ELSEVIER

Journal of Banking & Finance 20 (1996) 419–438

Journal of BANKING & FINANCE

# An evaluation of volatility forecasting techniques

Timothy J. Brailsford [a,*], Robert W. Faff [b]

[a] Department of Accounting and Finance, University of Melbourne, Parkville 3052, Australia
[b] Department of Accounting and Finance, Monash University, Clayton 3168, Australia

Received 15 March 1992; accepted 15 January 1995

## Abstract

The existing literature contains conflicting evidence regarding the relative quality of stock market volatility forecasts. Evidence can be found supporting the superiority of relatively complex models (including ARCH class models), while there is also evidence supporting the superiority of more simple alternatives. These inconsistencies are of particular concern because of the use of, and reliance on, volatility forecasts in key economic decision-making and analysis, and in asset/option pricing. This paper employs daily Australian data to examine this issue. The results suggest that the ARCH class of models and a simple regression model provide superior forecasts of volatility. However, the various model rankings are shown to be sensitive to the error statistic used to assess the accuracy of the forecasts. Nevertheless, a clear message is that volatility forecasting is a notoriously difficult task.

*JEL classification:* G12; G15

*Keywords:* Forecasting; Stock market volatility; ARCH models; Australia

## 1. Introduction

While traditional financial economics research has tended to focus upon the mean of stock market returns, in recent times the emphasis has shifted to focus

* Tel.: 61-3-9344-7662; fax: 61-3-9344-6681.


*SSDI* 0378-4266(95)00015-1

upon the volatility of these returns. Moreover, the international stock market crash of 1987 has increased the focus of regulators, practitioners and researchers upon volatility. Large swings in price movements have apparently become more prevalent and some observers have blamed institutional changes for this apparent increase in volatility. [1] These concerns have led researchers to examine the level and stationarity of volatility over time. Specifically, research has been directed toward examining the accuracy of volatility forecasts obtained from various econometric models including the autoregressive conditional heteroscedasticity (ARCH) family of models. [2]

Volatility forecasts have many practical applications such as use in the analysis of market timing decisions, aid with portfolio selection and the provision of estimates of variance for use in asset (and option) pricing models. Thus, it follows that it is important to distinguish between various models in order to find the model which provides the most accurate forecasts. This information is clearly of particular value in economic decision-making. To this end, this study compares various volatility forecasting models, including the ARCH class of models, in the Australian stock market.

The investigation of conditional volatility of the US stock market has been extensively undertaken. However, only recently have models of conditional volatility been tested in other stock markets. Specifically, the application of these models has been investigated by de Jong et al. (1992) (Holland), Tse (1991) (Japan), Tse and Tung (1992) (Singapore), Poon and Taylor (1992) (United Kingdom) and Brailsford and Faff (1993) and Kearns and Pagan (1993) (Australia). The evidence in these non-US markets is limited and therefore examining the Australian market, which has different institutional features, provides an important opportunity to add to the accumulated evidence to date. [3]

Brailsford and Faff (1993) using the same Australian data as those employed in the current paper, conducted an extensive model fitting exercise of stock market volatility. On the basis of several model selection techniques, they initially found in favour of the GARCH(3,1) model. However, the results from the asymmetry based diagnostic tests of Engle and Ng (1993) showed that this model was unable to adequately capture asymmetric responses in volatility to past innovations. Consequently, Brailsford and Faff (1993) examined asymmetric volatility models and found support for the Glosten, Jagannathan and Runkle (GJR; 1993) modified

[1] Program trading, the advent of trading strategies such as arbitrage trading, the introduction of futures and options trading and the increased influence of institutional investors, have all been put forward as possible causes. See Schwert (1990) for further discussion.

[2] See Bollerslev et al. (1992) for a comprehensive review of the theory and empirical evidence of modelling ARCH in finance.

[3] Further support is provided by Ang (1991, pp. 201–203) who argues strongly in favour of conducting financial research in Pacific Basin markets. Moreover, Bollerslev et al. (1992, p. 31) stress the potential insights to be gained from examining volatility using international data.

GARCH model. Specifically, they concluded that the GJR-GARCH(3,1) specification was the superior fitting model.

While such model fitting investigations provide useful insights into volatility, the models are usually selected on the basis of full sample information. For practical forecasting purposes, the predictive ability of these models needs to be examined out-of-sample. The few papers that have tested the forecasting ability of ARCH models out-of-sample have reached inconsistent conclusions. Akgiray (1989) found that a GARCH(1,1) specification exhibited superior forecasting ability of monthly US stock market volatility compared to more traditional models. However, Tse (1991) and Tse and Tung (1992) questioned the superiority of the GARCH model in the Japanese and Singaporean markets, respectively. These latter two studies found evidence strongly in favour of an exponentially weighted moving average (EWMA) model. Dimson and Marsh (1990) in an examination of the UK equity market, concluded that the simple models provide more accurate forecasts, and recommended the exponential smoothing and simple regression models. However, Dimson and Marsh did not subject ARCH models to examination. [4] Nevertheless, the conclusions of Dimson and Marsh have important implications for forecasts obtained from the relatively complex GARCH model: "With the increasing interest in using complicated econometric techniques for volatility forecasting, our research strikes a warning bell. For those who are interested in forecasts with reasonable predictive accuracy, the best forecasting models may well be the simplest ones" (Dimson and Marsh (1990, p. 420)).

Using daily Australian data, the aim of this paper is to examine the relative ability of various models to forecast monthly stock market volatility. The forecasting models which are employed range from naive models to the relatively complex ARCH class of models. The various model rankings are shown to be sensitive to the error statistic used to assess the accuracy of the forecasts. Notwithstanding this sensitivity, the results suggest superior forecasts of volatility are provided by the ARCH class of models and a simple regression model. Of further note is the poor performance of some other models which have previously been identified in the literature as good performers.

## 2. The Australian environment [5]

Brailsford and Faff (1993) identify several distinguishing features of the Australian stock market which are worthy of further discussion. Specifically, these

[4] Dimson and Marsh (1990) investigated five volatility forecasting models: (1) a random walk model, (2) a long-term mean, (3) a moving average model, (4) an exponential smoothing model and (5) regression models.

