.F. Kroner, 1992, ARCH modelling in finance: A review of the theory and empirical evidence, Journal of Econometrics 52, 5–59.

Bollerslev, T., R.F. Engle and J.M. Wooldridge, 1988, A capital asset pricing model with time-varying covariances, Journal of Political Economy 96, 116–131.

Campbell, J. and L. Hentschel, 1992, No news is good news: An asymmetric model of changing volatility in stock returns, Journal of Financial Economics 31, 281–318.

Chong, Y.Y. and D.F. Hendry, 1986, Econometric evaluation of linear macroeconomic models, Review of Economic Studies 53, 671–690.

Diebold, F.X. and R. Mariano, 1995, Comparing predictive accuracy, Journal of Business and Economic Statistics 13, 253–264.
Diebold, F.X., S.C. Im and C. Lee, 1993, A note on conditional heteroskedasticity in the market model, Journal of Accounting, Auditing and Finance 8, 141–150.
Donaldson, R.G. and M. Kamstra, 1996a, A new dividend forecasting procedure that rejects bubbles in asset prices, Review of Financial Studies 9, 333–383.
Donaldson, R.G. and M. Kamstra, 1996b, Forecast combining with neural networks, Journal of Forecasting 15, 49–61.
Dutta, S. and S. Shekhar, 1988, Bond-rating: A non-conservative application of neural networks, IEEE International Conference on Neural Networks, 443–450.
Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50, 987–1007.
Engle, R.F. and T. Bollerslev, 1986, Modelling the persistence of conditional variances, Econometric Reviews 5, 1–50.
Engle, R.F. and V.K. Ng, 1993, Measuring and testing the impact of news on volatility, Journal of Finance 48, 1749–1778.
Engle, R.F., C.H. Hong, A. Kane and J. Noh, 1992, Arbitrage valuation of variance forecasts with simulated options markets, Manuscript (UCSD, San Diego, CA).
Ferson, W.E. and C.R. Harvey, 1991, The variation of economic risk premiums, Journal of Political Economy 99, 385–415.
French, K.R., G.W. Schwert and R.F. Stambaugh, 1987, Expected stock returns and volatility, Journal of Financial Economics 19, 3–29.
Glosten, L.R., R. Jagannathan and D.E. Runkle, 1993, The relationship between expected value and the volatility of the nominal excess return on stocks, Journal of Finance 48, 1779–1801.
Hannan, E.J., 1980, The estimation of the order of an ARMA process, The Annals of Statistics 8, 1071–1081.
Hertz, J., A. Krogh and R.G. Palmer, 1991, Introduction to the theory of neural computing (Addison-Wesley, Redwood City, CA).
Hornik, K., 1991, Approximation capabilities of multilayer feedforward networks, Neural Networks 4, 251–257.
Hornik, K., M. Stinchcombe and H. White, 1989, Multilayer feedforward networks are universal approximators, Neural Network 2, 359–366.
Hornik, K., M. Stinchcombe and H. White, 1990, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, Neural Networks 3, 551–560.
Hutchinson, J.M., A.W. Lo and T. Poggio, 1994, A nonparametric approach to pricing and hedging derivative securities via learning networks, Journal of Finance 49, 851–889.
Kimijo, K. and T. Tanigawa, 1990, Stock price pattern recognition: A recurrent neural network approach, Proceedings of the International Joint Conference on Neural Networks, San Diego, CA.
Kroner, K. and J. Sultan, 1993, Time-varying distributions and dynamic hedging with foreign currency futures, Journal of Financial and Quantitative Analysis 28, 535–551.
Kuan, C.M. and H. White, 1994, Artificial neural networks: An econometric perspective, Econometric Reviews 13, 1–91.
Ljung, G.M. and G.E.P. Box, 1978, On a measure of lack of fit in time series models, Biometrika 65, 297–303.
Lopez, J.A., 1995, Evaluating the predictive accuracy of volatility models, Research paper no. 9524 (Federal Reserve Bank of New York, New York).
MacKinlay, A.C. and M.P. Richardson, 1991, Using generalized method of moments to test mean–variance efficiency, Journal of Finance 46, 511–527.
Nelson, D.B., 1991, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59, 347–370.
Pagan, A. and W. Schwert, 1990, Alternative models for conditional stock volatility, Journal of Econometrics 45, 267–290.

Schwert, W.G., 1989, Why does stock market volatility change over time? Journal of Finance 44, 1115–1153.

Stinchcombe, M. and H. White, 1994, Using feedforward networks to distinguish multivariate populations, Proceedings of the International Joint Conference on Neural Networks.

Tam, K.Y. and M.Y. Kiang, 1992, Managerial applications of neural networks: The case of bank failure predictions, Management Science 38, 926–947.

White, H., 1989, Some asymptotic results for learning in single hidden-layer feedforward network models, Journal of the American Statistical Association 84, 1003–1013.

White, H., 1990, Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings, Neural Networks 3, 535–549.