Forecasting Volatility in Financial Markets:

A Review[*]

Ser-Huang Poon[†]   Clive Granger[‡]

September 18, 2002

[†]Corresponding author. Please send all correspondence to Department of Accounting and Finance, Strathclyde University, 100 Cathedral Street, Glasgow G4 0LN, UK. Tel: +44 141 548 4983, Fax: +44 141 552 3547, Email: s.poon@strath.ac.uk.

[‡]Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508. Fax: (619) 534-7040, Email: cgranger@ucsd.edu.

Abstract

Financial market volatility is an important input for investment, option pricing and financial market regulation. In this review article, we compare the volatility forecasting findings in 93 papers published and written in the last two decades. This article is written for general readers in Economics, and its emphasis is on forecasting instead of modelling. We separate the literature into two main streams; the first consists of research papers that formulate volatility forecasts based on historical price information only, while the second includes research papers that make use of volatility implied in option prices.

Provided in this paper as well are volatility definitions, insights into problematic issues of forecast evaluation, the effect of data frequency on volatility forecast accuracy, measurement of "actual" volatility, and the confounding effect of extreme values on volatility forecasting performance. We compare volatility forecasting results across different asset classes, and markets in different geographical regions. Suggestions are made for future research.

Keywords: Volatility, ARCH, Option Implied, High Frequency Data, Forecast Evaluation.

JEL classification: A10, C10, C50, G10.

1   Introduction

Volatility forecasting is an important task in financial markets and it has occupied the attention of academicians and practitioners over the last two decades. At the time of writing, there are at least 93 published and working papers that study forecasting performance of various volatility models, and several times that number have been written on the subject of volatility modelling alone without the forecasting aspect. This extensive volatility research reflects the importance of volatility in investment, security valuation, risk management and monetary policy making.

Volatility is not the same as risk. When it is interpreted as uncertainty, it becomes a key input to many investment decisions and portfolios creations. Investors and portfolio managers have certain levels of risk threshold at which they could bear. A good forecast of volatility of prices of asset over the investment holding period is a good starting point for assessing investment risk.

Volatility is the most important variable in the pricing of derivative securities, of which trading volume has been quadrupled in recent years. To price an option, we need to know the volatility of the underlying asset from now till the option expires. In fact, the market convention is to list option price in term of volatility units. Nowadays, one can buy derivatives that are written on volatility itself, in which case the definition and measurement of volatility will be clearly specified in the derivative contracts. In these new contracts, volatility now becomes the underlying "asset". So volatility forecast and a second prediction on the volatility of volatility over the defined period will be needed in order to price such derivative contracts.

Financial risk management has taken a centre role since the first Basle Accord was established in 1996. This effectively makes volatility forecasting a compulsory risk management exercise for many financial institutions around the world. Banks and trading houses have to set aside reserve capital of at least three times that of value-at-risk (VaR), which is defined as the minimum expected loss with a 1% confidence level for a given time horizon (usually 1 or 10 days). Sometimes, a 5% critical value is used. Such VaR estimates are readily available given volatility forecast, mean estimate and a normal distribution assumption for the changes in total asset value. When the normal distribution assumption is disputed, which is very often the case; volatility is still needed in the simulation process used to produce the VaR figures.

Financial market volatility can have a wide repercussion on the economy as a whole. The incidents caused by the terrorists' attack on September 11, 2001, and the recent financial reporting scandals in the US have caused great turmoil to the financial markets over several continents and a negative impact on the world economy. This is a clear

evidence of the important link between financial market uncertainty and public confidence. For this reason, policy makers often rely on market's estimates of volatility as a barometer for the vulnerability of financial markets and the economy. In the US, the Federal Reserve explicitly takes into account the volatility of stocks, bonds, currencies and commodities in establishing its monetary policy (Nasar (1992)). The Bank of England is also known to make frequent references to market sentiment and option implied densities of key financial variables in its monetary policy meetings.

Given the important role of volatility forecast and that so much has been written on volatility forecasting, this paper aims to provide a comprehensive coverage of the status of this research. Taking a utilitarian viewpoint, we believe that the success of a volatility model lies in its out-of-sample forecasting power. It is impossible, in practice, to perform tests on all volatility forecasting models on a large number of data sets and over many different periods. By carefully reviewing the methodologies and empirical findings in 93 papers, the contribution of this review is to provide a birdseye view of the whole volatility forecasting literature and to provide some recommendations for the practice and future research. Knight and Satchell (1998) is the first book to cover many issues and early empirical results related to volatility forecasting, which we draw upon frequently. However, our focus here is on the 93 papers and the collective findings in this pool of research effort. We have excluded in this review all papers that do not produce out-of-sample volatility forecasts, and papers that forecast correlations although these might be useful for forecasting portfolio risk.

The remaining sections are organized as follows. Section 2 provides some preliminaries such as the definition and the measurement of volatility, and lists some confounding factors such as forecast horizons and data frequency. Section 3 introduces the two broad categories of methods widely used in making volatility forecasts, viz. time series models and option ISD (implied standard deviation). Section 4 lists a number of forecast performance measures and raises various issues related to forecasts evaluation. Sections 5 and 6 are the core sections of this paper. Section 5 reviews research papers that forecast volatility based on historical price information only. Section 6 reviews research papers that use option ISD alone, or as an addition to historical price information, to forecast future volatility. Section 7 discusses our view about research and achievement in volatility forecasting and provides some directions for future research. Section 8 summarises and concludes. The technical specifications of volatility models are listed in an Appendix. A list containing a short summary for each of the 93 papers are provided at the end.

2 Some preliminaries

2.1 Volatility, standard deviation and risk

Many investors and generations of finance students often have an incomplete appreciation of the differences between volatility, standard deviation and risk. It is worth elucidating some of the conceptual issues here. In finance, volatility is often used to refer to standard deviation, σ, or variance, σ^2, computed from a set of observations as

$$\widehat{\sigma}^2 = \frac{1}{N-1} \sum_{t=1}^{N} \left(R_t - \overline{R} \right)^2 \;, \tag{1}$$

where $\overline{R}$ is the mean return. The sample standard deviation statistic $\widehat{\sigma}$ is a distribution free parameter representing the second moment characteristic of the sample. Only when σ is attached to a standard distribution, such as a normal or a Student-t distributions, the required probability density and cumulative probability density can be derived analytically. Indeed, σ can be calculated from any irregular shape distributions, in which case the probability density will have to be derived empirically. In the continuous time setting, σ is a scale parameter that multiplies or reduces the size of the fluctuations generated by the standard wiener process. Depending on the dynamic of the underlying stochastic process and whether or not the parameters are time varying, very different shapes of returns distributions might result. So it is meaningless to use σ as a risk measure unless it is attached to a distribution or a pricing dynamic. When σ is used to measure uncertainty, the users usually have in mind, maybe implicitly, a normal distribution for the returns distribution.

