Predicting stock index volatility: can market volume help?

Article

Accepted Version

Predicting Stock Index Volatility: Can Market Volume Help?

Chris Brooks
ISMA Centre, Department of Economics, University of Reading
PO Box 218, Whiteknights, Reading, RG6 6AA.

Tel: (+44) 118 931 67 68 (direct)
Fax: (+44) 118 931 47 41
E-mail: C.Brooks@reading.ac.uk

This Draft. 19 / 4 / 97

Keywords: Stock index volatility, volume, prediction, time series models, GARCH

Author's biography:

Chris Brooks is a Lecturer in Finance at the ISMA Centre, University of Reading, where he also obtained his PhD. His research interests are in the field of financial econometrics, particularly in time series modelling and forecasting.

Predicting Stock Index Volatility: Can Market Volume Help?

Abstract

This paper explores a number of statistical models for predicting the daily stock return volatility of an aggregate of all stocks traded on the NYSE. An application of linear and non-linear Granger causality tests highlights evidence of bi-directional causality, although the relationship is stronger from volatility to volume than the other way around. The out-of-sample forecasting performance of various linear, GARCH, EGARCH, GJR and neural network models of volatility are evaluated and compared. The models are also augmented by the addition of a measure of lagged volume to form more general ex ante forecasting models. The results indicate that augmenting models of volatility with measures of lagged volume leads only to very modest improvements, if any, in forecasting performance.

Introduction

Modelling and forecasting stock market volatility has been the subject of much recent empirical and theoretical investigation by academics and practitioners alike. There are a number of motivations for this line of inquiry. First, volatility, as measured by the standard deviation or variance of returns, is often used as a crude measure of the total risk of financial assets. Second, the volatility of stock market prices enters directly into the Black-Scholes formula for deriving the prices of traded options. Although in the past, historical measures of volatility have been used as estimates of future volatility, there is a growing body of evidence that suggests that the use of volatility predicted from more sophisticated time series models will lead to more accurate option valuations (see, for example, Akgiray, 1989; or Chu and Freund, 1996). Finally, using combinations of options, it is possible to trade volatility as if it were any other commodity, so that accurate predictions of future volatility give the forecaster the potential to make a more direct profit.

The vast majority of recent papers which attempt to forecast volatility out-of-sample have been entirely univariate in nature, using past realisations of volatility to predict its future path. Akgiray (1989), for example, finds the GARCH model superior to ARCH, exponentially weighted moving average, and historical mean models for forecasting monthly US stock index volatility. A similar result concerning the apparent superiority of GARCH is observed by West and Cho (1995) using one-step ahead forecasts of Dollar exchange rate volatility, although for longer horizons, the model behaves no better than their alternatives[i]. Pagan and Schwert (1990) compare GARCH, EGARCH, Markov switching regime and three non-parametric models for forecasting monthly US stock return volatilities. The EGARCH followed by the GARCH models perform moderately; the remaining models produce very poor predictions. Franses and van Dijk (1996) compare three members of the GARCH family (standard GARCH, QGARCH and the GJR model) for forecasting the weekly volatility of various European stock market indices. They find that the non-linear GARCH models were unable to beat the standard GARCH model. Finally, Brailsford and Faff (1996) find GJR and GARCH models slightly superior to various simpler models[ii] for predicting Australian monthly stock index volatility. The conclusion arising from this growing body of research is that forecasting volatility is a "notoriously difficult task" (Brailsfrod and Faff, 1996, p419), although it appears that conditional heteroscedasticity models are among the best that are currently available. In particular, more complex non-linear and non-parametric models are inferior in prediction to simpler models, a result echoed in an earlier paper by Dimson and Marsh (1990) in the context of relatively complex versus parsimonious linear models.

At the same time, a parallel but almost entirely separate literature has developed which seeks to consider the relationship between price and volume or between volatility and volume. Research by Hiemstra and Jones (1994) and Campbell, Grossman and Wang (1993) fall into the former category, although the latter relationship is the one of interest here. An early systematic survey of the relationship between stock market volume and volatility is undertaken by Karpoff (1987), who argues that it is likely there exists a positive contemporaneous correlation between the absolute price and volume measures; this conclusion is supported by evidence from a large number of previous studies using a variety of data sets of different frequencies and originating from different markets. None of these studies, however, appears to be concerned with whether one series leads or causes the other.

There have also been a number of theoretical studies into the relationship between stock market volume and return volatility. Although current asset pricing models do not have a place for volume data (see Ross, 1987), and researchers are still uncertain as to the precise role of volume in the analysis of financial markets as a whole, volume data may contain information useful for modelling other financial quantities of more direct interest, such as volatility or the returns themselves. Among the theoretical models of volume and volatility, the mixture of distributions hypothesis (MDH) suggests that the two quantities should be positively correlated as a consequence of their joint dependence on a common underlying (latent) "mixing" or directing variable, the rate of information flow (see Epps and Epps, 1975; Harris, 1984). Epps (1975) model proposes a model where volume tends to be higher when stock prices are rising than falling, although there is no strong reason why the relationship should be contemporaneous rather than lead-lag, since volume may react more quickly to changes in the directing variable than volatility, or vice versa. Another class of models, associated primarily with Copeland (1976), and developed by Smirlock and Starks (1984) are known as "sequential information" models. These models posit that new information flowing into the market is disseminated to investors one at a time, which creates a series of intermediate equilibria prior to the creation of a complete equilibrium. This sequential information flow could imply bi-directional causality or a positive contemporaneous relationship between volume and volatility. These theoretical models have paved the way and provided motivation for for empirical research into this relationship.

A further empirical advancement was made by Lamoureux and Lastrapes (1990). A number of previous studies had shown that GARCH models were useful descriptions of stock return volatility, and that shocks to the conditional variance showed a high degree of persistence (Engle and Bollerslev, 1986). Lamoureux and Lastrapes show, however, that this persistence becomes much less pronounced if the variance equation is augmented by the addition of a contemporaneous term in volume. They argue that if volume is viewed as a proxy for the rate of information flow to the stock market, then once this is taken into account, lagged squared residuals add little extra information about stock return volatilities. However, a major weakness of this empirical methodology (which they acknowledge in a footnote) is the possibility of simultaneity between volume and volatility. If stock volume, which is used as an exogenous RHS variable in the variance equation of the GARCH model, is part of a larger system of equations where volume is itself partly determined by volatility, failure to appropriately model the system as such will cause a simultaneity bias in the coefficient estimates. One potential solution to this problem is to use lagged measures of volume, which will be pre-determined and therefore not subject to the simultaneity problem. A necessary condition for this "quick fix" to work is that lagged volume is an appropriate instrument for contemporaneous volume, and the results on this question are mixed. Lamoureux and Lastrapes (1990) find lagged volume to be a poor instrument, while Najand and Yung (1991) find it to be quite acceptable in an analysis of price variability in Chicago Board of Trade futures data. In the latter paper, however, the effect on the persistence of shocks to the conditional variance equation of including either contemporaneous or lagged volume is minimal in both cases. These empirical findings are given theoretical explanation by the modified MDH model of Andersen (1996), which demonstrates the kind of reduction in volatility persistence observed by Lamoureux and Lastrapes.

A common factor running through all of the papers cited above is that they investigate the relationship between volume and volatility, but with the exception of Weigend and LeBaron (1994), none has attempted to use the information contained in one series to forecast the other. A natural and obvious extension to the existing literature is therefore to consider whether stock market volume has any predictive power for stock index volatility, and this is the subject of the present paper. To this end, section 3 describes a methodology for testing for both linear and non-linear Granger causality in time series data. To anticipate the findings of this part of the paper, displayed in section 4, the tests show some evidence of bi-directional feedback

between volatility and volume, although more evidence that changes in volume bring about changes in volatility.

Armed with this evidence, section 5 describes a number of forecasting models which have recently been used for forecasting volatility, and considers their augmentation to allow for lagged values of market volume as predictors of future volatility. Section 6 describes the results of this forecasting exercise, while section 7 concludes.