[5] This section extends a similar discussion in Brailsford and Faff (1993, pp. 111–112).

features relate to the size of the market; the relative dominance of the largest stocks; the significance of resource sector stocks; the dominance of institutional investors; the degree of regulation; and the observed empirical regularities.

The Australian stock market comprises 1030 stocks with a total capitalisation of around USD$135 billion (at December 1992). The Australian Stock Exchange represents a small market with a small number of listed companies, of small per unit economic size when compared to the USA. For example, at the end of 1992 the US market had in excess of 7,000 listed stocks with a total market capitalisation of USD$4,758 billion. [6] However, in the context of the Asia Pacific region within which the Australian market operates, with the exception of Japan it assumes a position of relative importance. For example, in comparison the Singapore stock market had 163 listed companies with a total capitalisation of USD$49 billion, in 1992. [7] An alternative measure of market size is the total value of stocks traded in any given year. In 1992, USD$2,679 billion, USD$46 billion and USD$14 billion worth of trading occurred in the US, Australian and Singapore markets, respectively. [8]

The second feature of the Australian market is the dominance of trading activity in a relatively small number of the largest corporations. For example, the largest stock in the Australian market is Broken Hill Proprietary Company Ltd which at any one time constitutes approximately 8 to 10 percent of total market capitalisation. Moreover, the largest 10 stocks represent some 40 percent of the market capitalisation and trading value. [9] In comparison, only 15 percent of the total market capitalisation in the US was concentrated in the ten largest stocks in 1992. [10] Furthermore, many of the smaller Australian companies are typified by prolonged periods of little or even no trading activity (see Hathaway (1986, p. 51)). This market characteristic suggests a need to recognise the potential effects of thin trading in any analysis of data.

The third feature of the Australian stock market is the substantial influence which resource based stocks have upon the market. Australia is a resource rich country and relies heavily upon its rural and mining sectors. Resource stocks comprise about one third of the total market capitalisation. The performance of these stocks is dependent on many highly variable factors including, for example, the variability of commodity prices, which tends to induce greater volatility in price changes compared to industrial companies. Indeed, Ball and Brown (1980) have shown that while Australian resource stocks earned approximately the same mean return as their industrial counterparts over the period 1958 to 1979, the standard deviation of resource sector stocks was almost double that of industrial

[6] Source: International Finance Corporation (1993).
[7] ibid.
[8] ibid.
[9] Source: Australian Stock Exchange, 1993.
[10] Source: International Finance Corporation (1993).

sector stocks. A similar result holds in the later period. [11] To the extent that resource stocks experience more volatile price changes, the Australian market is potentially more volatile than it otherwise would be.

The fourth feature of the Australian stock market is the composition of the investment community in terms of the mix between personal and institutional investors. The Australian market is typified by a relatively low incidence of direct personal investment, particularly since the occurrence of the Crash of October 1987. For example, only 10 percent of adult Australians were direct share owners at December 1991. Furthermore, 94 percent of the top 330 companies have at least 50 percent of their stock held by no more than 20 investors. [12]

The fifth feature of the Australian stock market is the degree of regulation. During the 1980s the Australian economy operated under a regime of continuing financial market deregulation. For example, in late 1983 the Australian dollar was floated, while in late 1984 foreign banks were granted licences to operate in Australia. Notwithstanding the significant progress toward general financial market deregulation over the past decade, it is true that some specific regulatory restrictions are more prevalent in Australia than in the USA. For example, the ability of investors to short sell securities in Australia is far more restrictive than in the USA. Indeed short selling was outlawed in Australia over the period 1971 to 1986. Since 1986 only restrictive short selling opportunities have been available. This constraint on trading could suggest, all other things equal, that the volatility of stock returns may be higher in Australia. For example, there is empirical evidence that the standard deviation of the Australian market is 30 percent larger than the standard deviation of the US market, (see Jaffe and Westerfield (1985)).

Arguably, the Australian financial system represents one of the more deregulated markets of the developed countries in the Asia-Pacific region. This is particularly the case during the last decade or so for the reasons discussed in the preceding paragraph. In contrast, the governments of countries such as Japan and Singapore are well known for their stringent market controls.

The final distinguishing feature of the Australian stock market is several observed empirical regularities which differ from many other stock markets. For example, the Australian market does not only exhibit a strong January seasonal as in the USA, but exhibits January, July and August seasonals (see Brown et al. (1983)). Of greater relevance to the current study is the finding of a significantly negative average Tuesday return in the Australian market. A similar Tuesday effect has also been documented in Japan (Kato, 1990), in Singapore (Condoyanni et al., 1987) and some other South-East Asian markets (Wong et al., 1992). This

[11] Over the period 1980 to 1993, the resource sector earned an average monthly return of 0.62 percent with a standard deviation of 8.70 percent while the industrial sector earned 1.48 percent with a standard deviation of only 6.05 percent.

[12] Source: Australian Stock Exchange, 1991.

Tuesday effect is in contrast to the widely observed Monday effect in many other markets, including the USA. [13] Consequently, any analysis of Australian data using the daily measurement interval needs to consider the impact of such empirical regularities.

## 3. Data

In order to investigate aggregate stock market volatility, a market index is required. The index used in our study is the Statex-Actuaries Accumulation Index. This index comprises the 50 most actively traded companies listed on the Australian Stock Exchange. Further, the index is an accumulation index and therefore yields a measure of total return. Our sample consists of over 4900 observations encompassing the period from 1 January 1974 to 30 June 1993. [14]

A preliminary investigation of the data reveals that the series is skewed, leptokurtic and exhibits a high degree of autocorrelation in both levels and squares. Specifically, the skewness estimate is $-4.81$, [15] the excess kurtosis estimate is 139.54, the $Q(20)$ statistic is 336.50 and the $Q^2(20)$ statistic is 137.26. [16]

To ensure that the volatility models use only data contained in the current information set, the raw monthly volatility series (which is used in forecasting), is defined as the sum of squared daily returns, viz: [17]

$$\sigma_T^2 = \sum_{t=1}^{N_T} r_t^2 \tag{1}$$

where $r_t$ is the daily rate of return and $N_T$ is the number of trading days in month $T$.

[13] For example, see Jaffe and Westerfield (1985) and Jaffe et al. (1989).