Standard deviation, σ, is the correct dispersion measure for the normal distribution and some other distributions, but not all. Other measures that were suggested and found useful including the mean absolute return and the inter-quantile range. However, the link between volatility and risk is tenuous, in particular risk is more often associated with small, or negative, returns whereas most measures of dispersion make no such distinction. The Sharpe ratio for example, defined as return in excess of risk free rate divided by standard deviation, is frequently used as an investment performance measure. It incorrectly penalizes occasional high returns. The idea of "semi-variance", an early suggestion by Markowitz (1991), which only uses the squares of returns below the mean, has not been widely used, largely because it is not operationally easy to apply in portfolio construction.

2.2 Volatility definition and measurement

As mentioned previously, volatility is often calculated as the sample standard deviation which is the square root of equation (1). Figlewski (1997) notes that since the statistical

properties of sample mean make it a very inaccurate estimate of the true mean especially for small samples, taking deviations around zero instead of the sample mean as in equation (1) typically increases volatility forecast accuracy. There are methods for estimating volatility that are designed to exploit or reduce the influence of extremes.[1] While equation (1) is an unbiased estimate of σ^2, the square root of $\widehat{\sigma}^2$ is a biased estimate of σ due to Jensen inequality.[2] Ding, Granger and Engle (1993) suggest measuring volatility directly from absolute returns.[3]

To understand the continuous time analogue of (1), we assume for the ease of exposition that the instantaneous returns are generated by the continuous time martingale,

$$dp_t = \sigma_t \; dW_{p,t} \tag{2}$$

where $dW_{p,t}$ denotes a standard wiener process. From (2) the conditional variance for the one-period returns, $r_{t+1} \equiv p_{t+1} - p_t$, is $\int_0^1 \sigma_{t+\tau}^2 d\tau$ which is known as the *integrated volatility* over the period t to $t + 1$. This quantity is also of central importance in the pricing of derivative securities under stochastic volatility (see Hull and White (1987)). While p_t can be observed at time t, σ_t is an unobservable latent variable that scales the stochastic process $dW_{p,t}$ continuously through time.

Let m be the sampling frequency and that there are m continuously compounded returns in one unit time such that

$$r_{m,t} \equiv p_t - p_{t-1/m} \; .$$

If the discretely sampled returns are serially uncorrelated and that the sample path for σ_t is continuous, it follows from the theory of quadratic variation (Karatzas and Shreve (1988)) that

$$p \lim_{m \to \infty} \left(\int_0^1 \sigma_{t+\tau}^2 d\tau - \sum_{j=1,\cdots,m} r_{m,t+j/m}^2 \right) = 0.$$

Hence time t volatility is theoretically observable from the sample path of the return process so long as the sampling process is frequent enough. The term *realized volatility* has been used in Fung and Hsieh (1991) and Andersen and Bollerslev (1998) to mean

---

[1] For example, the Maximum likelihood method proposed by Ball and Torous (1984), the high-low method proposed by Parkinson (1980) and Garman and Klass (1980).

[2] See Fleming 1998, footnote 10. Cox and Rubinstein (1985, p.256) explain how this bias can be corrected assuming a normal distribution for R_t. However, in most cases, the impact of this adjustment is small.

[3] Davidian and Carroll (1987) show absolute returns volatility specification is more robust against asymmetry and nonnormality. There are some empirical evidence that deviations or absolute returns based models produce better volatility forecasts than models based on squared returns (Taylor (1986), Ederinton and Guan (2000a) and McKenzie (1999)) but the majority of time series volatility models are squared returns models.

the sum of intraday squared returns at short intervals such as 15- or 5-minutes.[4] Such a volatility estimator has been shown to provide an accurate estimate of the latent process that defines volatility. Characteristics of financial market data used in these studies suggest that returns measured at an interval shorter than 5 minutes are plagued by spurious serial correlation caused by various market microstructure effects including nonsynchronous trading, discrete price observations, intraday periodic volatility pattern and bid-ask bounce. Andersen and Bollerslev (1998) and Christodoulakis and Satchell (1988) show how the inherent noise in the approximation of actual and unobservable volatility by square returns results in misleading forecast evaluation. These theoretical results turn out to have a major implication on volatility forecasting research. We shall return to this important issue in Section 4.4.

2.3   Stylized facts about financial market volatility

There are several salient features about financial time series and financial market volatility that are now well documented. These include fat tail distributions of risky asset returns, volatility clustering, asymmetry and mean reversion, and co-movements of volatilities across assets and financial markets. More recent research finds correlation among volatility is stronger than that among returns and both tend to increase during bear markets and financial crises. Since volatility of financial time series has complex structure, Diebold, Hickman, Inoue and Schuermann (1998) warn that forecast estimates will differ depending on the current level of volatility, volatility structure (e.g. the degree of persistence and mean reversion etc.) and the forecast horizon. These will be made clearer in the discussions below.

If returns are *iid* (independent and identically distributed, or strict white noise), then variance of returns over a long horizon can be derived as a simple multiple of single period variance. But, this is clearly not the case for many financial time series because of the stylized facts listed above. While a point forecast of $\widehat{\sigma}_{T-1,T\,|\,t-1}$ becomes very noisy as $T \to \infty$, a cumulative forecast, $\widehat{\sigma}_{t,T\,|\,t-1}$, becomes more accurate because of errors cancellation and volatility mean reversion unless there is a fundamental change in the volatility level or structure.[5]

---

[4]In the foreign exchange markets, quotes for major exchange rates are available round the clock. In the case of stock markets, close-to-open square return is used in the volatility aggregation process during market close.

[5]$\widehat{\sigma}_{t,T\,|\,t-1}$ denotes a volatility forecast formulated at time $t-1$ for volatility over the period from t to T. In pricing options, the required volatility parameter is the expected volatility over the life of the option. The pricing model relies on a riskless hedge to be followed through until the option reaches maturity. Therefore the required volatility input, or the implied volatility derived, is a cumulative volatility forecast over the option maturity and not a point forecast of volatility at option maturity. The interest in

Some studies find volatility time series appears to have a unit root (Perry (1982) and Pagan and Schwert (1990)). That is

$$\sigma_t = \phi\sigma_{t-1} + \varepsilon_t,$$

with ϕ indistinguishable from 1. Other papers find some volatility measures of daily and intra-day returns have a long memory property (see Granger, Ding, and Spear (2000) for examples and references). The autocorrelations of variances, and particularly those of mean absolute deviations, stay positive and significantly above zero for lags up to a thousand or more. These findings are important because they imply that a shock in the volatility process will have a long lasting impact.

Complication in relation to the choice of forecast horizon is partly due to volatility mean reversion. In general, volatility forecast accuracy improves as data sampling frequency increases relative to forecast horizon (Andersen, Bollerslev and Lange (1999)). However, for forecasting volatility over a long horizon, Figlewski (1997) finds forecast error doubled in size when daily data, instead of monthly data, is used to forecast volatility over 24 months. In some cases, where application is of very long horizon e.g. over 10 years, volatility estimate calculated using weekly or monthly data is better because volatility mean reversion is difficult to adjust using high frequency data. In general, model based forecasts lose supremacy when the forecast horizon increases with respect to the data frequency. For forecast horizons that are longer than 6 months, a simple historical method using low frequency data over a period at least as long as the forecast horizon works best (Alford and Boatsman (1995) and Figlewski (1997)).