2. The Data

The data used in this study comprises 2431 daily observations on New York Stock Exchange (NYSE) aggregate volume together with daily observations on the Dow Jones composite, running from 17 November 1978 to 30 June 1988. The data is taken from the data set used in Weigend and LeBaron (1994). Stock returns are calculated as the first difference of the logarithm of the level of the index, whilst the daily volatility measure used here is simply the square of the day's return. This is a measure commonly used (e.g. Chan, Christie and Schultz, 1995; Day and Lewis, 1992; West and Cho, 1995), although others such as the difference between the highest and lowest daily prices are possible. The latter is a more efficient volatility estimator in terms of approximating the true (unknown) diffusion constant (Parkinson, 1980), particularly for small samples, although it is subject to more biases, for example due to the closure of the stock exchange over night (Garman and Klass, 1980).

The measure of volume used here as a starting point is the proportion of shares traded each day (i.e. the total daily aggregate market turnover divided by the total number of shares outstanding). A fundamental problem with this measure is, however, that the series is non-stationary (see Gallant *et al*., 1992) . Stationarity is therefore induced by detrending the series by dividing by a 100-day moving average of turnover and then taking the natural logarithm. This is the measure of volume used by Weigend and LeBaron (1994) and Campbell *et al*. (1995), and is the measure used for analysis in this study.

3A. Linear Granger Causality

The linear Granger causality test is usually constructed in the context of a reduced-form bivariate vector autoregression (VAR). Let $\{X_t\}$ and $\{Y_t\}$ be two stationary time series of length T,

$$X_t = A(L)X_t + B(L)Y_t + \varepsilon_{X,t}$$
$$Y_t = C(L)X_t + D(L)Y_t + \varepsilon_{Y,t} \qquad t = 1,2,\ldots,T \qquad (1)$$

where $A(L)$, $B(L)$, $C(L)$, and $D(L)$ are all polynomials in the lag operator with all roots outside the unit circle. The error terms, $\varepsilon_{X,t}$ and $\varepsilon_{Y,t}$ are separate i.i.d. processes with zero mean and constant variance. The test of whether *Y* strictly Granger causes *X* is simply a test of the joint restriction that all of the coefficients contained in the lag polynomial $B(L)$ are zero. Similarly, a test of whether *X* strictly Granger causes *Y* is a test of the restriction that all of the coefficients contained in the lag polynomial $C(L)$ are jointly zero. In each case, the null hypothesis of no Granger causality is rejected if the exclusion restriction is rejected. Bi-directional feedback, or causality running in both directions, exists if the elements in both the lag polynomials ($B(L)$ and $C(L)$) are jointly significantly different from zero.

3B. Non-Linear Granger Causality

Baek and Brock (1992) recently proposed a test for multivariate non-linearity, which can be considered an extension of the BDS test (Brock *et al*., 1996), which uses the concept of the correlation integral based upon the closeness of points in hyperspace. If rejections are restricted to one tail of the distribution, a rearrangement of the test can be viewed as a non-linear test for Granger causality. This is an important generalisation, for there is no reason why causality should be of the linear type, and it is likely that linear Granger causality tests will have low power against many types of non-linear causality (Brock, 1991). A further modification which improves the small sample properties of the test and relaxes the assumption that the series to which the test is applied are i.i.d. is due to Hiemstra and Jones (1994). In particular, the original Baek and Brock test is over-sized. Results of Monte Carlo simulations in Hiemstra and Jones (1993) also show that their modified test is robust to the presence of structural breaks in the series and contemporaneous correlations in the errors of the VAR model used to filter out linear cross- and auto-dependence.

Using the notation above (closely following Hiemstra and Jones, 1994), and additionally, letting X_t^m, $X_{t-Lx}^{Lx}, Y_{t-Ly}^{Ly}$ denote a lead vector for *X* of length *m*, and lag vectors for *X* and *Y* of length *Lx* and *Ly* respectively, i.e.

$X_t^m = (X_t, X_{t+1}, \ldots, X_{t+m-1})$, $\quad m = 1,2,\ldots \quad t = 1,2,\ldots$

$X_{t-Lx}^{Lx} = (X_{t-Lx}, X_{t-Lx+1}, \ldots, X_{t-1})$, $\quad Lx = 1,2,\ldots \quad t = Lx+1, Lx+2, \ldots$

$Y_{t-Ly}^{Ly} = (Y_{t-Ly}, Y_{t-Ly+1}, \ldots, Y_{t-1}),$ $\quad Ly = 1,2, \ldots \quad t = Ly + 1, Ly + 2, \ldots$

Then for given values of *m*, *Lx*, and *Ly* all ≥ 1 and $e > 0$, if Y does not strictly Granger cause X, then we can write

$$Pr\left(\left\|X_t^m - X_s^m\right\| < e \middle| \left\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\right\| < e, \left\|Y_{t-Ly}^{Ly} - Y_{s-Ly}^{Ly}\right\| < e\right) = Pr\left(\left\|X_t^m - X_s^m\right\| < e \middle| \left\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\right\| < e\right) \tag{2}$$

where Pr(•) denotes a probability measure and $\|\bullet\|$ denotes a distance measure (in this case the supremum norm). Hence non-Granger causality implies that the probability that two arbitrary lead vectors of length *m* are within a distance *e* of each other is the same conditional upon the two lag vectors of $\{X_t\}$ being within a distance *e* of each other and the two lag vectors of $\{Y_t\}$ being within a distance *e* of each other, and conditional upon the lag vectors of $\{X_t\}$ only being within a distance *e* of each other. In other words, no Granger causality means that the probability that the lead vectors are within distance *e* is the same whether we have information about the distances between the $\{Y_t\}$ lag vectors or not.

The conditional probabilities in (2) can be expressed as ratios of a joint and marginal probabilities. Thus

$$\frac{\Pr\left(\left\|X_t^m - X_s^m\right\| < e, \left\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\right\| < e, \left\|Y_{t-Ly}^{Ly} - Y_{s-Ly}^{Ly}\right\| > e\right)}{\Pr\left(\left\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\right\| < e, \left\|Y_{t-Ly}^{Ly} - Y_{s-Ly}^{Ly}\right\| > e\right)} = \frac{\Pr\left(\left\|X_t^m - X_s^m\right\| < e, \left\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\right\| < e\right)}{\Pr\left(\left\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\right\| < e\right)} \tag{3}$$

Whether the probability in (3) holds can evaluated by calculating a set of four correlation integrals:

$$\frac{C1(m+Lx, Ly, e)}{C2(Lx, Ly, e)} = \frac{C3(m+Lx, e)}{C4(Lx, e)} \tag{4}$$

Letting $\{x_t\}$ and $\{y_t\}$ denote the actual realisations of the process and *I*(*A*,*B*,*e*) denoting an indicator function which takes the value 1 if the vectors *A* and *B* are within a distance *e* of each other and zero otherwise and noting that the properties of the supremum norm allow us to write $\Pr\left(\left\|X_t^m - X_s^m\right\| < e, \left\|X_{t-Lx}^{Lx} - X_{s-Lx}^{Lx}\right\| < e\right)$ as $\Pr\left(\left\|X_{t-Lx}^{m+Lx} - X_{s-Lx}^{m+Lx}\right\| < e\right)$, then the estimates of the correlation integrals in (4) can be expressed as

$$C1(m+Lx, Ly, e, n) \equiv \frac{2}{n(n-1)} \sum_{t<s} \sum I(x_{t-Lx}^{m+Lx}, x_{s-Lx}^{m+Lx}, e) \cdot I(y_{t-Ly}^{Ly}, y_{s-Ly}^{Ly}, e) \tag{5}$$

$$C2(Lx, Ly, e, n) \equiv \frac{2}{n(n-1)} \sum_{t<s} \sum I(x_{t-Lx}^{Lx}, x_{s-Lx}^{Lx}, e) \cdot I(y_{t-Ly}^{Ly}, y_{s-Ly}^{Ly}, e) \tag{6}$$

$$C3(m+Lx,e,n) \equiv \frac{2}{n(n-1)} \sum\sum I(x_{t-Lx}^{m+Lx}, x_{s-Lx}^{m+Lx}, e) \quad (7)$$

$$C4(Lx,e,n) \equiv \frac{2}{n(n-1)} \sum\sum I(x_{t-Lx}^{Lx}, x_{s-Lx}^{Lx}, e) \quad (8)$$

for t,s = max(Lx,Ly)+1, ... , T-m+1, $n = T+1-m$-max(Lx,Ly).