[14] Further details on the data are provided in Brailsford and Faff (1993).

[15] When data from October 1987 crash are excluded the skewness estimate falls to 0.11.

[16] The $Q(20)$ and $Q^2(20)$ statistics are from the Box–Pierce–Ljung test for first to twentieth-order autocorrelation in the returns and squared returns, respectively. The test statistics are approximate chi-squares with 20 degrees of freedom.

[17] As the true underlying monthly volatility series is unobservable, the entire analysis was also conducted on other volatility series and very similar results were obtained. In particular, these alternative series were: (1) the daily returns were adjusted for the within-month mean return, and (2) a series based on the daily return data adjusted for the effect of non-synchronous trading through the addition of cross-product terms (see French et al. (1987, p. 5)). Furthermore, as partial autocorrelation estimates for the daily rate of return indicated significant autocorrelation, a moving average process (MA(1)) was used to take account of the autocorrelation which may be induced by non-synchronous trading in the daily return series. The empirical regularities of the day of the week and holiday effects were also accounted for in a return model by running an OLS regression of returns on independent dichotomous dummy variables for days of the week and before and after holidays.

## 4. Volatility forecasting models

The focus of this paper is on the forecasting accuracy of monthly stock market volatility from various statistical models. The basic methodology involves the estimation of the various models' parameters using an initial set of data and the application of these parameters to later data, thus forming out-of-sample forecasts. Following Akgiray (1989), Dimson and Marsh (1990), [18] Tse (1991) and Tse and Tung (1992), several forecasting models of volatility are investigated. These are (1) a random walk model, (2) an historical mean model, (3) a moving average model, (4) an exponential smoothing model, (5) an exponentially weighted moving average model, (6) a simple regression model, (7) two standard GARCH models and (8) two Glosten–Jagannathan–Runkle (GJR) asymmetric GARCH models. [19]

For all forecasting approaches, the initial data used for the estimation of the models' parameters are drawn from the period 1974 to 1985 (months $T = 1,2,\ldots,144$). Thus, the first month for which out-of-sample forecasts are obtained is January 1986 ($T = 145$). As the sample period covers 234 months, out-of-sample forecasts are constructed for months 145 to 234. Note that this period includes 1987 which requires the models to predict volatility in a period when actual volatility was extremely high.

### 4.1. Random walk model

Under a random walk model, the best forecast of this month's volatility is last month's observed volatility.

$$\hat{\sigma}_T^2(\mathrm{RW}) = \sigma_{T-1}^2 \quad T = 145, 146,\ldots, 234. \tag{2}$$

where $\sigma_T^2$ is the monthly volatility measure defined in expression (1).

### 4.2. Historical mean model

Under the assumption of a stationary mean, the best forecast of this month's volatility is a long-term average of past observed volatilities.

$$\hat{\sigma}_T^2(\mathrm{LTM}) = \frac{1}{T-1} \sum_{j=1}^{T-1} \sigma_j^2 \quad T = 145, 146,\ldots, 234. \tag{3}$$

[18] Note that Dimson and Marsh (1990) examine models of standard deviation.

[19] An alternative approach would entail calculating implied standard deviations from option prices (see Day and Lewis (1992)). However, such an approach requires an active options market on the index which unfortunately, does not exist in Australia.

### 4.3. Moving average models

A moving average is often used by market analysts as a predictor of mean returns. Further, this technique is often used in traditional time-series analysis. Thus, a moving average model is employed. The choice of the moving average estimation period is arbitrary. In this paper, five (mid-term) and twelve years (long-term) are chosen as estimation periods to ensure consistency with the estimation period of later models which require large samples for estimation. The twelve-year moving average model can be expressed as:

$$\hat{\sigma}_T^2(\mathrm{MA}) = \frac{1}{144}\sum_{j=1}^{144}\sigma_{T-j}^2 \qquad T = 145, 146, \ldots, 234. \tag{4}$$

A similar formulation is used for the five-year moving average model where the summation is conducted over 60 months.

### 4.4. Exponential smoothing model

Following Dimson and Marsh (1990), an exponential smoothing model is used to forecast volatility. In this model, the forecast of volatility is posited to be a function of the immediate past forecast and the immediate past observed volatility.

$$\hat{\sigma}_T^2(\mathrm{ES}) = \phi\hat{\sigma}_{T-1}^2(\mathrm{ES}) + (1-\phi)\sigma_{T-1}^2 \qquad T = 145, 146, \ldots, 234. \tag{5}$$

The smoothing parameter ($\phi$) is constrained to lie between zero and one. The optimal value of $\phi$ must be determined empirically. If $\phi$ is zero, then the exponential smoothing model collapses to the random walk model. As $\phi$ approaches unity, the major weight is given to the prior period forecast which itself is heavily influenced by its immediate past forecast and so on. When forecasting, the optimal value of $\phi$ is chosen by a search of values between zero and one using data from $T = 1, 2, \ldots, 144$ and selecting the value of $\phi$ which corresponds to the minimum prediction error. [20] Three error metrics were used to assess the minimum prediction error, namely the root mean squared error, the mean absolute error and the mean absolute percentage error. Based on the similarity of selected values of $\phi$ generated by the three error metrics, the root mean squared error metric is employed for the selection of $\phi$. The smoothing parameter selection procedure is performed annually, whereby the estimation period is updated, thus resulting in a rolling 12-year estimation window. [21]

[20] The initial exponential smoothing forecast is taken to be the estimated volatility over the first month of the estimation period.

[21] Dimson and Marsh (1990) in their study of UK stock market volatility found that the optimal value of $\phi$ was 0.76. In this study, the value of $\phi$ ranges between 0.51 and 0.98.