As far as sampling frequency is concerned, Drost and Nijman (1993) prove, theoretically and for a special case (i.e. the GARCH(1,1) process, which will be introduced in Section 3.1.2 later), that volatility structure should be preserved through intertemporal aggregation. This means that whether one models volatility at the hourly, daily or monthly intervals, the volatility structure should be the same. But, it is well known that this is not the case in practice; volatility persistence, which is highly significant in daily data, weakens as the frequency of data decreases.[6] This further complicates any attempt to generalize volatility patterns and forecasting results.

forecasting $\sigma_{t,T\,|\,t-1}$ goes beyond the riskless hedge argument however.

[6]See Diebold (1988), Baillie and Bollerslev (1989) and Poon and Taylor (1992) for examples. Note that Nelson (1992) points out separately that as the sampling frequency becomes shorter, volatility modelled using discrete time model approaches its diffusion limit and persistence is to be expected provided that the underlying returns is a diffusion or a near diffusion process with no jumps.

3 Models used in volatility forecasting

In this section, we first describe various popular time series volatility models that use the historical information set to formulate volatility forecasts and a second approach that derives market estimates of future volatility from traded option prices. Nonparametric methods for volatility forecasting have been suggested. But, as nonparametric methods were reported to perform poorly (Pagan and Schwert (1990) and West and Cho (1995)), they will not be discussed here. Also excluded from discussion here are volatility models that are based on neural network (Hu and Tsoukalas (1999)), genetic programming (e.g. Zumbach, Pictet and Masutti (2001), time change and duration (e.g. Cho and Frees (1988) and Engle and Russell (1998)).

3.1 Times series volatility forecasting models

Brown (1990), Engle (1993) and Aydemir (1998) contain lists of time series models for estimating and modelling volatility. Kroner (1996) explains how volatility forecasts can be created and used. In this section, we narrow our discussion to models that are used in the 93 papers reviewed here. The specifications of these volatility models are provided in Appendix A.

All models described in this section capture volatility persistence or clustering. Others take into account volatility asymmetry also. It is quite easy to construct a supply and demand model for financial assets, with supply a constant and demand partly driven by an external instrument that enters non-linearity, that will produce a model for financial returns that is heteroskedastic. Such a model is to some extent "theory based" but is not necessarily realistic. The pure time series models discussed in this section are not based on theoretical foundations but are selected to capture the main features of volatility found with actual returns. If successful in this, it is reasonable to expect that they will have some forecasting ability.

3.1.1 Predictions based on past standard deviations

This group of models start on the basis that $\sigma_{t-\tau}$ for all $\tau > 0$ is known or can be estimated at time $t - 1$. The simplest historical price model is the *Random Walk* model, where σ_{t-1} is used as a forecast for σ_t. Extending this idea, we have the *Historical Average* method, the simple *Moving Average* method, the *Exponential Smoothing* method and the *Exponentially Weighted Moving Average* method. The *Historical Average* method makes use of all historical standard deviations while the *Moving Average* method discards the older estimates. Similarly the *Exponential Smoothing* method uses all historical estimates and the *Exponentially Weighted Moving Average (*EWMA *)* method uses only the more

9

recent ones. But unlike the previous two, the two exponential methods place greater weights on the more recent volatility estimates. All together, the four methods reflect a tradeoff between increasing the number of observations and sampling nearer to time t.

The $Riskmetrics^{TM}$ model is a procedure that uses the EWMA method. *Smooth Transition Exponential Smoothing* model, proposed by Taylor (2001), is a more flexible version of exponential smoothing where the weight depends on the size, and sometimes the sign as well, of the previous return. Next we have the *Simple Regression* method which expresses volatility as a function of its past values and an error term. The *Simple Regression* method is principally autoregressive. If past volatility errors are also included, one gets the ARMA model for volatility. Introducing a differencing order I(d), we get ARIMA when $d = 1$ and ARFIMA when $d < 1$. Finally, we have the *Threshold Autoregressive* model, where the thresholds separate volatility into states with independent *simple regression* models and noise processes for volatility in each state.

Apart from *Random Walk* and *Historical Average*, successful applications of models described in this section normally involve searching for the optimal lag length or weighting scheme in an estimation period for out-of-sample forecasting. Such optimisation generally involves minimising in-sample volatility forecast errors. A more sophisticated forecasting procedure would involve constant updating of parameter estimates when new information is observed and absorbed into the estimation period.

3.1.2 ARCH class conditional volatility models

A more sophisticated group of time series models is the ARCH family, which is extensively surveyed in Bera and Higgins (1993), Bollerslev, Chou and Kroner (1992), Bollerslev, Engle and Nelson (194) and Diebold and Lopez (1995). In contrast to models described in Section 3.1.1, ARCH class models do not make use of sample standard deviations, but formulate conditional variance, h_t, of returns via maximum likelihood procedure. Moreover, because of the way ARCH class models are constructed, h_t is known at time $t - 1$. So the one-step ahead forecast is readily available. Forecasts that are more than one-step ahead can be formulated based on an iterative procedure.

The first example of ARCH model is ARCH(q) (Engle (1982)) where h_t is a function of q past squared returns. In GARCH (p, q) (Bollerslev (1986) and Taylor (1986)), additional dependencies are permitted on p lags of past h_t. Empirical findings suggest that GARCH is a more parsimonious model than ARCH, and GARCH(1,1) is the most popular structure for many financial time series. It turns out that $Riskmetrics^{TM}$ EWMA is a non-stationary version of GARCH(1,1) where the persistence parameters sum to 1 and there is no finite fourth moment. Such a model is often called an integrated model, which should not be confused with integrated volatility described in Section 2.2.

While unconvincing theoretically as a volatility generating process, an integrated model for volatility can nevertheless be estimated and has been shown to be powerful for prediction over short horizon as it is not conditioned on a mean level of volatility and it adjusts to changes in unconditional volatility quickly.

The EGARCH (Exponential GARCH) model (Nelson (1991)) specifies conditional variance in logarithmic form, which means that there is no need to impose estimation constraint in order to avoid negative variance. With appropriate conditioning of the parameters, this specification captures the stylized fact that a negative shock lead to a higher conditional variance in the subsequent period than a positive shock would. Other models that allow for non-symmetrical dependencies are the TGARCH (Threshold GARCH) which is similar to the GJR-GARCH (Glosten, Jagannathan and Runkle (1993)), QGARCH (Quadratic GARCH) and various other non-linear GARCH reviewed in Franses and van Dijk (2000).

Both ARCH and GARCH models have been implemented with a Hamilton (1989) type regime switching framework, where volatility persistence can take different values depending on whether it is in high or low volatility regimes. The most generalised form of regime switching model is the RS-GARCH(1,1) model used in Gray (1996) and Klaassen (1998).