Under the null hypothesis that $\{Y_t\}$ does not Granger cause $\{X_t\}$, then Hiemstra and Jones (1994, appendix) show that the test statistic $\sqrt{n}\left(\frac{C1(m+Lx,Ly,e,n)}{C2(Lx,Ly,e,n)} - \frac{C3(m+Lx,e,n)}{C4(Lx,e,n)}\right)$ is asymptotically distributed as a normal variate with mean zero and variance that is a complicated function of (m,Lx, Ly, e, n).

4. Granger Causality Results

A. Linear Granger Causality

Bivariate VAR models with up to 4 lags of each series are estimated and the joint coefficient restrictions representing the linear Granger causality test are calculated and presented in table 1 together with marginal significance levels in parentheses. The results clearly show the existence of a bi-directional feedback relationship between the two series, although causality from volatility to volume is the greater of the two, and this result is also more robust to increases in the numbers of lags used in the VAR.

The first panel of table 1 shows the linear Granger causality test results for the cases where the same number of lags are used in all components of the VAR model for comparison with the nonlinear Granger causality results and also to avoid any pre-test problem associated with the choice of lag lengths prior to joint hypothesis testing. The null hypothesis that volatility does not Granger cause volume is rejected at the 1% level for all the VAR structures considered. The opposite null hypothesis (that volume does not Granger cause volatility) is rejected at the 1% level when only one lag of each variable is used in each equation of the VAR, but the marginal significance level is increased to nearly 10% when three lags are included, and becomes insignificant, even at the 10% level, when 4 lags are added.

Panels B and C of table 1 show the linear Granger causality test results for numbers of lags of each variable chosen by minimising Akaike's or Schwarz's Bayesian information criterion respectively for each equation

of the VAR. The joint hypothesis tests again show strong evidence of rejection (at the 5% level or better) for both sets of null hypothesis, although once more causality is stronger from volatility to volume than vice-versa.

B. Non-linear Granger Causality

The results of applying the Hiemstra and Jones procedure to the residuals of a VAR(4,4) model[iii] are presented in table 2. The value of *e* is set to values of 0.5, 1, and 1.5 times the standard deviation of the data, and *m* is set to 1 as suggested by the Monte Carlo experiments of Hiemstra and Jones (1993). The picture is similar for that of the linear causality test: that is, there is extensive evidence of bi-directional feedback between volume and volatility, but with the line of causation running from volatility to volume being the stronger of the two. In contrast to the results of linear the linear causality tests however, this result is not qualitatively altered by changes in the length of the lag-vectors used in estimation of the test statistic.

5. Forecasting Volatility

The previous section has shown linear and non-linear Granger causality runs between volume and volatility in both directions. This finding has two important implications for forecasting volatility. First, there is evidence of linear causality running from volume to volatility, which suggests that it may be possible to use lagged values of volume to predict volatility. Second, there exists incremental non-linear causality once linear cross- and auto-dependence have been removed from the series. This suggests that the set of forecasting models containing lagged volume measures should not be limited to the linear class. The relationship is stronger from volatility to volume than the other way around, but it is not at all clear how one might proceed to turn this relationship into any useful forecasting model since it is not possible to trade volume as it is volatility. Hence, although an investigation of the forecastability of measures of stock market volume is an interesting exercise for the researcher interested in learning more about the volume series for its own sake, or as a route to understanding more about how financial markets operate more generally (as Weigend and LeBaron suggest), the remainder of this paper examines the effects of using lagged volume as a predictor for volatility. A number of the papers surveyed in section 1 above have attempted to forecast volatility, but all have been essentially univariate in nature. Various forecasting models used in recent studies are brought together and presented here, and their forecasting properties are evaluated and compared with similar models that are augmented by the addition of lagged measures of volume to the forecasting equations.

Although the models do not have any rigorous theoretical motivation, they could represent reduced-forms of a number of theoretical models incorporating volume.

5A. Construction of Forecasts and Notation

The total sample of 2,431 observations is split into two parts: the first 2000 observations (approximately 8 years of daily trading data) are used for estimation of the parameters of the model, and then a one step ahead forecast is calculated. The sample is then rolled forward by removing the first observation of the sample and adding one to the end, and another one step ahead forecast of the next day's volatility is made. This "recursive" modelling and forecasting procedure is repeated until a forecast for observation 2431 has been made using data available at time 2430. Thus the initial in-sample modelling period for generation of the first forecast runs from 17 November, 1978 until 16 October, 1986. Computation of forecasts using a rolling window of data should ensure that the forecasts are made using models whose parameters have been estimated using all of the information available at that time, while not incorporating old data that is probably no longer relevant in the context of a dynamic, rapidly evolving financial market. The procedure used in this study closely follows that of earlier studies (e.g. Akgiray, 1989; Pagan and Schwert, 1990; Brailsford and Faff, 1996), although the present study differs from its predecessors in an important respect: the latter all compute monthly volatility estimates by taking some measure of the added daily volatilities, while this study uses the daily data directly. This should have two important advantages: first, it is likely to be daily forecasts which are of more use to traders or market-makers who have shorter time-horizons, and second, the use of monthly forecasts implies that there will be few observations to compare with actual values (as few as 24 in Akgiray), leading to the possibility that a small number of extreme errors for a given model will have a profound influence on the accuracy measure.

5B. Forecasting Models

1. The random walk model

The simplest possible forecast is one computed using a naive or random walk model, which formulates tomorrow's forecast of volatility as being equal to the realised value today. Letting $\sigma_{f,t}^2 \left| \Omega_{t-1} \right.$ denote the one step ahead forecast for the variance conditional upon information available at time *t-1*, where *t* always runs from observation 2001 to observation 2431, then

$$\sigma_{f,t}^2 = \sigma_{t-1}^2 \tag{9}$$

2. The long term mean

Assuming that the conditional mean is constant, the best forecast of future volatility would be an average of past volatilities:

$$\sigma_{f,t}^2 = \frac{1}{2000} \sum_{j=t-2000}^{t-1} \sigma_{t-j}^2 \tag{10}$$

3. Moving Average Models

Since the forecasts here are computed using a rolling window, the long term mean is, in a sense, a very long term moving average. For comparison, two shorter moving averages, of length 5 days and 100 days (equivalent to one trading week and nearly four trading months respectively) are also estimated:

$$\sigma_{f,t}^2 = \frac{1}{L} \sum_{j=1}^{L} \sigma_{t-j}^2 \tag{11}$$

where L is the length of the moving average (5 or 100 observations).

4. Exponential Smoothing

Exponential smoothing models volatility as being a weighted average of the previous forecast and realised values of volatility:

$$\sigma_{f,t}^2 = \alpha \sigma_{f,t-1}^2 + (1-\alpha)\sigma_{t-1}^2 \tag{12}$$

The value of the smoothing parameter, α is chosen to produce the best fit by minimising the sum of the squared in-sample forecast errors.

5. Exponentially Weighted Moving Average

The exponentially weighted moving average model is essentially an exponential smoothing model in the moving average rather than the previous realised value of volatility. The smoothing constant is estimated in the same way as for the exponential smoothing parameter, and L is chosen as for equation (11) above.

$$\sigma_{f,t}^2 = \beta \sigma_{f,t-1}^2 + (1-\beta) \frac{1}{L} \sum_{j=1}^{L} \sigma_{t-j}^2 \tag{13}$$

Equivalently, the model (13) can be written to make the exponential weighting more transparent

$$\sigma_{f,t}^2 = (1-\beta)\sum_{j=1}^{L}\beta^{j-1}\sigma_{t-j}^2 \qquad (14)$$

6. Autoregressive Volatility Models

This is a simple OLS regression of the current realised value of volatility on *p* own lags and either a constant term or five daily dummy variables which take the value one if the volatility estimate is for that day, and zero otherwise.

$$\sigma_{f,t}^2 = \beta_0 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2 + \sum_{i=1}^{5}\gamma_i D_{i,t} + \varepsilon_t \qquad i = 1,2,3,4,5 \text{ , and either } \beta_0 = 0 \text{ or } D_i = 0 \;\forall\; i. \qquad (15)$$

This model therefore considers whether any predictability can be derived from well-documented day-of-the-week anomalies or recent past volatility realisations. Values of p = 1,3 and 10 are utilised which covers many models used in other research, although this choice is somewhat arbitrary. It is not feasible to use any objective criteria (e.g. Akaike's or Schwarz's information criteria) to determine the appropriate lag-length which optimises the in-sample fit, since the data, and probably the "best" model with it, are changing as the observation window moves through the data. In any case, it is not clear that the use of these metrics to optimise in-sample model adequacy is conducive to maximising out-of-sample forecast accuracy.