### 4.5. Exponentially weighted moving average model

Following the work of Tse (1991) and Tse and Tung (1992), an exponentially weighted moving average model (EWMA) is examined. This model is similar to the exponential smoothing model except that the past observed volatility in expression (5) is replaced by the five-year moving average forecast which can be formally expressed as: [22]

$$\hat{\sigma}_T^2(\text{EWMA}) = \psi\hat{\sigma}_{T-1}^2(\text{EWMA}) + (1-\psi)\hat{\sigma}_T^2(\text{MA})$$
$$T = 145, 146,..., 234. \quad (6)$$

Again, the selection of the smoothing parameter ($\psi$) value is an empirical issue. The optimal value of $\psi$ is chosen by a search of values between zero and one using pre-sample data and selecting the value of $\psi$ which corresponds to the minimum prediction error. Similar to the exponential smoothing model, the selected values of $\psi$ do not vary considerably across error metrics and therefore, the root mean squared error is employed. The parameter selection procedure is also updated annually, employing a rolling 12-year window. [23]

### 4.6. Simple regression model

This model employs an OLS regression of observed volatilities on immediate past observed volatility, and the resulting volatility forecasts are given by:

$$\hat{\sigma}_T^2(\text{SR}) = \hat{\gamma}_0 + \hat{\gamma}_1\sigma_{T-1}^2 \quad T = 145, 146,..., 234. \quad (7)$$

The model parameters are initially estimated over the 12-year period (from 1974 to 1985). These parameter estimates are subsequently used to forecast the volatility of January 1986 ($T = 145$). Two approaches are then used to obtain the new parameter estimates. Under the first approach, the parameter estimation window is anchored at January 1974 ($T = 1$), while the end of the window is continually updated as the most recent month's data become available. The second approach uses a constant window length of 12 years. Under this latter approach, the window is rolled forward one month at a time as the more recent data become available resulting in a 12-year rolling estimation period. As the results from the two estimation procedures are very similar, only the results using the estimation period of a 12-year rolling window are reported.

[22] A twelve-year moving average forecast was also used in the EWMA model in expression (6). These results do not substantially differ from those of the EWMA model forecasts reported in the text.

[23] Tse (1991) found the optimal value of $\psi$ to be 0.86 in his study on the Japanese market, while Tse and Tung (1992) do not report the optimal value of $\psi$ in their study on the Singapore market. In this study, the value of $\psi$ ranges between 0.0 (following October 1987) and 0.9.

### 4.7. Standard GARCH models

The GARCH model involves the joint estimation of a conditional mean and a conditional variance equation. As the GARCH(1,1) model has generally been found to be the most appropriate of the standard ARCH family of models for stock return data, [24] this model is employed, viz: [25]

$$r_t = \gamma + \epsilon_t \tag{8}$$

where $\epsilon_t \sim N(0,h_t)$ and

$$h_t = \omega + \beta_1 h_{t-1} + \alpha_1 \epsilon_{t-1}^2. \tag{9}$$

The conditional variance equation (expression (9)) models the time varying nature of the volatility of the errors derived from the conditional mean equation. [26] An initial test for ARCH errors using Engle's LM procedure (Engle, 1982) yields a test statistic significant at the 0.001 level.

Following Engle and Bollerslev (1986) a daily $s$-step ahead forecast can be formed based on the GARCH(1,1) model as follows:

$$\hat{h}_{t+s}(G) = \hat{\omega} \sum_{i=0}^{s-2} \left(\hat{\alpha}_1 + \hat{\beta}_1\right)^i + \left(\hat{\alpha}_1 + \hat{\beta}_1\right)^{s-1} \hat{h}_{t+1} \quad s = 1, 2, \ldots, N_T. \tag{10}$$

where $\hat{h}_{t+1}$ is the one-day ahead volatility forecast for the first day of each month generated by the empirical counterpart of expression (9).

Monthly volatility forecasts are then formed by aggregating the s-step ahead daily forecasts across trading days in each month as follows:

$$\hat{\sigma}_T^2(G) = \sum_{s=1}^{N_T} \hat{h}_{t+s}(G) \tag{11}$$

where a given month ($T$) has $N_T$ daily observations.

The monthly GARCH(1,1) volatility forecasts therefore become: [27]

$$\hat{\sigma}_T^2(G) = \hat{\omega} \sum_{s=1}^{N_T} \sum_{i=0}^{s-2} \left(\hat{\alpha}_1 + \hat{\beta}_1\right)^i + \sum_{s=1}^{N_T} \left(\hat{\alpha}_1 + \hat{\beta}_1\right)^{s-1} \hat{h}_{t+1} \tag{12}$$

$T = 145, 146, \ldots, 234.$

[24] This conclusion is consistent with the majority of research in this area such as Akgiray (1989), Baillie and DeGennaro (1990), Lamoureux and Lastrapes (1990) and Schwert and Seguin (1990).

[25] The conditional mean specification was also used taking account of predictable influences which had been identified as the effects of non-synchronous trading, days of the week and holidays. These results do not substantially differ from those reported in the text.

[26] In this paper, all GARCH models are estimated using maximum likelihood techniques and the Berndt et al. (1974) algorithm employing numerical derivatives and the assumption of conditional normality.

[27] Alternatively, monthly returns could be used in the GARCH model to obtain one-step ahead forecasts of monthly volatility. However, there are no ARCH effects in the monthly return series (LM test is insignificant at the 0.001 level). Hence, such a procedure is not used in this paper.

Initially the GARCH model is estimated over the 12-year period from 1974 to 1985. The parameter estimates of $\omega$, $\alpha_1$ and $\beta_1$ are then used in (10) to obtain daily s-step ahead forecasts for each month. These volatility forecasts initially cover the trading days in January 1986 ($T = 145$) and their summation yields the monthly forecast for that month according to (11). The start and end dates of the parameter estimation period are then rolled forward one calendar month and the model parameters are re-estimated. These new estimates are used to forecast daily $s$-step ahead volatilities for each trading day over the next month (February 1986 or $T = 146$) and their summation yields that month's volatility forecast. This procedure is repeated, rolling forward the estimation window one calendar month at a time until the forecast for the final calendar month (June 1993 or $T = 234$) is obtained.

In a general investigation of fitting conditional volatility models in the Australian stock market, Brailsford and Faff (1993), using the same data as the current paper, considered higher order ARCH models, but concluded that the GARCH(1,1) specification was preferred. However, when higher order GARCH($p,q$) models were examined, it was found that a GARCH(3,1) specification was preferred to the GARCH(1,1) counterpart. Hence, the forecasting ability of the GARCH(3,1) model is also examined here. The conditional variance equation becomes: [28]

$$h_t = \omega + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \beta_3 h_{t-3} + \alpha_1 \epsilon_{t-1}^2. \tag{13}$$

A similar forecasting procedure to the GARCH(1,1) model is used to obtain daily volatility forecasts for the GARCH(3,1) model which are then summed over trading days in each month to obtain monthly volatility forecasts. Again, a rolling 12-year window is used as the estimation period commencing with the period 1974 to 1985, and forecasts are obtained for months $T = 145, 146, \ldots, 234$.