As mentioned before volatility persistence is a feature that many time series models are designed to capture. A GARCH model features an exponential decay in the autocorrelation of conditional variances. However, it has been noted that squared and absolute returns of financial assets typically have serial correlations that are slow to decay similar to those of an I(d) process. A shock in the volatility series seems to have very "long memory" and impact on future volatility over a long horizon. The Integrated GARCH (IGARCH) model of Engle and Bollerslev (1986) captures this effect but a shock in this model impacts upon future volatility over an infinite horizon and the unconditional variance does not exist for this model. This gives rise to FIGARCH(p, d, q) in Baillie, Bollerslev and Mikkelsen (1996) and FIEGARCH(p, d, q) in Bollerslev and Mikkelsen (1996) with $d \geq 0$. Provided that $d < 0.5$, the fractional integrated model is covariance stationary. However, as Hwang and Satchell (1998) and Granger (2001) point out, positive I(d) process has a positive drift term or a time trend in volatility level which is not observed in practice. This is a major weakness of the fractionally integrated model for it to be adopted as a theoretically sound model for volatility.

It is important to note that there are many data generating processes, other than an I(d) process, that also exhibit long memory in covariances. The short-memory stationary series with occasional breaks in mean in Granger and Hyung (2000) is an example. Diebold and Inoue (2001) show stochastic regime switching can be easily confused with

long memory if only a small amount of regime switching occurs. Zumbach (2002), on the other hand, captures long memory using IGARCH(2) (i.e. the sum of two IGARCH) and a LM model which aggregates high frequency squared returns with a set of power law weights.

3.1.3 Stochastic volatility models

In the stochastic volatility (SV) modelling framework, volatility is subject to a source of innovations that may or may not be related to those that drive returns. Modelling volatility as a stochastic variable immediately leads to fat tail distributions for returns. Autoregressive term in the volatility process introduces persistence, and correlation between the two innovative terms in the volatility process and the return process produces volatility asymmetry (Hull and White (1987, 1988)). Long memory SV models have also been proposed by allowing volatility to have a fractional integrated order (see Harvey, 1998).

For an excellent survey of SV work see Ghysels, Harvey and Renault (1996) but the subject is rapidly changing. The volatility noise term makes the SV model a lot more flexible, but as a result the SV model has no closed form, and hence cannot be estimated directly by maximum likelihood. The quasi-maximum likelihood estimation (QMLE) approach of Harvey, Ruiz and Shephard (1994) is inefficient if volatility proxies are non-Gaussian (Andersen and Sorensen (1997)). The alternatives are the generalized method of moments (GMM) approach through simulations (Duffie and Singleton (1993)), or analytical solutions (Singleton (2001)), and the likelihood approach through numerical integration (Fridman and Harris (1988)) or Monte Carlo integration using either importance sampling (Danielsson (1994), Pitt and Shephard (1997), Durbin and Koopman (2000)) or Markov Chain (e.g. Jacquier, Polson and Rossi (1994), Kim, Shephard and Chib (1998)).

3.2 Options based volatility forecasts

A European style call (put) option is a right, but not an obligation, to purchase (sell) an asset at a strike price on option maturity date, T. An American style option is a European option that can be exercised prior to T. The Black-Scholes model for pricing European equity options (Black and Scholes (1973)) assumes stock price has the following dynamics

$$dS = \mu S dt + \sigma S dz \ , \tag{3}$$

and for the growth rate on stock

$$\frac{dS}{S} = \mu dt + \sigma dz \ . \tag{4}$$

From ito lemma, the logarithmic of stock price has the following dynamics

$$d \ln S = \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dz \ , \qquad (5)$$

which means that stock price has a lognormal distribution or the logarithm of stock price has a normal distribution. Using a riskless hedge argument, Black-Scholes proved that under certain assumptions, options prices can be derived using a risk neutral valuation relationship where all derivative assets generate only risk free returns. Under this risk neutral setting, investor risk preference and the required rate of returns on stock, μ in (4), are irrelevant as far as the pricing of derivatives is concerned. The Black-Scholes assumptions include constant volatility, σ, short sell with full use of proceeds, no transaction costs or taxes, divisible securities, no dividend before option maturity, no arbitrage, continuous trading and a constant risk free interest rate, r.

Empirical findings suggest that option pricing is not sensitive to the assumption of a constant interest rate. There are now good approximating solutions for pricing American style options that can be exercised early and options that encounter dividend payments before option maturity. The impact of stochastic volatility on option pricing is much more profound, an issue which we shall return to shortly. Apart from the constant volatility assumption, the violation of any of the remaining assumptions will result in the option price being traded within a band instead of at the theoretical price.

The Black-Scholes European option pricing formula states that option price at time t is a function of S_t (the price of the underlying asset), X (the strike price), r (the risk free interest rate), T (time to option maturity) and σ (volatility of the underlying asset over the period from t to T). Given that S_t, X, r and T are observable, once the market has produced a price (either a quote or a transaction price) for the option, we could use a backward induction technique to derive σ that the market used as an input. Such a volatility estimate is called *option implied volatility*. Since the reference period is from t to T in the future, option implied volatility is often interpreted as a market's expectation of volatility over the option's maturity, i.e. the period from t to T.

Given that each asset can have only one σ, it is a well known puzzle that options of the same time to maturity, but differ in strikes, appeared to produce different implied volatility estimates for the same underlying asset. Volatility smile, smirk and sneer are names given to non-linear shapes of implied volatility plots (against strike price). There are at least two theoretical explanations (viz. distributional assumption and stochastic volatility) for this puzzle. Other explanations that are based on market microstructure and measurement errors (e.g. liquidity, bid-ask spread and tick size) and investor risk preference (e.g. model risk, lottery premium and portfolio insurance) have also been proposed.

3.2.1 Distributional assumption

To understand how Black-Scholes distributional assumption produces volatility smile, we need to make use of the positive relationship between volatility and option price, and the put-call parity[7]

$$c_t + Xe^{r(T-t)} = p_t + S_t \tag{6}$$

which established the positive relationship between call and put option prices. Since implied volatility is positively related to option price, equation (6) suggests there is also a positive relationship between implied volatilities derived from call and put options that have the same strike price and the same time to maturity.

As mentioned before Black-Scholes requires stock price in (5) to follow a lognormal distribution or the logarithmic stock returns to have a normal distribution. It is now a widely documented empirical evidence that risky financial asset returns have leptokurtic tails. In the case where strike price is very high, the call option is deep-out-of-the-money[8] and the probability for this option to be exercised is very low. Nevertheless, a leptokurtic right tail will give this option a higher probability, than that from a normal distribution, for the terminal asset price to exceed the strike price and the call option to finish in the money. This higher probability leads to a higher call price and a higher Black-Scholes implied volatility at high strike.