7. Symmetric GARCH Models

Unlike the models considered thus far, the GARCH family of models entails a joint estimation of the conditional mean and conditional variance equations. The model, due to Bollerslev (1986) is formulated as

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim N(0,\sigma_t^2) \qquad (16)$$

and

$$\sigma_{f,t}^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \qquad (17)$$

Since both variables on the RHS of the variance equation (17) are known at time *t*, then a one step ahead conditional forecast can be made by simply iterating through the model without the need for successive substitutions or complex iterations of the conditional expectations operator.

8. Asymmetric GARCH Models

A major criticism of (17) as it stands is that positive and negative innovations have an identical effect upon the conditional variance since their sign becomes lost upon taking the square. There is a body of evidence that suggests that this restriction is not empirically valid; in other words, it has been noted that often negative shocks to the conditional mean equation have a larger effect upon volatility than positive shocks[iv]. Two models which remove the assumption of symmetric responses of volatility to shocks of different sign are the EGARCH model due to Nelson (1991), and the GJR model due to Glosten, Jaganathan and Runkle (1993). Under these formulations, the conditional variance equations (17) become

$$\sigma_{f,t}^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma S_t^- \varepsilon_{t-1}^2 \tag{18}$$

$$\log(\sigma_{f,t}^2) = \omega + \beta \log(\sigma_{t-1}^2) + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] \tag{19}$$

for the GJR-GARCH and EGARCH models respectively. In (18) the asymmetry arises from the inclusion of a dummy variable, S_t^-, which takes the value one when $\varepsilon_{t-1} \leq 0$ and zero otherwise, while in (19) the asymmetry arises from the direct inclusion of the term in ε_{t-1}, normalised by the standard deviation of the data. The latter model also has the advantage that no non-negativity constraints are required of the coefficients as they are for the other forms of GARCH model outlined here, since even negative parameter values would not cause the variance itself ($\sigma_{f,t}^2$) to be negative.

9. A Neural Network

Artificial neural networks (ANN's) are a class of nonlinear regression models inspired by the way computation is performed by the brain. Their primary advantage over more conventional econometric techniques lies in their ability to model complex, possibly nonlinear processes without assuming any prior knowledge about the underlying data generating process. The fully flexible functional form makes them particularly suited to a financial application where nonlinear patterns are often deemed to be present, but an adequate structural model is conspicuously absent.

The success of neural networks in point prediction of financial variables (i.e. pattern recognition in time series) is by no means clear. Kuan and Lim (1994), for example, forecast five daily Dollar-denominated

exchange rates based on around five years of daily data. Using feed-forward and recurrent networks, they find the network models generally perform significantly better than ARMA models, although there were considerable differences between the series. In some cases, the ARMA model performed particularly poorly, and was unable to predict even 50% of sign changes. Haefke and Helmenstein (1994) find that an ANN can out-perform an AR(2) in predicting the Austrian stock index. Tsibouris and Zeidenberg (1995) find up to 60% correct sign predictions for four U.S. stocks using a neural network with nine inputs and five hidden layers. Weigend and LeBaron (1994), however, find no significant improvement over linear predictors. They argue that they "...can be fairly confident that this is a fairly general result for the time series and models we considered...there is probably little hope of fine tuning the networks we used."

By far the most popular type of model, and the one studied here, is known as a single hidden layer feed-forward neural network. The model can be specified as follows; the structure consists of three layers: the inputs (akin to regressors in a linear regression model), which are connected to the output(s) (the regressand) via a hidden or intermediate layer. From an econometric perspective, the problem reduces to one of estimating the synaptic weights or connection strengths between the layers. Formally the network model can be written

$$\hat{x}_{N,m}(X;\beta,w,b) = \sum_{j=1}^{N} \beta_j \phi(\sum_{i=1}^{m} w_{ij} Z_i + b_j) \tag{20}$$

where the number of hidden units in the intermediate layer is *N*. The inputs were selected as own lagged values of the series from *t-1* to *t-m*, where *m* is the number of inputs. $\hat{x}$ is a vector of fitted values, *Z* is the input, β represents the hidden to output weights, and w and b represent the input to hidden weights. Let

$$x_t^m = (x_{t+m-1}, x_{t+m-2}, \ldots, x_t) \tag{21}$$

The multivariate nonlinear least squares minimisation problem is then given by

$$\min_{\beta,w,b} \sum_{t=0}^{T-m-1} [x_{t+m} - \hat{x}_{N,m}(x_t^m;\beta,w,b)]^2 \tag{22}$$

Nonlinear least squares (NLS) estimates are computed using an application of the Levenberg-Marquardt algorithm (see Marquardt, 1963). The activation function for the hidden layer is the sigmoid

$$\phi(p) = \frac{1}{1+\exp(-p)} \tag{23}$$

The number of inputs was fixed at one (a single lag of volatility) or two (one lag of volatility and one of volume), and the number of hidden units in the single hidden layer at 8. Hornik *et al.* (1989) have shown that a neural network model with one hidden layer and a sufficient number of hidden nodes can approximate any continuous function to an arbitrary degree of accuracy. Hence it is unlikely that any additional hidden layer would add to predictive power, and is likely to represent an over-parameterisation[v].

10. Augmentation of Forecasting Models using Lagged Volume

The autoregressive volatility, GARCH, EGARCH, and GJR equations can all easily be augmented by the addition of lags of market volume as predictor variables; $\delta_1 V_{t-1}$ or $\delta_1 V_{t-1} + \delta_2 V_{t-2}$ are added to the RHS of equations (15), (17)-(20). Only up to two lags are considered since it is likely that these will have the largest effect upon the current value of volatility (and these lags are the most important in causing rejections in the Granger causality tests).

5C. Forecast Evaluation

Three criteria are used here to evaluate the accuracy of the forecasts: mean squared error (MSE), mean absolute error (MAE), and proportion of over-predictions. Mean squared error provides a quadratic loss function:

$$MSE = \frac{1}{T_f}\sum_{t=1}^{T_f}(\sigma_t^2 - \sigma_{f,t}^2)^2 \qquad (24)$$

where T_f is the total number of forecasts computed (in this case, 431), σ_t^2 and $\sigma_{f,t}^2$ represent the realised volatility at time *t* and the model's forecast of it respectively. Mean absolute error is calculated as

$$MSE = \frac{1}{T_f}\sum_{t=1}^{T_f}\left|\sigma_t^2 - \sigma_{f,t}^2\right| \qquad (25)$$

with notation as (24). The MSE criterion disproportionately weights large forecast errors more heavily relative to MAE, and hence the former may be particularly useful in forecasting situations when large forecast errors are disproportionately more serious than small errors. The proportion of over-predictions should give a rough indication of the average direction of the forecast error (compared with the two previous measures which only give some measure of the average size) and whether the models are persistently over- or under-predicting the "true" value of volatility. Hence this measure gives an approximate guide as to whether the forecasts are biased.

6. Forecasting Results

The results of the volatility forecasting exercise are given in table 3. The proportion of over-predictions given in the last column show that almost all models consistently over-predict the true realised value of volatility. The model which comes closest to neither over- nor under-predicting more than half the time is the simple random walk; the worst models are the exponential smoothing and exponentially weighted long moving average models, which over-predict the true value of volatility 80% of the time.

The mean squared error and mean absolute error are given in the second and fourth columns respectively of table 3. The third and fifth columns give relative MSE and MAE respectively, which are calculated by dividing the actual error measure by the value of the error measure for the "best" model according to that criterion. For example, the model with the lowest MSE and MAE is the first order autoregressive model with

daily dummy variables and hence these have relative MSE and MAE of one. The GJR-GARCH model with two lags of volume thus has a MSE of 639 times and a MAE 51 times that of the best model.