### 4.8. GJR-GARCH models

The standard GARCH model is symmetric in its response to past innovations. However, there are theoretical arguments which suggest a differential response in conditional variance to past positive and negative innovations. The two main arguments are related to corporate leverage and information arrival. [29] Several alternative GARCH model specifications have been proposed in an attempt to capture the asymmetric nature of volatility responses. Engle and Ng (1993) in a test of volatility models on Japanese stock return data find strong support for the GJR-GARCH (Glosten et al., 1993) model which explicitly incorporates the

[28] The conditional mean equation of the GARCH(3,1) model remains the same as Eq. (8).

[29] Refer to Black (1976), Christie (1982) and Nelson (1991) for a discussion of the effects of corporate leverage on volatility, and Campbell and Hentschel (1992) for a discussion of the relationship between information arrival and volatility.

potential for asymmetry in the conditional variance equation. Similarly, Brailsford and Faff (1993) also find the GJR-GARCH models to be preferable to standard GARCH models using Australian stock return data.

Specifically, the GJR-GARCH model augments the conditional variance equation in the GARCH($p,q$) model with a variable equal to the product of $S_t^-$ and $\epsilon_{t-1}^2$, where $S_t^-$ is a dichotomous dummy variable that takes the value of unity if $\epsilon_{t-1}$ is negative and zero otherwise. In the case of the GJR-GARCH(1,1) version of the model, the conditional variance equation becomes: [30]

$$h_t = \omega + \beta_1 h_{t-1} + \alpha_1 \epsilon_{t-1}^2 + \gamma S_t^- \epsilon_{t-1}^2. \tag{14}$$

Brailsford and Faff (1993) also examined the fit of higher order GJR-GARCH($p,q$) specifications. They found that the GJR-GARCH(3,1) model was preferred to all other models investigated. Hence, the forecasting ability of this model is also examined and the conditional variance equation becomes: [31]

$$h_t = \omega + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \beta_3 h_{t-3} + \alpha_1 \epsilon_{t-1}^2 + \gamma S_t^- \epsilon_{t-1}^2 \tag{15}$$

A similar forecasting procedure to the standard GARCH models in Section 4.7 is used to obtain daily volatility forecasts for both GJR-GARCH specifications which are then summed over trading days in each month to obtain monthly volatility forecasts. Again, a rolling 12-year window is used as the estimation period commencing with the period 1974 to 1985, and forecasts are obtained for months $T = 145, 146, \ldots, 234$.

## 5. Out-of-sample model forecast results

### 5.1. Definition of the forecast error statistics

Previous papers have used a variety of statistics to evaluate and compare forecast errors. [32] Consistent with this research, the 90 monthly forecast errors generated from each model in this study are compared by the mean error (ME), the mean absolute error (MAE), the root mean squared error (RMSE) and the mean absolute percentage error (MAPE) which are defined as follows:

$$\text{ME} = \frac{1}{90} \sum_{T=1}^{90} \left( \hat{\sigma}_T^2 - \sigma_T^2 \right) \tag{16}$$

$$\text{MAE} = \frac{1}{90} \sum_{T=1}^{90} \left| \hat{\sigma}_T^2 - \sigma_T^2 \right| \tag{17}$$

[30] The conditional mean equation of the GJR-GARCH(1,1) model remains the same as Eq. (8).

[31] The conditional mean equation of the GJR-GARCH(3,1) model remains the same as Eq. (8).

[32] See Akgiray (1989), Dimson and Marsh (1990), Tse (1991) and Tse and Tung (1992).

Table 1
Error statistics from forecasting monthly volatility

| | ME | MAE | | RMSE | | MAPE | |
|---|---|---|---|---|---|---|---|
| | Actual | Actual | Relative | Actual | Relative | Actual | Relative |
| Random walk | −0.00001 | 0.00427 | 0.951 | 0.01870 | 1.000 | 1.06022 | 0.449 |
| Historical mean | −0.00102 | 0.00318 | 0.708 | 0.01441 | 0.771 | 1.39229 | 0.589 |
| Moving average (5-year) | −0.00000 | 0.00405 | 0.902 | 0.01455 | 0.778 | 2.36392 | 1.000 |
| Moving average (12-year) | −0.00101 | 0.00327 | 0.728 | 0.01446 | 0.773 | 1.47063 | 0.622 |
| Exponential smoothing | 0.00094 | 0.00449 | 1.000 | 0.01477 | 0.790 | 2.29353 | 0.970 |
| EWMA | −0.00086 | 0.00361 | 0.804 | 0.01453 | 0.777 | 1.74513 | 0.738 |
| Simple regression | −0.00107 | 0.00315 | 0.702 | 0.01441 | 0.771 | 1.37149 | 0.580 |
| GARCH(1,1) | −0.00181 | 0.00324 | 0.722 | 0.01542 | 0.825 | 0.57398 | 0.243 |
| GARCH(3,1) | −0.00082 | 0.00317 | 0.706 | 0.01537 | 0.822 | 0.86086 | 0.364 |
| GJR-GARCH(1,1) | −0.00255 | 0.00292 | 0.650 | 0.01449 | 0.775 | 0.56895 | 0.241 |
| GJR-GARCH(3,1) | −0.00108 | 0.00310 | 0.690 | 0.01527 | 0.817 | 0.76393 | 0.323 |

Calculated values are provided for four different error statistics across eleven models used to forecast monthly volatility. ME is a mean error statistic defined by expression (16); MAE is a mean absolute error statistic defined by expression (17); RMSE is a root mean squared error statistic defined by expression (18), and MAPE is a mean absolute percentage error statistic defined by expression (19). The error statistics are applied to forecasts obtained over the period January 1986 to June 1993. The relative error statistics are obtained by expressing the actual statistic as a ratio relative to the worst performing model for a given error measure.

$$\mathrm{RMSE} = \sqrt{\frac{1}{90}\sum_{T=1}^{90}\left(\hat{\sigma}_T^2 - \sigma_T^2\right)^2} \tag{18}$$

$$\mathrm{MAPE} = \frac{1}{90}\sum_{T=1}^{90}\left|\left(\hat{\sigma}_T^2 - \sigma_T^2\right)/\sigma_T^2\right|. \tag{19}$$

Dimson and Marsh (1990) further standardise each error statistic by the value of the error statistic obtained from the random walk forecast. The advantage of such a procedure is that the statistics can be more easily interpreted relative to a benchmark forecast. In this study, each error statistic is also expressed on a relative basis where the benchmark is the value of the statistic for the worst performing model.