Next, we look at the case when strike price is low. First note that option value has two components; intrinsic value and time value. Intrinsic value reflects how deep the option is in the money. Time value reflects the amount of uncertainty before the option expires, hence it is most influenced by volatility. Deep-in-the-money call option has high intrinsic value and little time value, and a small amount of bid-ask spread or transaction tick size is sufficient to perturb the implied volatility estimation. We could however make use of the previous argument but apply it to out-of-the-money (OTM) put option at low strike price. OTM put option has a close to nil intrinsic value and the put option price is due mainly to time value. Again because of the thicker tail on the left, we expect the probability that OTM put option finishes in the money to be higher than that for a normal distribution. Hence the put option price (and hence the call option price through put-call parity) should be greater than that predicted by Black-Scholes. If we use Black-Scholes to invert volatility estimates from these option prices, the Black-Scholes implied will be

---

[7]Much of the discussion here is from Hull (2000).

[8]In option terminology, an option is *out of the money* when it is not profitable to exercise the option. For call option, this happens when $S < X$, and in the case of put, the condition is $S > X$. The reverse is true for *in-the-money* option. A call or a put is said to be *at the money* (ATM) when $S = X$. *Near-the-money* option is an option that is not exactly ATM, but close to being ATM. Sometimes, discounted values of S and X are used in the conditions.

higher than actual volatility. This results in volatility smile where implied volatility is much higher at very low and very high strikes.

The above arguments apply readily to the currency market where exchange rate returns exhibit thick tail distributions that are approximately symmetrical. In the stock market, volatility skew (i.e. low implied at high strike but high implied at low strike) is more common than volatility smile after the October 1987 stock market crash. Since the distribution is skewed to the far left, the right tail can be thinner than the normal distribution. In this case implied volatility at high strike will be lower than that expected from a volatility smile.

3.2.2 Effect of stochastic volatility

The thick tail and non-symmetrical distribution referred to in the previous section could be a result of volatility being stochastic. First, we rewrite (3) as

$$dS_t = \mu_s \ S_t \ dt + \sigma_t \ S_t \ dz_s \tag{7}$$

and now σ_t has its own dynamics

$$d\sigma_t^2 = \left(\mu_v - \beta\sigma_t^2\right) \ dt + \sigma_v \ \sigma_t \ dz_v \tag{8}$$

where β is the speed of the volatility process mean reverting to the long run average $\left(\frac{\mu_v}{\beta}\right)$, σ_v is the *volatility of volatility*, and ρ, not shown above, is the correlation between dz_s and dz_v.

When $\rho = 0$, the price process and the volatility process are not correlated, σ_v alone is enough to produce kurtosis and Black-Scholes volatility smile. When $\rho < 0$, large negative returns corresponds to high volatility stretching the left tail further into the left. On the other hand, when return is very high, volatility is low, "squashing" the right tail nearer to the centre. This will give rise to low implied volatility at high strikes and volatility skew. The reverse is true when $\rho > 0$.

Given that volatility is not a directly tradable asset, the hedging mechanism used in Black-Scholes may not apply and the risk neutral valuation principle has to be modified since volatility may command a risk premium. Different approaches to this problem have been adopted. Hull and White (1987) assume volatility risk is not priced. Wiggins (1987) derives various specifications of volatility risk premium according to different assumptions for risk preference. Heston (1993) provides a specification where volatility risk premium is proportional to variance and extract this volatility risk premium from option prices in the same manner as implied volatility is extracted. Despite the variety of approaches adopted, consensus emerge on the degree of Black-Scholes pricing bias as a result of stochastic volatility. In the case where volatility is stochastic and $\rho = 0$, Black-Scholes

15

overprices near-the-money (NTM) or at-the-money (ATM) options and the degree of overpricing increases with maturity. On the other hand, Black-Scholes underprices both in- and out-of-the-money options. In term of implied volatility, ATM implied volatility will be lower than actual volatility while implied volatility of far-from-the-money options (i.e. either very high or very low strikes) will be higher than actual volatility. The pattern of pricing bias will be much harder to predict if ρ is not zero, there is a premium for bearing volatility risk, and if either or both values vary through time.

Some of the early work on option implied volatility focus on finding an optimal weighting scheme to aggregate implied volatility of options across strikes. (See Bates (1996) for a comprehensive survey of these weighting schemes.) Since the plot of implied volatility against strikes can take many shapes, it is not likely that a single weighting scheme will remove all pricing errors consistently. For this reason and together with the liquidity argument presented below, ATM option implied volatility is often used for volatility forecast but not implied volatilities at other strikes.

3.2.3 Market microstructure and measurement errors

Early studies of option implied volatility suffered many estimation problems[9] such as the improper use of the Black-Scholes model for American style option, the omission of dividend payments, the option price and the underlying asset prices were not recorded at the same time, or stale prices were used. Since transactions may take place at bid or ask prices, transaction prices of option and the underlying assets are subject to bid-ask bounce making the implied volatility estimation unstable. Finally, in the case of S&P 100 OEX option, the privilege of a wildcard option is often omitted.[10] In more recent studies, much of these measurement errors have been taken into account. Many studies use futures and options futures because these markets are more active than the cash markets and hence the smaller risk of prices being stale.

Conditions in the Black-Scholes model include no arbitrage, transaction cost is zero and continuous trading. The lack of such a trading environment will result in option being traded within a band around the theoretical price. This means that implied volatility estimates extracted from market option prices will also lie within a band even without the complications described in Sections 3.2.1 and 3.2.2. Figlewski (1997) shows that implied volatility estimates can differ by several percentage points due to bid-ask spread

---

[9]Mayhew (1995) gives a detailed discussion on such complications involved in estimating implied volatility from option prices, and Hentschel (2001) provides a discussion of the confidence intervals for implied volatility estimates.

[10]This wildcard option arises because the stock market closes later than the option market. Option trader is given the choice to decide, before the stock market closes, whether or not to trade on an option whose price is fixed at an earlier closing time.

and discrete tick size alone. To smooth out errors caused by bid-ask bounce, Harvey and Whaley (1992) use a nonlinear regression of ATM option prices observed in a ten-minute interval before the market close on model prices.

Indication of non-ideal trading environment is usually reflected in poor trading volume. This means implied volatility of options written on different underlying assets will have different forecasting power. For most option contracts, ATM option has the largest trading volume. This supports the popularity of ATM implied volatility referred to in Section 3.2.2.

3.2.4   Investor risk preference

In the Black-Scholes's world, investor risk preference is irrelevant in pricing options. Given that some of the Black-Scholes assumptions have been shown to be invalid, there is now a model risk. Figlewski and Green (1999) simulate option writer's positions in the S&P 500, DM/$, US LIBOR and T-Bond markets using actual cash data over a 25 year period. The most striking result from the simulations is that delta hedged short maturity options, with no transaction costs and a perfect knowledge of realised volatility, finished with losses on average in all four markets. This is a clear evidence of Black-Scholes model risk. If option writers are aware of this model risk and mark up option prices accordingly, the Black-Scholes implied volatility will be greater than the true volatility.