Figure 1 gives a plot of the actual volatility over the entire sample period. The salient feature of this graph is that volatility is very considerably greater around the October 1987 stock market crash than at any other time over the life of the modelling and forecasting sample. This reveals a potential flaw in a number of recent studies in empirical finance[vi] in their use of historical time series data which spans the period surrounding the October 1987 crash of world equity markets. According to Gallant *et al*. (1992, p200), on 19 October 1987, the S&P Composite Index dropped 22.9% whilst volume hit the second highest figure ever recorded (604 million shares). The next day, the value of the S&P Index increased by 5.2% on an even higher daily volume of 608 million shares. This extreme series of fluctuations, which persisted for some time after that date, was exceptional and has not been repeated at such magnitude since. Therefore, whether or not we accept the proposition that these observations are "outliers", it is clear that such huge swings in volume and volatility are likely to have a profound influence on models and forecasts of the latter. Thus evaluation of forecasts of volatility during this period would represent a severe and probably inappropriate test of the models since almost the entire total sum of squares to be explained are attributable to these few observations.

6A. The Possible Effects of the October '87 Crash

In order to ensure that the results of this forecasting study are not unduly influenced by these extreme observations, the error measures are re-calculated disregarding all data from October 1, 1987 onwards[vii]. This leaves a total modelling and forecasting sample running from 17 November 1978 - 30 September 1987, and a hold-out sample of 231 pre-crash observations for evaluation of forecasts in the manner described in section 5A. The results of table 4 demonstrate that the MSE and MAE are at least halved for all models once the crash period, where by definition we would expect the model errors to be very large, is removed. Furthermore, the relative rankings of each of the models is also altered. In particular, the EGARCH models, which had performed rather poorly for the whole sample, are now among the best models. Also notably, the regression with one lag of volatility and one lag of volume has the lowest MSE of all 29 models. This model and the actual values are plotted in figure 2. Adding further lags of volume and/or volatility however slightly reduces forecasting accuracy. The autoregressive models and pure volume models of volatility of the same

order have similar aggregate forecast errors. This gives further evidence to the arguments in Karpoff (1987) that the two are strongly contemporanously correlated.

For the GARCH and the GJR models, lagged volume increases the average forecast errors by orders of magnitude. Although this result is perhaps surprising, it may be attributed to the fact that the measure of volume used in this study has been transformed in order to induce stationarity, as detailed above, and the transformed series can take on negative values. The simpler autoregressive models seem to take this in their stride, producing reasonable forecasts; for the GARCH and GJR models, the negativity appears problematic, leading to very erratic forecasts, although the forecasts themselves are never negative. The EGARCH model is not affected in the same way due to the formulation of the model in logarithms of the variance; thus negative values on the RHS of equation (19) need not present a problem for forecasting volatility which, by definition, must be positive[viii]. The neural network models give a reasonable performance, but the complexity and loss of any diagnostic information associated with the use of this technique renders them hardly worth the additional effort in this case.

6B. The Relative Predictive Power of Alternative Forecasts

The previous section has highlighted which of the models have relatively low or high values of the aggregate forecast error measures but although many models, which are very different in formulation, have approximately the same average (squared or absolute) forecast error, it is possible that the information contained in the forecast series for each model is different. This is likely to become obscured in the error aggregation process. Fair and Shiller (1990) suggest a simple encompassing method for evaluating this proposition. The actual realised value of volatility is regressed upon a constant and two or more forecasts from different models:

$$\sigma_t^2 = \beta_0 + \sum_{j=1}^{J} \beta_j \sigma_{f,t}^2(j) + u_t \qquad (26)$$

where $\sigma_{f,t}^2(j)$ is the forecast for volatility at time t for model j (and J such models are compared), β_0 , β_j are parameters to be estimated, and u_t is an i.i.d. error term. If the models all contain independent information that has power in predicting volatility, then all of the β_j should be significantly different from

zero. If, however, the information contained in one forecast is simply a subset of that contained in the other(s), then the coefficient on the former should be insignificant.

The results from application of the encompassing test are given in table 5 for the whole forecast sample, and table 6 for the pre-crash sample. The results are fairly inconclusive since the most common result is that none of the models have significant explanatory power (a result also observed by Day & Lewis, 1992). Table entries in columns 2 to 7 are coefficient estimates followed by heteroscedasticity-consistent *t*-ratios in parentheses; the unadjusted value of R^2 for each model is given in the last column. The first row in tables 5 and 6 compare a variety of different models (long term mean, exponential smoothing, daily dummy variables, regression on one lag of volume and an EGARCH model). For the entire hold-out sample, only the dummy variable forecasts have significant power in "explaining" variations of the actual volatility series, although since the coefficient is actually negative, it implies that the dummies significantly confound the forecasting of volatility; no forecast series have significant coefficients for the pre-crash sample (table 6). Rows 3 to 6 of tables 5 and 6 offer additional comparative evidence on the consequences of adding lagged volume measures to volatility forecasting equations. If volume has any additional explanatory power, we would expect the coefficient on the forecast series incorporating volume to be statistically significant. The only forecast containing volume which is significant is that for the AR(1) model and for the EGARCH model, both with a single lag of volume. This result gives additional weight to the observation that volume has a very minor role to play in predicting volatility. The R^2 values for the forecast series are also interesting. For the sample including the crash, the R^2 values are never above 5%, indicating that the models are explaining virtually none of the variability in the actual volatility series. Most of this is entirely due to the observations around the October 1987 stock market crash, and once we move to the pre-crash sample, the fit of the forecasts to the data improves substantially. The values of R^2 rise to just over 20%, indicating that still none of the forecasts, even taken in groups, have satisfactory explanatory power. This result adds further weight to the argument that volatility forecasts with similar MSE's constructed in this and numerous other studies are uniformly poor rather than uniformly good.

7. Summary and Conclusions

This paper has examined the inter-temporal relationship between stock index volatility and market trading volume. An application of linear and non-linear Granger causality shows evidence of bi-directional causality, although the relationship is stronger from volatility to volume than vice versa, particularly in the linear case. These findings demonstrate that predicting volume using volatility is likely to be more fruitful from the point of view of forecasting accuracy (see also Weigend and LeBaron, 1994), but forecasts of stock index volume are of no direct use in terms of an implication for trading or risk management. Hence the second part of this study has examined the effect on out-of-sample forecasting performance of including measures of lagged volume in equations for forecasting volatility.

The main conclusion of the second part of this study is that lagged stock market volume measures have little role to play in improving the out-of-sample forecasting performance of volatility models, almost irrespective of the criterion used to evaluate the forecasts. Although the best model for the pre-crash data according to the MSE criterion included volume, occasionally, the inclusion of lagged volume can lead to very erratic forecasts which are off the mark by orders of magnitude. In terms of the univariate forecasting models, the findings of this paper are very much in the spirit of Dimson and Marsh (1990) and Franses and van Dijk (1996) in that simpler models are generally preferable, although there is very little to choose between many of the models, and there is certainly not a single model which is universally superior. In particular, the GJR model recommended by Brailsford and Faff (1996), but maligned by Franses and van Dijk (1996), is inferior to the EGARCH formulation which neither paper considers.

These findings have a number of important implications for future research in this area. First, although linear Granger causality implies that a linear model should be used for forecasting, no such inference can be made for the case of non-linear causality. In fact, evidence of non-linear Granger causality gives the researcher no clue as to the appropriate functional form for the non-linear forecasting model, and this may go some way to explaining the reason for the disappointing forecasting performances of many non-linear models (see Ramsey, 1996): it may be that the variables are correct, but the functional form is wrong. Second, the forecasting performance of the entire GARCH family of models is worse than one might have expected. One explanation for this may be that although the returns are posited to be generated by some stochastic process, the variances, which we use as the measure of volatility, are formulated to be entirely deterministic in nature. It is just as likely that volatilities have a stochastic data generating component as the returns themselves, so

this is probably a gross over-simplification. It may therefore be the case that stochastic volatility models (see, for example, Harvey, Ruiz and Shephard, 1994 or Shephard, 1995) are more useful in this context, although no comparisons of their forecasting accuracy have yet been made.

Another explanation for the lack of additional explanatory power of the volume series could be that the measure of volume used here has been transformed (in common with many other studies) to induce stationarity, but that the newly created series bears insufficient relationship to "true" volume series. Moreover, even the literal number of shares traded (the real definition of volume) has a certain degree of arbitrariness about it since, if a company decides to issue half as many shares at $10 rather than $5, then volume will automatically be halved. Perhaps volume measures should have some value-weighting to bring them closer into line with the underlying quantity of financial interest, whatever that might be. More theoretical and empirical research on the relationship between volatility and volume, and in particular, what useful information (if any) is contained in the latter, is clearly warranted.