### 5.2. Forecast results

Table 1 presents the actual and relative forecast error statistics for each model across the four error measures. An examination of Table 1 reveals that no single model is clearly superior. The ME does not allow for the offsetting effect of errors of different signs and as such, little credence should be placed upon it. However, the mean error can be used as a general guide as to the direction of over/under-prediction. All models are found to under-predict volatility, with the exception of the exponential smoothing model.

The MAE statistic indicates that the GJR-GARCH(1,1) model provides the most accurate forecasts. This forecast model is 35 percent more accurate than the benchmark model which, for this error statistic, is the exponential smoothing model. The GJR-GARCH(3,1) model ranks a close second. However, the MAE statistic does not allow for a clear distinction between the higher ranking models which is evidenced by the marginal (2.6 percent) difference in relative accuracy which separates the next five ranked models.

The RMSE statistic equally favours the historical mean and simple regression models which are 23 percent more accurate than the benchmark model which is the random walk model. The exponential smoothing model now ranks seventh. The GJR-GARCH(1,1) model now ranks fourth, but it is only slightly less accurate than the best model and is still 22.5 percent more accurate than the benchmark model. Furthermore, while the GJR-GARCH(3,1) model now ranks eighth, it too is not substantially worse than the best ranked model. Indeed, it is difficult to distinguish between the higher ranking models as evidenced by the top six ranking models which are separated by only 0.7 percent in forecasting accuracy.

The MAPE statistic gives a relative indication of overall forecasting performance. The GJR-GARCH(1,1) model has the best (actual) MAPE of 56.9 percent. The standard GARCH(1,1) model ranks second and is only marginally less accurate. Notably, the four GARCH models rank as the top four models. While the MAPE estimates for all four GARCH models may be considered as high in absolute terms, the corresponding estimates for the other models are all substantially higher. The worst performing model, as assessed by the MAPE statistic, is the moving average (5-year) model. The best GARCH class models are 76 percent more accurate than the benchmark model while the random walk model which is the next ranking non-GARCH model, is only 55 percent more accurate than the benchmark. Thus, the MAPE statistic clearly identifies the GARCH class models as superior.

In summary, the ranking of any one forecasting model varies depending upon the choice of error statistic. This sensitivity in rankings highlights the potential hazard of selecting the best model on the basis of an arbitrarily chosen error statistic. Notably, the RMSE statistic is equivalent to the primary error measure used by Dimson and Marsh (1990) to reach their conclusion that the simple regression model is superior. [33] Consistent with this finding, our results also indicate that the simple regression model is ranked (equal) first by the RMSE statistic. Furthermore, Dimson and Marsh find that the superiority of the simple

[33] Dimson and Marsh (1990) use the MSE (mean squared error) statistic.

regression model is insensitive to the use of the MAE statistic which is again generally consistent with our MAE results. However, while Dimson and Marsh find an equivalent ranking across all models between their error statistics, our model rankings, while similar, are not entirely robust between the RMSE and MAE statistics. This inconsistency in rankings is exacerbated further when other error statistics, such as the MAPE statistic, are considered. For example, the random walk model ranks second last and last for the MAE and RMSE metrics respectively, yet ranks fifth (only behind the GARCH class models) for the MAPE metric. Hence, the results of Dimson and Marsh may not be robust to other error statistics, contrary to their claim (see Dimson and Marsh (1990, pp. 416–418)). Nevertheless, if the GARCH class model results are excluded from Table 1, the simple regression model ranks first for the MAE and RMSE metrics and second for the MAPE metric, generally consistent with the results of Dimson and Marsh (1990).

Also of note from Table 1 is the consistently poor performance of the EWMA model which ranks eighth, fifth and ninth out of the eleven models under the MAE, RMSE and MAPE statistics, respectively. This evidence is in direct conflict to the results of Tse (1991) and Tse and Tung (1992) which show the EWMA to be clearly superior under the same error metrics. [34] In conclusion, while it is difficult to claim superiority of any one model, the GJR-GARCH(1,1) model is our choice.

### 5.3. Asymmetric loss functions

The error statistics reported in Table 1 assume that the underlying loss function is symmetric. From a practical viewpoint, it is conceivable that many investors will not attribute equal importance to both over- and under-predictions of volatility of similar magnitude. For example, consider the positive relationship between the volatility of underlying stock prices and call option prices. An under-prediction of stock price volatility will lead to a downward biased estimate of the call option price. This under-estimate of the price is more likely to be of greater concern to a seller than a buyer. The reverse is true of over-predictions of stock price volatility. In the spirit of Pagan and Schwert (1990), [35] to account for the potential

[34] However, it should be noted that in highly regulated markets, such as Japan and Singapore, that the objectives of the regulators is often to dampen volatility. Hence, forecasting models which use a smoothing parameter may be expected to perform better in such markets than in less regulated markets. This may explain the inconsistency between our results and those of Tse (1991) and Tse and Tung (1992).

[35] The restrictions of the standard assumptions on loss functions has been recognised and relaxed in previous work, such as Pagan and Schwert (1990, p. 280) who employ a proportional loss function measure as opposed to a quadratic loss function.

asymmetry in the loss function, we construct an error statistic which penalises under-predictions more heavily and is called the mean mixed error (MME($U$)): [36]

$$\mathrm{MME}(U)=\frac{1}{90}\left[\sum_{T=1}^{O}|\hat{\sigma}_T^2-\sigma_T^2|+\sum_{T=1}^{U}\sqrt{|\hat{\sigma}_T^2-\sigma_T^2|}\right] \tag{20}$$

where $O$ is the number of over-predictions and $U$ is the number of under-predictions.