In some situations, investor risk preference may override the risk neutral valuation relationship. Figlewski (1997), for example, compares the purchase of an OTM option to buying a lottery ticket. Investors are willing to pay a price that is higher than the fair price because they like the potential payoff and the option premium is so low that mispricing becomes negligible. On the other hand, we also have fund managers who are willing to buy comparatively expensive put options for fear of the collapse of their portfolio value. Both types of behaviour could cause market price of option to be higher than the Black-Scholes price, translating into a higher Black-Scholes implied volatility. Arbitrage argument does not apply here because these are unique risk preference (or aversion) associated with some groups of individuals. Franke, Stapleton and Subrahmanyam (1998) provide a theoretical framework in which such option trading behaviour may be analysed.

3.2.5   Option implied volatility measure and forecast

From the discussion above, we may deduce that the construction of VIX by the Chicago Board of Options Exchange is an example of good practice. VIX, short for volatility index, is an implied volatility composite compiled from eight options written on the S&P 100. It is constructed such that it is at-the-money (by combining just-in and just-out-of-the-money options) and has a constant 28 calendar days to expiry (by combining the first

nearby and second nearby options around the targeted 28 calendar days to maturity). Eight option prices are used, including four calls and four puts, to reduce any pricing bias and measurement errors caused by staleness in the recorded index level. Since options written on S&P 100 are American style, a cash-dividend adjusted binomial model was used to capture the effect of early exercise. The mid bid-ask option price is used instead of traded price because transaction prices are subject to bid-ask bounce. (See Whaley (1993) and Fleming, Ostdiek and Whaley (1995) for further details.) Due to the calendar day adjustment, VIX is about 1.2 times (i.e. $\sqrt{365/252}$) greater than historical volatility computed using trading day data.

Once an implied volatility estimate is obtained, it is usually scaled by $\sqrt{n}$ to get an n-day ahead volatility forecast. In some cases, a regression model may be used to adjust for historical bias (e.g. Ederington and Guan (2000)), or the implied volatility may be parameterized within a GARCH/ARFIMA model with or without its own persistence adjustment (e.g. Day and Lewis (1992), Blair, Poon & Taylor (2001), Hwang and Satchell (1998)).

As mentioned earlier, option implied volatility is perceived as a market's expectation of future volatility and hence it is a market based volatility forecast. Arguably it should be superior than time series volatility forecast. On the other hand, we explained before that option model based forecast requires a number of assumptions to hold for the option theory to produce a useful volatility estimate. Moreover, option implied also suffers from many market driven pricing irregularities detailed above. Nevertheless, as we will learn in Section 6, option implied volatility appears to have superior forecasting capability, out performing many historical price volatility models and matching the performance of forecasts generated from time series models that use a large amount of high frequency data.

4 Forecast Evaluation

Comparing forecasting performance of competing models is one of the most important aspects of any forecasting exercise. In contrast to the efforts made in the construction of volatility models and forecasts, little attention has been paid to forecast evaluation in the volatility forecasting literature.

4.1 Measuring forecast errors

Ideally an evaluation exercise should measure the relative or absolute usefulness of a volatility forecast to investors. However, to do that one needs to know the decision process that will include these forecasts and the costs or benefits that result from using

these forecasts. Other utility-based criterion, such as that used in West, Edison and Cho (1993), requires some assumptions about the shape and property of the utility function. In practice these costs, benefits and utility function are not known and it is usual to simply use measures suggested by statisticians.

Popular evaluation measures used in the literature include *Mean Error* (ME), *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), and *Mean Absolute Percent Error* (MAPE). Other less commonly used measures include *Mean logarithm of absolute errors* (MLAE), Theil-U statistic and LINEX. Except for the last two performance measures, all the other performance measures are self-explanatory. Assume that the subject of interest is X_i, $\widehat{X}_i$ is the forecast of X_i, and that there are N forecasts. The Theil-U measure is:

$$\text{Theil-}U = \frac{\sum_{i=1}^{N} \left(\widehat{X}_i - X_i \right)^2}{\sum_{i=1}^{N} \left(\widehat{X}_i^{BM} - X_i \right)^2} \quad , \tag{9}$$

where $\widehat{X}_i^{BM}$ is the Benchmark forecast, used here to remove the effect of any scalar transformation applied to X.

In the LINEX loss function below the positive errors are weighted differently from the negative errors:

$$\text{LINEX} = \frac{1}{N} \sum_{i=1}^{N} \left[\exp \left\{ -a \left(\widehat{X}_i - X_i \right) \right\} + a \left(\widehat{X}_i - X_i \right) - 1 \right] \quad . \tag{10}$$

The choice of the parameter a is subjective. If $a > 0$, the function is approximately linear for over-prediction and exponential for under-prediction. Granger (1999) describes a variety of non-symmetric cost, or loss, functions of which the LINEX is an example. Given that most investors would treat gains and losses differently, use of such functions may be advisable, but their use is not common in the literature.

4.2 Comparing forecast errors of different models

In the special case where the error distribution of one forecasting model dominates that of another forecasting model, the comparison is straightforward (Granger (1999)). In practice, this is rarely the case, and most comparisons are made based on the average figure of some statistical measures described in Section 4.1. For statistical inference, West (1996), West and Cho (1995) and West and McCracken (1998) show how standard errors for ME, MSE, MAE and RMSE may be derived taking into account serial correlation in the forecast errors and uncertainty inherent in model parameters estimates that were used to produce the forecasts. In general, West (1996) asymptotic theory works for recursive scheme only, where newly observed data is used to expand the estimation period.

However, a rolling fixed estimation-period method, where the oldest data is dropped whenever a new data is added, might be more appropriate if there is non-stationarity or time variation in model parameters estimates.

Diebold and Mariano (1995) propose three tests for "equal accuracy" between two forecasting models. The tests relate prediction error to some very general loss function and analyse loss differential derived from errors produced by two competing models. The three tests include an asymptotic test that corrects for series correlation and two exact finite sample tests based on sign test and the Wilcoxon's signed-rank test. Simulation results show that the three tests are robust against non-Gaussian, nonzero mean, serially and contemporaneously correlated forecast errors. The two sign based tests in particular continue to work well among small samples.

Instead of striving to make some statistical inference, model performance could be judged on some measures of economic significance. Examples of such an approach include portfolio improvement based on volatility forecasts (Fleming, Kirby, Ostdiek (2000, 2002)). Some papers test forecast accuracy by measuring the impact on option pricing errors (Karolyi (1993)). In this case, if there is any pricing error in the option model, the mistake in volatility forecast will be cancelled out when the option implied is re-introduced into the pricing formula. So it is not surprising that evaluation which involves comparing option pricing errors often prefers the implied volatility method to all other time series methods.

What has not yet been done in the literature is to separate the forecasting period into "normal" and "exceptional" periods. It is conceivable that different forecasting methods are suited for different trading environment.