Acknowledgements

I am grateful to Blake LeBaron for making the data available, and to Craig Hiemstra for providing the software to compute the modified Baek and Brock test. I would also like to thank Simon Burke, Peter Hart and seminar participants at a University of Reading staff workshop and a local econometrics group workshop for helpful discussions whilst absolving them from any remaining errors.

References

Akgiray, V. 'Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts', Journal of Business **62**(1) (1989), 55-80

Andersen, T.G. 'Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility', Journal of Finance **51**(1) (1996), 169-204

Baek, E. and Brock, W. 'A Nonparameteric Test for Independence of a Multivariate Time Series', Statistica Sinica **2** (1992), 137-156

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

Bollerslev, T. 'Generalised Autoregressive Conditional Heteroskedasticity', Journal of Econometrics **31** (1986), 307-327

Brailsford, T.J. and Faff, R.W. 'An Evaluation of Volatility Forecasting Techniques', Journal of Banking and Finance **20** (1996), 419-438

Brock, W.A. 'Causality, Chaos, Explanation, and Prediction in Economics and Finance', in Casti, J. and Karlqvist (eds.) Beyond Belief: Randomness, Prediction and Explanation on Science CRC Press, Boca Raton, Florida, 1991

Brock, W.A., Dechert, W.D., Scheinkman, J.A. and LeBaron, B. 'A Test for Independence Based on the Correlation Dimension', Econometric Reviews **15**(3) (1996), 197-235

Campbell, J.Y., Grossman, S.J. and Wang, J. 'Trading Volume and Serial Correlation in Stock Returns', Quarterly Journal of Economics **108** (1993), 905-939

Cao, C.Q. and Tsay, R.S. 'Nonlinear Time-Series Analysis of Stock Volatilities', Journal of Applied Econometrics **7** (1992), S165-S185

Chan, K.C., Christie, W.G. and Schultz, P.H. 'Market Structure and the Intraday Pattern of Bid-Ask Spreads for NASDAQ Securities', Journal of Business **68**(1) (1995), 35-40

Chu, S-H. and Freund, S. 'Volatility Estimation for Stock Index Options: A GARCH Approach', Quarterly Review of Economics and Finance **36**(4) (1996), 431-450

Copeland, T.E. 'A Model of Asset Trading under the Assumption of Sequential Information Arrival', Journal of Finance **31** (1976), 1149-1168

Day, T.E. and Lewis, C.M. 'Stock Market Volatility and the Information Content of Stock Index Options', Journal of Econometrics **52** (1992), 267-287

Dimson, E. and Marsh, P. 'Volatility Forecasting without Data-Snooping', Journal of Banking and Finance **14** (1990), 399-421

Engle, R.F. and Bollerslev, T. 'Modelling the Persistence of Conditional Variances', Econometric Reviews **5**(1) (1986), 1-50

Epps, T.W. 'Security Price Changes and Transaction Volumes: Theory and Evidence', American Economic Review **65** (1975), 586-597

Epps, T.W. and Epps, M.L. 'The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture of Distributions Hypothesis', Econometrica **44** (1975), 305-321

Fair, R.C. and Shiller, R.J. 'Comparing Information in Forecasts from Econometric Models', American Economic Review **80** (1990), 375-389

Franses, P.H. and van Dijk, D. 'Forecasting Stock Market Volatility Using Non-Linear GARCH Models', Journal of Forecasting **15** (1996), 229-235

Gallant, A.R., Hsieh, D.A., and Tauchen, G. 'Stock Prices and Volume', Review of Financial Studies **5**(2) (1992), 199-242

Garman, M.B. and Klass, M.J. 'On the Estimation of Security Price Volatilities from Historical Data', Journal of Business **53**(1) (1980), 67-78

Glosten, L.R., Jagannathan, R. and Runkle, D.E. 'On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks', The Journal of Finance **48**(5) (1993), 1779-1801

Haefke, C. and Helmenstein, C. 'Stock Price Forecasting of Austrian Initial Public Offerings Using Artificial Neural Networks', Presented at the IFAC Workshop on Computational Methods in Economics and Finance, Amsterdam, 8-10 June, 1994

Harris, L. 'Transactions Data Tests of the Mixture of Distributions Hypothesis', Working Paper University of Southern California, 1984

Harvey, A.C., Ruiz, E., and Shephard, N. 'Multivariate Stochastic Variance Models', Review of Economic Studies **61** (1994), 247-264

Hiemstra, C. and Jones, J.D. 'Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation', Journal of Finance **49**(5) (1994), 1639-1664

Hiemstra, C.and Jones, J.D. 'Some Information Relating to the Finite Sample Properties of the Modified Baek and Brock Nonlinear Granger Causality Test', Working Paper, University of Strathclyde and Securities and Exchange Commission, 1993

Hornik, K., Stinchcome, Maxwell, and White, H. 'Multilayer Feedforward Networks are Universal Approximators', Neural Networks **2**(5) (1989), 359-366

Karpoff, J.M. 'The Relation Between Price Changes and Trading Volume: A Survey', Journal of Financial and Quantitative Analysis **22**(1) (1987), 109-126

Kuan, C-M. and Lim, T. 'Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks', Journal of Applied Econometrics **10** (1994), 347-364

Lamoureux, C.G. and Lastrapes, W.D. 'Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects', Journal of Finance **45**(1) (1990), 221-229

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

Najand, M. and Yung, K. 'A GARCH Examination of the Relationship between Volume and Price Variability in Futures Markets', Journal of Futures Markets **11**(5) (1991), 613-621

Nelson, D.B. 'Conditional Heteroskedasticity in Asset Returns: A New Approach', Econometrica **59**(2) (1991), 347-370

Pagan, A.R. and Schwert, G.W. 'Alternative Models for Conditional Stock Volatilities', Journal of Econometrics **45** (1990), 267-290

Parkinson, M. 'The Extreme Value Method for Estimating the Variance of the Rate of Return', Journal of Business **53**(1) (1980), 61-65

Ramsey, J.B. 'If Nonlinear Models Cannot Forecast, What Use are They?', SNDE 1(2) (1996), 1-43

Ross, S.A. 'The Interrelations of Finance and Economics: Theoretical Perspectives', American Economic Review **77** (1987), 29-34

Shephard, N. 'Statistical Aspects of ARCH and Stochastic Volatility', in Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen (eds). Likelihood, Time Series with Econometric and Other Applications Chapman & Hall, 1995

Smirlock, M. and Starks, L. 'A Transactions Approach to Testing Information Arrival Models', Working Paper Washington University, 1984

Tsibouris, G. and Zeidenberg, M. 'Testing the Efficient Market Hypothesis with Gradient Descent Algorithms', in Refenes, A-P. (ed.) Neural Networks in the Capital Markets John Wiley and Sons, Chichester, England, 127-136, 1995

Weigend, A.S. and LeBaron, B. 'Evaluating Neural Network Predictors by Bootstrapping', Working Paper presented at ICONIP '94, Seoul, 1994

West, K.D. and Cho, D. 'The Predictive Ability of Several Models of Exchange Rate Volatility', Journal of Econometrics **69** (1995), 367-391

Table 1

Results for Linear Granger Causality

Number of Lags of volume, volatility	H_0: Volatility Does not Granger cause Volume	Number of Lags of volume, volatility	H_0: Volume Does not Granger cause Volatility
Panel A: Equal Numbers of lags of each variable			
1,1	15.795 (0.0001)	1,1	12.395 (0.0004)
2,2	8.043 (0.0003)	2,2	3.515 (0.0299)
3,3	5.349 (0.0011)	3,3	2.096 (0.0986)
4,4	5.4217 (0.0002)	4,4	1.716 (0.1436)
Panel B: Lag length chosen using AIC			
10,1	18.673 (0.0000)	1,3	6.259 (0.0124)
Panel C: Lag Length Chosen using SBIC			
1,1	15.795 (0.0001)	1,1	12.395 (0.0004)

Notes: Entries in the table are F-test statistics followed by marginal significance levels in parentheses. In each case, the null distribution of the test statistic is an F(k,2430-3k), where k is the equal number of lags of each variable in the VAR.