Similarly, the above statistic can be redefined so to weight over-predictions more heavily and is constructed as follows:

$$\mathrm{MME}(O)=\frac{1}{90}\left[\sum_{T=1}^{O}\sqrt{|\hat{\sigma}_T^2-\sigma_T^2|}+\sum_{T=1}^{U}|\hat{\sigma}_T^2-\sigma_T^2|\right]. \tag{21}$$

A 'biased' forecast model can be viewed as one which systematically over- or under-predicts, whereas an 'unbiased' forecast model should, when not providing a perfect forecast, over-predict 50 percent of the time and under-predict 50 percent of the time. Table 2 presents the MME statistics and the number of times that the models over- and under-predict. From the number of over- and under-predictions reported in Table 2, only the random walk and the GJR-GARCH(3,1) models can be claimed to provide 'unbiased' forecasts. These models provide the only cases in which the null hypothesis of an equal number of over- and under-predictions can be accepted at standard significance levels. With the exception of the GJR-GARCH(1,1) model, the remaining models all systematically over-predict volatility. However, it is notable that systematic over-prediction of the other models is a result of decreasing average volatility in the out-of-sample test period compared to the estimation period. This feature arises because of the Crash of October 1987. [37] No doubt a different period could result in systematic under-prediction.

Recall that in Table 1, according to the ME statistic, with the exception of the exponential smoothing model, all models were found to under-predict volatility. This apparently contrasts with the results presented in Table 2 which indicate that all models except the random walk and GJR-GARCH(1,1) models over-predict volatility. The paradox is explained as the ME statistic takes into account the *magnitude* of forecasting errors, whereas Table 2 presents the *number* of over/under-predictions and ignores the magnitude of forecast errors. Hence, most of the models over-predict volatility more frequently but with relatively small

[36] As the absolute values of all forecast errors are less than unity, taking their square root will place a heavier weighting on the under-predictions. If the absolute value of all forecast errors were greater than unity, the MME($U$) would need to square the errors in order to achieve the desired penalty.

[37] October 1987 is the 22nd month in the 90 month out-of-sample forecasting period. The surge in volatility over this month creates a dramatic increase in average volatility which subsequently declines from then on as volatility never reaches the same level again.

Table 2
Mean mixed error statistics from forecasting monthly volatility and results of under- and over-prediction

| | MME($U$) | | MME($O$) | | Under-predictions | Over-predictions | Binomial prob. |
|---|---|---|---|---|---|---|---|
| | Actual | Relative | Actual | Relative | | | |
| Random walk | 0.02137 | 0.659 | 0.02284 | 0.516 | 45 | 45 | 0.458 |
| Historical mean | 0.01482 | 0.457 | 0.02944 | 0.665 | 23 | 67 | 0.000 |
| Moving average (5-year) | 0.01484 | 0.458 | 0.03965 | 0.895 | 21 | 69 | 0.000 |
| Moving average (12-year) | 0.01558 | 0.481 | 0.02979 | 0.673 | 24 | 66 | 0.000 |
| Exponential smoothing | 0.01281 | 0.395 | 0.04429 | 1.000 | 17 | 73 | 0.000 |
| EWMA | 0.01759 | 0.543 | 0.03241 | 0.732 | 26 | 64 | 0.000 |
| Simple regression | 0.01445 | 0.446 | 0.02773 | 0.626 | 24 | 66 | 0.000 |
| GARCH(1,1) | 0.01067 | 0.329 | 0.02847 | 0.643 | 27 | 63 | 0.000 |
| GARCH(3,1) | 0.01669 | 0.515 | 0.02202 | 0.497 | 34 | 56 | 0.007 |
| GJR-GARCH(1,1) | 0.03241 | 1.000 | 0.00654 | 0.148 | 76 | 14 | 0.000 |
| GJR-GARCH(3,1) | 0.01875 | 0.579 | 0.01922 | 0.434 | 39 | 51 | 0.085 |

Calculated values are provided for two different error statistics across eleven models used to forecast monthly volatility. MME($U$) is a mean mixed error which penalises under-predictions more heavily and is defined by expression (20). MME($O$) is a mean mixed error which penalises over-predictions more heavily and is defined by expression (21). These statistics are designed to capture potential asymmetry in the loss function. The error statistics are applied to forecasts obtained over the period January 1986 to June 1993. The number of under- and over-predictions are provided for each set of forecasts. The associated binomial probability is based on the test that the number of under-predictions and over-predictions are equal.

errors. In contrast, when the models under-predict volatility, they do so by a relatively large amount. Also note that in Table 1, the GJR-GARCH(1,1) model had the largest negative ME measure and it is the only model which under-predicts more frequently in Table 2. Conversely, the exponential smoothing model is the only model which had a positive ME measure in Table 1 and it has the highest frequency of over-predictions in Table 2.

The MME($U$) statistics in Table 2 indicate superiority of the standard GARCH(1,1) model, while the GJR-GARCH(1,1) and random walk models provide the worst forecasts. This is not surprising as these models provide the greatest number of under-predictions which receive a heavier penalty than over-predictions in the computation of the MME(U) statistic. Also of note is the relatively good performance of models previously identified (in Table 1) as poor performers. For example, the exponential smoothing model was ranked last, seventh and tenth by the MAE, RMSE and MAPE statistics respectively, but is the second best forecasting model according to the MME($U$) statistic.

The MME($O$) statistic which penalises over-prediction errors more heavily, ranks the GJR-GARCH(1,1) model first while the exponential smoothing model is now ranked last.[38] This contrast in model rankings again illustrates that the forecasts are highly sensitive to the assessment criteria. Hence, caution should be exercised in the interpretation of the obtained rankings. Moreover, the choice of error measure should reflect an appropriate underlying loss function which in turn depends on the ultimate purpose of the forecasting procedure. For example, again using the call option context, a buyer of a call option being more concerned with over-predictions, would prefer the MME(O) statistic and hence would favour the GJR-GARCH model.