4.3 Regression based forecast efficiency and orthogonality test

The regression based method for examining the informational content of forecasts entails regressing the "actual", X_i, on the forecasts, $\widehat{X}_i$, as shown below

$$X_i = \alpha + \beta \widehat{X}_i + \upsilon_t . \tag{11}$$

Conditioning upon the forecast, the prediction is unbiased only if $\alpha = 0$ and $\beta = 1$. The standard errors of the parameter estimates are often computed based on Hansen and Hodrick (1980) since the error term, υ_t, is heteroskedastic and serially correlated when overlapping forecasts are evaluated. In cases where there are more than one forecasting models, additional forecasts are added to the right hand side of (11) to check for incremental explanatory power. Such forecast encompassing test dates back to Theil (1966). Chong and Hendry (1986) and Fair and Shiller (1989, 1990) provide further theoretical

exposition of such method for testing forecast efficiency. The first forecast is said to subsume information contained in other forecasts if these additional forecasts do not significantly increase the adjusted regression R^2. Alternatively, an orthogonality test may be conducted by regressing the residuals from (11) on other forecasts. If these forecasts are orthogonal, i.e. do not contain additional information, then the regression coefficients will not be different from zero.

While it is useful to have an unbiased forecast, it is important to distinguish between biasness and predictive power. A biased forecast can have predictive power if the bias can be corrected. An unbiased forecast is useless if forecast errors are always big. For $\widehat{X}_i$ to be considered as a good forecast, $Var\left(v_t\right)$ should be small and R^2 for the regression should tend to 100%. Blair, Poon and Taylor (2001) use the proportion of explained variability, P, to measure explanatory power

$$P = 1 - \frac{\sum\left(X_i - \widehat{X}_i\right)^2}{\sum\left(X_i - \mu_X\right)^2} . \tag{12}$$

The ratio in the right hand side of (12) compares the sum of squared prediction errors (assuming $\alpha = 0$ and $\beta = 1$ in (11)) with the sum of squared variations of X_i. P compares the amount of variations in the forecast errors with that in actual volatility. If prediction errors are small, P is closer to 1. Given that regression model that produces (12) is more restrictive than (11), P is likel to be smaller than conventional R^2. P can even be negative since the ratio in the right hand side of (12) can be greater than 1. A negative P means that the forecast errors have a greater amount of variations than the actual volatility, which is not a desirable characteristic for a well behave forecasting model.

4.4 Using squared return to proxy actual volatility

Given that volatility is a latent variable, the actual volatility X is often estimated from a sample using equation (1) which is not entirely satisfactory when the sample size is small. Before high frequency data becomes widely available, many researchers have resorted to using daily squared return, calculated from market closing prices, to proxy daily volatility. As shown in Lopez (2001), while ε_t^2 is an unbiased estimator of σ_t^2, it is very imprecise due to its asymmetric distribution. Let

$$Y_t = \mu + \varepsilon_t , \qquad \varepsilon_t = \sigma_t z_t , \tag{13}$$

and $z_t \sim N\left(0,1\right)$. Then

$$E\left[\varepsilon_t^2\,\middle|\,\Phi_{t-1}\right] = \sigma_t^2 E\left[z_t^2\,\middle|\,\Phi_{t-1}\right] = \sigma_t^2$$

since $z_t^2 \sim \chi_{(1)}^2$. However, since the median of a $\chi_{(1)}^2$ distribution is 0.455, $\varepsilon_t^2 < \frac{1}{2}\sigma_t^2$ more than 50% of the time. In fact

$$\Pr\left(\varepsilon_t^2 \in \left[\frac{1}{2}\sigma_t^2, \frac{3}{2}\sigma_t^2\right]\right) = \Pr\left(z_t^2 \in \left[\frac{1}{2}, \frac{3}{2}\right]\right) = 0.2588 \,,$$

which means that ε_t^2 is 50% greater or smaller than σ_t^2 by nearly 75% of the time.

Under the null hypothesis that returns in (13) is generated by a GARCH(1,1) process, Andersen and Bollerslev (1998) show that the population R^2 for the regression

$$\varepsilon_t^2 = \alpha + \beta \widehat{\sigma}_t^2 + \upsilon_t$$

is equal to κ^{-1} where κ is the kurtosis of the standardized innovations and κ is finite. For conditional Gaussian error, the R^2 from a correctly specified GARCH(1,1) model is bounded from above by $\frac{1}{3}$. Christodoulakis and Satchell (1998) extend the results to include compound normals and the Gram-Charlier class of distributions and show that the misestimation of forecast performance is likely to be worsened by non-normality known to present in financial data.

Hence, the use of ε_t^2 as volatility proxy will lead to low R^2 and undermine the inference regarding forecast accuracy. Blair, Poon and Taylor (2001) report an increase of R^2 by 3 to 4 times for the one-day ahead forecast when intra-day 5-minutes square returns instead of daily square returns are used to proxy the actual volatility. Extra caution is called for when interpreting empirical findings in studies that adopt such a noisy volatility estimator.

4.5 Further issues in forecast evaluation

In all forecast evaluations, it is important to distinguish in-sample and out-of-sample forecasts. In-sample forecast, which is based on parameters estimated using all data in the sample, implicitly assumes parameter estimates are stable through time. In practice, time variation of parameter estimates is a critical issue in forecasting. A good forecasting model should be one that can withstand the robustness of out-of-sample test; a test design that is closer to reality. In our analyses of empirical findings in Sections 5 and 6, we focus our attention only to studies that implement out-of-sample forecasts.

An issue not addressed previously is whether volatility X in (9), (10), (11) and (12) should be standard deviation or variance? The complication generated by this choice could be further compounded with the choice of performance measure (e.g. MAE or MSE). The square of a variance error is the 4th power of the same error measured from standard deviation. This can complicate the task of forecast evaluation, given the difficulty in estimating fourth moments with common distributions let alone the thick tailed ones in finance. The confidence interval of the mean error statistic can be very wide when forecast errors are measured from variances and worse if they are squared.

This leads to difficulty in finding significant differences between alternative forecasting methods. For this reason, one may even consider using a logarithmic transformation (as in Pagan and Schwert (1990)) to reduce the impact of outliers.

Davidian and Carroll (1987) make similar observations in their study of variance function estimation for heteroskedastic regression. Using high order theory, they show that the use of square returns for modelling variance is appropriate only for approximately normally distributed data, and becomes non-robust when there is a small departure from normality. Estimation of the variance function that is based on logarithmic transformation or absolute returns is more robust against asymmetry and non-normality. More recently, Andersen, Bollerslev, Diebold and Labys (2001) and Andersen, Bollerslev, Diebold and Ebens (2001) find realized volatility estimated from high frequency currency and stock returns are approximately lognormal. These findings are generally consistent with X being logarithmic volatility.

Bollerslev and Ghysels (1996) further suggest a heteroskedasticity-adjusted version of MSE called HMSE where

$$\text{HMSE} = \frac{1}{N} \sum_{t=1}^{N} \left[\frac{X_{T+t}}{\widehat{X}_{T+t}} - 1 \right]^2$$

In this case, the forecast error is effectively scaled by actual volatility. This type of performance measure is not appropriate if the absolute magnitude of the forecast error is a major concern.