Table 2

Results for Non-linear Granger Causality

Number of Lags ($Lx = Ly$)	H_0: Volatility Does not Granger cause Volume			H_0: Volume Does not Granger cause Volatility		
e / σ	0.5	1	1.5	0.5	1	15
1	0.0068 (4.683)**	0.0042 (3.553)**	0.0022 (2.473)**	0.0016 (2.095)*	0.0013 (2.642)**	0.007 (1.996)*
2	0.0056 (4.376)**	0.0034 (3.478)**	0.0017 (2.433)**	0.0018 (1.839)*	0.0016 (2.448)**	0.0005 (1.506)
3	0.0057 (4.247)**	0.0035 (3.525)**	0.0016 (2.396)**	0.0013 (1.177)	0.0012 (2.080)*	0.0005 (1.482)
4	0.0058 (4.005)**	0.0032 (3.425)**	0.0015 (2.448)**	0.0023 (2.055)*	0.0012 (2.183)*	0.0003 (1.078)

Notes: The first entry in each cell refers to the difference between the two conditional probabilities in equation (4), and the second is a standardised test statistic which is asymptotically distributed as a standard normal variate. * and ** denote significance at the 5% and 1% levels respectively using a one-sided test.

Table 3

Results for Forecasts of Volatility

Model Description	MSE	Relative	MAE	Relative	% Over-predi-ctions
1. Random Walk	1.91	1.85	6.17	1.65	49.42
2. Long term mean	1.05	1.02	3.78	1.01	67.52
3. Short moving average	1.16	1.13	5.11	1.37	67.98
4. Long moving average	1.06	1.03	5.37	1.44	78.88
5. Exponential smoothing	3.15	3.06	7.69	2.06	79.58
6. Short EWMA	1.22	1.18	5.34	1.43	67.75
7. Long EWMA	1.06	1.03	5.39	1.45	79.12
8. AR(1)	5.45	5.29	6.89	1.85	67.52
9. AR(3)	8.79	8.53	9.36	2.51	68.45
10. AR(10)	8.79	8.53	11.82	3.17	68.21
11. Daily dummies only	1.05	1.02	3.84	1.03	65.43
12. AR(1) + dummies	1.03	1.00	3.73	1.00	67.05
13. AR(3) + dummies	1.04	1.01	3.94	1.06	67.75
14. AR(10) + dummies	1.06	1.03	4.28	1.15	67.52
15. Regression on 1 lag of volume	1.05	1.02	4.05	1.09	56.15
16. Regression on 3 lags of volume	1.05	1.02	4.03	1.08	55.22
17. Regression on 10 lags of volume	1.05	1.02	4.13	1.11	57.08
18. Regression on 1 lag of volume and 1 lag of volatility	5.31	5.16	7.07	1.90	57.08
19. Regression on 3 lags of volume and 3 lags of volatility	7.66	7.44	9.50	2.55	61.48
20. Regression on 10 lags of volume and 10 lags of volatility	9.34	9.07	12.40	3.32	62.18
21. GARCH(1,1)	1.06	1.03	4.79	1.28	74.01
22. GARCH(1,1) with 1 lag of volume	20.64	20.04	100.06	26.83	36.66
23. GARCH(1,1) with 2 lags of volume	63.75	61.89	187.97	50.39	40.14
24. EGARCH(1,1)	2.30	2.23	6.22	1.67	65.66
25. EGARCH(1,1) with 1 lag of volume	9.59	9.31	21.80	5.84	66.13
26. EGARCH(1,1) with 2 lags of volume	32.28	31.34	274.14	73.50	64.97
27. GJR-GARCH(1,1)	1.10	1.07	5.10	1.37	74.25
28. GJR-GARCH(1,1) with 1 lag of volume	194.84	189.17	94.72	25.39	35.73
29. GJR-GARCH(1,1) with 2 lags of volume	658.20	639.03	191.40	51.31	40.37
30. Neural network	1.06	1.03	3.82	1.05	25.60
31. Neural network with 1 lag of volume	1.05	1.02	3.75	1.03	28.11

Notes: MSE and MAE denote mean squared forecast error and mean absolute forecast error respectively. The former have been multiplied by 10^5 and the latter by 10^4 for ease of comparison. Constants are included in any regression which does not include the dummy variables.

Table 4

Results for Forecasts of Volatility for the Pre-Crash Period Only

Model Description	MSE	Relative	MAE	Relative	% Over-predictions
Random Walk	7.57	1.63	1.55	1.37	50.21
Long term mean	4.73	1.02	1.15	1.02	70.95
Short moving average	5.57	1.20	1.38	1.22	66.39
Long moving average	4.68	1.01	1.26	1.12	75.93
Exponential smoothing	4.75	1.03	1.27	1.12	74.27
Short EWMA	5.84	1.26	1.38	1.22	65.98
Long EWMA	4.70	1.02	1.26	1.12	75.93
AR(1)	4.67	1.01	1.16	1.03	70.54
AR(3)	4.71	1.02	1.17	1.04	70.95
AR(10)	4.82	1.04	1.22	1.08	69.71
Daily dummies only	4.72	1.02	1.15	1.02	69.29
AR(1) + dummies	4.75	1.03	1.17	1.04	69.29
AR(3) + dummies	4.75	1.03	1.18	1.04	69.71
AR(10) + dummies	4.75	1.03	1.19	1.05	70.12
Regression on 1 lag of volume	4.71	1.02	1.15	1.02	71.37
Regression on 3 lags of volume	4.71	1.02	1.15	1.02	70.54
Regression on 10 lags of volume	4.75	1.03	1.15	1.02	70.12
Regression on 1 lag of volume and 1 lag of volatility	4.63	1.00	1.15	1.02	70.54
Regression on 3 lags of volume and 3 lags of volatility	4.68	1.01	1.17	1.04	70.54
Regression on 10 lags of volume and 10 lags of volatility	4.88	1.05	1.24	1.10	69.71
GARCH(1,1)	4.78	1.03	1.23	1.09	72.61
GARCH(1,1) with 1 lag of volume	127.52	27.54	76.73	67.90	47.72
GARCH(1,1) with 2 lags of volume	457.58	98.83	153.4	135.75	51.04
EGARCH(1,1)	4.73	1.02	1.13	1.00	70.54
EGARCH(1,1) with 1 lag of volume	6.69	1.44	1.25	1.11	71.37
EGARCH(1,1) with 2 lags of volume	4.73	1.02	1.15	1.02	70.54
GJR-GARCH(1,1)	4.78	1.03	1.23	1.09	72.20
GJR-GARCH(1,1) with 1 lags of volume	1220.19	263.54	69.13	61.18	45.64
GJR-GARCH(1,1) with 2 lag of volume	4692.41	1013.48	157.0	138.94	51.45
Neural network	4.74	1.03	1.18	1.05	28.92
Neural network with 1 lag of volume	4.71	1.02	1.16	1.03	29.17

Notes: MSE and MAE denote mean squared forecast error and mean absolute forecast error respectively. The former have been multiplied by 10^8 and the latter by 10^4 for ease of comparison. Constants are included in any regression which does not include the dummy variables.

Table 5

Forecast Comparisons Using Encompassing Test

Models Compared	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	R^2
2, 5, 11, 15, 21	0.0003	0.6284	0.0132	-1.2069	1.5202	0.2669	0.0349
	(0.8584)	(0.3613)	(1.056)	(-1.9880)*	(1.5480)	(1.5992)	
8, 18	0.0004	-0.9415	0.9780	-	-	-	0.0170
	(2.3730)*	(-0.9786)	(1.0009)				
21, 22, 23	0.0003	0.3238	0.0059	0.0034	-	-	0.0248
	(1.5533)	(1.8895)*	(0.6734)	(0.7014)			
24, 25, 26	0.0004	-0.0060	0.0019	-0.0000	-	-	0.0141
	(2.4196)*	(-1.2168)	(0.4373)	(-0.0126)			
27, 28, 29	0.0004	0.1448	-0.0329	0.02878	-	-	0.0422
	(1.4948)	(0.7134)	(-0.9694)	(1.0348)			
30, 31	0.0004	0.5183	0.0049	-	-	-	0.0312
	(1.2819)	(0.5819)	(0.5397)				

Notes: Model numbers refer to those given in table 3. Figures in parentheses are *t*-ratios computed from standard errors corrected for heteroscedasticity according to White(1980). * and ** denote significance at the 5% and 1% levels respectively.