## 6. Summary

This paper has examined the ability of various models to forecast aggregate monthly stock market volatility in Australia. The models which were tested included a random walk model, an historical mean model, a moving average model, an exponential smoothing model, an exponentially weighted moving average (EWMA) model, a simple regression model, two standard GARCH models and two asymmetric GJR-GARCH models. Contrary to prior evidence, the results in this paper suggest that no single model is clearly superior. The rankings of the various model forecasts are sensitive to the choice of error statistic.

In addition to the traditional error measures, a mean mixed error statistic was constructed to allow for the asymmetry in the loss function of investors, and this measure also provided a conflict in model rankings. Given the conflicting nature of the results, we urge caution in the interpretation of various error measures and recommend that use be determined according to the purpose for which the forecasts are provided.

Nevertheless, consistent with the findings of Dimson and Marsh (1990), we find some support for a simple regression model. However, when ARCH class models are considered, they are found to be at least equal, if not superior to the simple regression model. In particular, the evidence favours the GJR-GARCH(1,1) specification.

## Acknowledgements

We wish to thank Richard Baillie, Fischer Black, Tim Bollerslev, Jon Kendall, Keith McLaren, Alan Ramsay, anonymous referees of this journal, G.P. Szego (the

[38] Note that while the GJR-GARCH(1,1) and exponential smoothing models appear to almost reverse their extreme ranks, there is no consistent negative correlation in the comparative rankings of other models resulting from the MME($U$) and MME($O$) statistics. For example, the EWMA model ranks eighth and ninth under the respective MME($U$) and MME($O$) statistics.

editor), seminar participants at the Australian National University, Monash University, the Universities of Melbourne, New South Wales, Queensland, Southern Queensland and Tasmania, and participants at the Fourth Australasian Finance and Banking Conference (Sydney), the Third Asia Pacific Finance Conference (Singapore) and the 1993 Annual Conference of the Accounting Association of Australia and New Zealand (Darwin) for helpful comments on earlier versions of this paper. The first author gratefully acknowledges the financial assistance of a Coopers and Lybrand Travel Grant.

## References

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

Ang, J.S., 1991, Agenda for research in pacific-basin finance, in: S.G. Rhee and R.P. Chang, eds., Pacific-Basin capital markets research, Volume II (Elsevier Science, North-Holland), 201–210.

Australian Stock Exchange, 1991, Australian share ownership.

Australian Stock Exchange, 1993, Fact book.

Baillie, R.T. and R.P. DeGennaro, 1990, Stock returns and volatility, Journal of Financial and Quantitative Analysis 25, 203–214.

Ball, R. and P. Brown, 1980, Risk and return from equity investments in the Australian mining industry: January 1958 to February 1979, Australian Journal of Management 5, 45–66.

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

Black, F., 1976, Studies of stock price volatility changes, American Statistical Association – 1976, Proceedings of the Business and Economic Section, Boston, 177–181.

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, 61–90.

Brailsford, T.J. and R.W. Faff, 1993, Modelling Australian stock market volatility, Australian Journal of Management 18, 109–132.

Brown, P., D. Keim, A. Kleidon and T. Marsh, 1983, Stock return seasonalities and the tax-loss selling hypothesis: Analysis of the arguments and Australian evidence, Journal of Financial Economics 12, 105–127.

Campbell, J.Y. 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.

Christie, A.A., 1982, The stochastic behaviour of common stock variances: Value, leverage and interest rate effects, Journal of Financial Economics 10, 407–432.

Condoyanni, L., J. O'Hanlon and C.W.R. Ward, 1987, Day of the week effects on stock returns: International evidence, Journal of Business Finance and Accounting 14, 159–174.

Day, T.E. and C.M. Lewis, 1992, Stock market volatility and the information content of stock index options, Journal of Econometrics 52, 267–287.

de Jong, F., A. Kemna and T. Kloek, 1992, A contribution to event study methodology with an application to the Dutch stock market, Journal of Banking and Finance 16, 11–36.

Dimson, E. and P. Marsh, 1990, Volatility forecasting without data-snooping, Journal of Banking and Finance 14, 399–421.

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 Ng, V.K., 1993, Measuring and testing the impact of news on volatility, Journal of Finance 48, 1749–1778.

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, On the relation between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance 48, 1779–1801.

Hathaway, N., 1986, The non-stationarity of share price volatility, Accounting and Finance 26, 35–54.

International Finance Corporation, 1993, Emerging stock markets factbook (IFC, Washington DC).

Jaffe, J. and R. Westerfield, 1985, The week-end effect in common stock returns: The international evidence, Journal of Finance 41, 433–454.

Jaffe, J., R. Westerfield and C. Ma, 1989, A twist on the Monday effect in stock prices: Evidence from the US and foreign stock markets, Journal of Banking and Finance 13, 641–650.

Kato, K., 1990, Weekly patterns in Japanese stock returns, Management Science 36, 1031–1043.

Kearns, P. and A.R. Pagan, 1993, Australian stock market volatility: 1875–1987, The Economic Record 69, 163–178.

Lamoureux, C.G. and W.D. Lastrapes, 1990, Persistence in variance, structural change, and the GARCH model, Journal of Business and Economic Statistics 8, 225–234.

Nelson, D.B., 1991, Conditional heteroscedasticity in asset returns: A new approach, Econometrica 59, 347–370.

Pagan, A.R. and G.W. Schwert, 1990, Alternative models for conditional stock volatility, Journal of Econometrics 45, 267–290.

Poon, S-H. and S.J. Taylor, 1992, Stock returns and volatility: An empirical study of the UK stock market, Journal of Banking and Finance 16, 37–59.

Schwert, G.W., 1990, Stock market volatility, Financial Analysts Journal 46, 23–34.

Schwert, G.W. and P.J. Seguin, 1990, Heteroskedasticity in stock returns, Journal of Finance 45, 1129–1155.

Tse, Y.K., 1991, Stock return volatility in the Tokyo stock exchange, Japan and the World Economy 3, 285–298.

Tse, Y.K. and S.H. Tung, 1992, Forecasting volatility in the Singapore stock market, Asia Pacific Journal of Management 9, 1–13.

Wong, K.A., T.K. Hui and C.Y. Chan, 1992, Day of the week effects: Evidence from developing stock markets, Applied Financial Economics 2, 49–56.