Table 6

Forecast Comparisons Using Encompassing Test on Pre-Crash Data

Models Compared	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	R^2
2, 5, 11, 15, 21	-0.0008	11.0438	0.3678	0.4116	2.2076	-3.4370	0.2328
	(-1.0632)	(1.0983)	(1.3554)	(0.3825)	(0.4580)	(-0.6676)	
8, 18	0.0000	-4.7195	6.2018	-	-	-	0.2435
	(-0.0968)	(-1.8660)	(2.7319)**				
21, 22, 23	0.0002	-0.4495	-0.0002	-0.0002	-	-	0.2141
	(2.0440)*	(-0.6700)	(-0.1270)	(-0.2439)			
24, 25, 26	0.0000	1.5919	-0.0719	0.2110	-	-	0.2148
	(-0.1923)	(1.1210)	(-3.719)**	(0.1112)			
27, 28, 29	0.0001	-0.3497	0.0006	-0.0005	-	-	0.2137
	(2.0642)*	(-0.5588)	(0.3264)	(-0.6968)			
30, 31	0.0006	0.9031	0.1485	-	-	-	0.2207
	(1.5634)	(1.1395)	(1.0521)				

Notes: Model numbers refer to those given in table 3. Figures in parentheses are *t*-ratios computed from standard errors corrected for heteroscedasticity according to White(1980). * and ** denote significance at the 5% and 1% levels respectively.

Figure 1

Actual Volatility over the Period October 1986 - October 1988

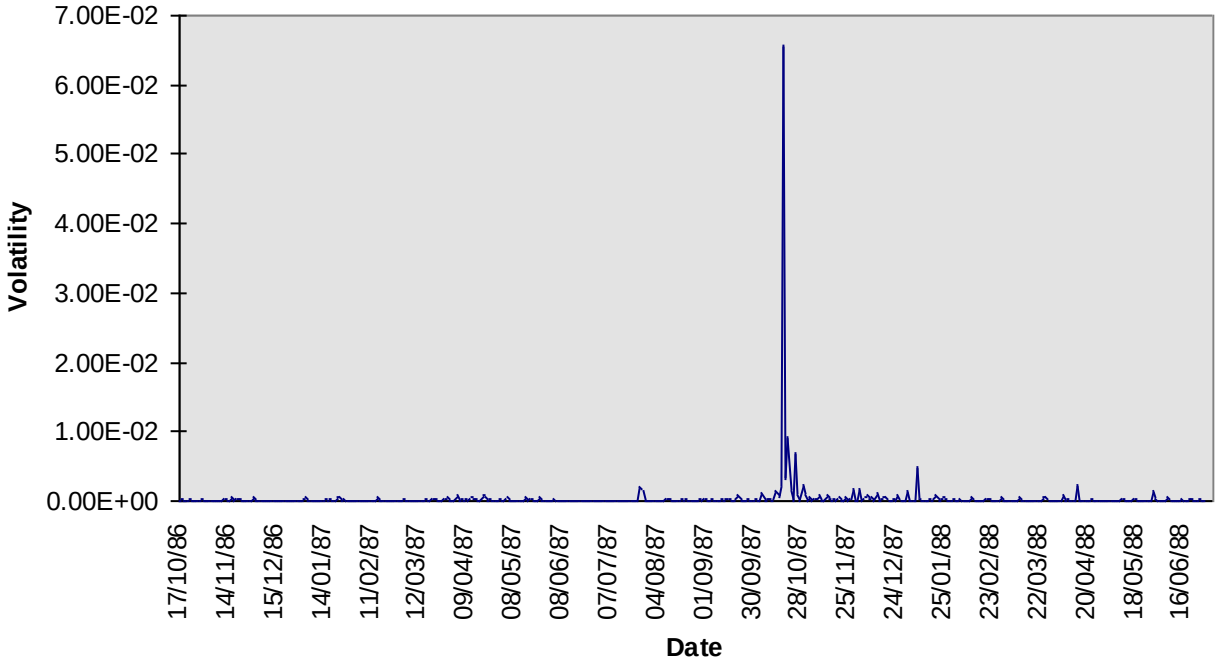
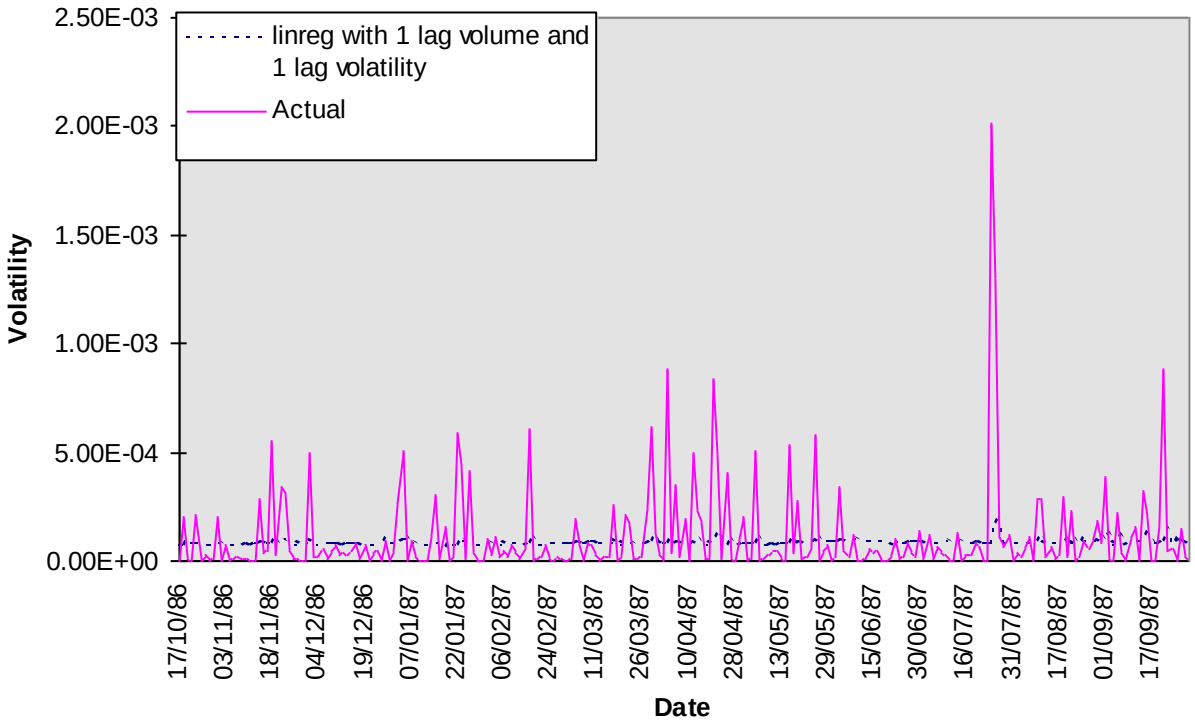


Figure 2

Actual Volatility and Forecasts using A Regression on One Lag of Volume and One Lag of Volatility for the Pre-Crash Data



[i] The alternative models are the long term mean, IGARCH, autoregressive models, and a nonparameteric model based on the Gaussian kernel.

[ii] The other models employed are the random walk, the historical mean, a short- and a long-term moving average, exponential smoothing, an exponentially weighted moving average model, and a linear regression.

[iii] An analysis of these residuals show that there is no remaining linear dependence in the series.

[iv] This has become known as the "leverage effect", first documented by Black (1976).

[v] The number of inputs and hidden nodes is seveerely constrained by their CPU requirement since these models are slow to estimate due to the large number of parameters required to be estimated. For example, the case with two inputs and eight hidden nodes entials the estimation of 30 coefficients, and the moving window used in the calculation of forecasts implies the estimation of 431 separate such models.

[vi] Examples in the volatility forecasting literature include Brailsford and Faff (1996), Dimson and Marsh (1990) and Cao and Tsay (1992).

[vii] These observations occur close to the end of the sample in any case, so that we lose little of the length by removing them.

[viii] Experimentation with squares or absolute values of the volume measure does not lead to improvements in the forecasting accuracies of these models.