5   Volatility forecasting based on time series models

In this section, we review major findings in 44 papers that construct volatility forecasts based on historical information only. We will make some references to implied volatility forecasts when we discuss forecasting performance of SV and long memory volatility models. Main findings regarding implied volatility forecasts will be discussed in Section 6.

5.1   Pre-ARCH era and non-ARCH debate

Taylor (1987) is one of the earliest to test time series volatility forecasting models before ARCH/GARCH permeates the volatility literature. Taylor (1987) studies the use of high, low and closing prices to forecast 1 to 20 days DM/$ futures volatility and find a weighted average composite forecast to perform best. Wiggins (1992) also gives support to extreme value volatility estimators.

In the pre-ARCH era, there are many other findings covering a wide range of issues. Dimson and Marsh (1990) find ex ante time-varying optimized weighting schemes do not always work well in out-of-sample forecasts. Sill (1993) finds S&P 500 volatility is

higher during recession and that commercial T-Bill spread helps to predict stock market volatility. Alford and Boatman (1995) find, from a sample of 6,879 stocks, that adjusting historical volatility towards volatility estimates of comparable firms in the same industry and size provides a better 5-year ahead volatility forecast. Alford and Boatman (1995), Figlewski (1997) and Figlewski and Green (1999) all stress the importance of having a long enough estimation period to make good volatility forecast over a long horizon.

5.2 The explosion of ARCH/GARCH forecasting contests

Akigray (1989) is one of the earliest to test predictive power of GARCH and is commonly cited in many GARCH studies later, although an earlier investigation has appeared in Taylor (1986). In the following decade, there were no less than 20 papers testing GARCH predictive power against other time series methods and against option implied volatility forecasts. Majority of these forecast volatility of major stock indices and exchange rates.

The ARCH/GARCH models, and their variants, have many supporters. Akgiray finds GARCH consistently outperforms EWMA and HIS (i.e. historical volatility derived from standard deviation of past returns over a fixed interval) in all subperiods and under all evaluation measures. Pagan and Schwert (1990) find EGARCH is best especially in contrast to nonparametric methods. Despite a low R^2, Cumby, Figlewski and Hasbrouck (1993) conclude that EGARCH is better than naïve historical methods. Figlewski (1997) finds GARCH superiority confines to stock market and for forecasting volatility over short horizon only. Cao and Tsay (1992) find TAR provides the best forecast for large stocks and EGARCH gives the best forecast for small stocks, and they suspect that the latter might be due to a leverage effect. Bali (2000) documents the usefulness of GARCH models, the nonlinear ones on particular, in forecasting 1-week ahead volatility of US T-Bill yields.

Other studies find no clear-cut result. These include Lee (1991), West and Cho (1995), Brooks (1998), and McMillan, Speight and Gwilym (2000). Some models work best under different error statistics (e.g. MAE, MSE), different sampling schemes (e.g. rolling fixed sample estimation, or recursive expanding sample estimation), different time periods and for different assets. Brailsford and Faff (1996) comment that the GJR-GARCH model has a marginally lead while Franses and Van Dijk (1996) claim the GJR forecast cannot be recommended. All these studies (and many other volatility forecasting studies) share one or more of the following characteristics: (i) they test a large number of very similar models all designed to capture volatility persistence, (ii) they use a large number of error statistics each of which has a very different loss function, (iii) they forecast and calculate error statistics for variance and not standard deviation, which makes the difference between forecasts of different models even smaller, (iv) they use squared daily, weekly or monthly returns to proxy daily, weekly or monthly "actual volatility", which result in extremely

noisy volatility estimates. The noise in the volatility estimates makes the small differences between forecasts of similar models indistinguishable.

Unlike the ARCH class model, the "simpler" methods, including the EWMA method, do not separate volatility persistence from volatility shocks and most of them do not incorporate volatility mean reversion. The GJR model allows the volatility persistence to change relatively quickly when return switches sign from positive to negative and *vice versa*. If unconditional volatility of all parametric volatility models is the same, then GJR will have the largest probability of an underforecast. The "simpler" methods tend to provide larger volatility forecasts most of the time because there is no constraint on stationarity or convergence to the unconditional variance, and may result in larger forecast errors.[11] This possibly explains why GJR was the worst performing model in Franses and Van Dijk (1996) because they use MedSE (median standard error) as their sole evaluation criteria. In Brailsford and Faff (1996), the GJR(1,1) model outperforms the other models when MAE, RMSE and MAPE are used.

There are some merits in using "simpler" methods, and especially models that include long distributed lags. As ARCH class models assume variance stationarity, the forecasting performance suffers when there are changes in volatility level. Parameters estimation becomes unstable when data period is short or when there is a change in volatility level. This has led to GARCH convergence problem in several studies (e.g. Tse and Tung (1992) and Walsh and Tsou (1998)). Taylor (1986), Tse (1991), Tse and Tung (1992), Boudoukh, Richardson and Whitelaw (1997) and Walsh and Tsou (1998), Ederington and Guan (1999), Ferreira (1999), and Taylor JW (2001) all favour some forms of exponential smoothing method to GARCH for forecasting volatility of a wide range of assets across equities, exchange rates and interest rates.

In general, models that allow for volatility asymmetry came out well in the forecasting contest because of the strong negative relationship between volatility and shock. Cao and Tsay (1992), Heynen and Kat (1994), Lee (1991) and Pagan and Schwert (1990) favour the EGARCH model for volatility of stock indices and exchange rates, whereas Brailsford and Faff (1996) and Taylor (2001) find GJR-GARCH to outperform GARCH in stock indices. Bali (200) find a range of nonlinear models work well for interest rate volatility. The fact that the sign of future return shocks is unknown could be a contributing factor to GJR's poorer performance in multi-step ahead forecasts. QGARCH, LSTGARCH and STES-EAE (see Appendix for details) are asymmetry volatility models that do not require the prediction of the sign of future return shocks and have been shown to work well.

---

[11] This characteristic is clearly evidenced in Table 2 of Brailsford and Faff (1996). The GJR(1,1) model under-forecasts 76 (out of 90) times. The RW model has an equal chance of under- and over-forecasts, whereas all the other methods overforecast more than 50 (out of 90) times.

5.3 The arrival of SV forecasts

The SV model has an additional innovative term in the volatility dynamics and, hence, is more flexible than ARCH class models and was found to fit financial market returns better and have residuals closer to standard normal. It is also closer to theoretical models in finance and especially those in derivatives pricing. However, largely due to the computation difficulty, volatility forecast based on SV model was not studied till mid 1990's, a decade later than ARCH/GARCH development. In a PhD thesis, Heynen (1995) finds SV forecast is best for a number of stock indices cross several continents. There are only six other SV studies and the view about SV forecasting performance is by no means unanimous at the time of writing.

Heynen and Kat (1994) forecast volatility for seven stock indices and five exchange rates and find SV provides the best forecast for indices but produces forecast errors that are ten times larger than EGARCH's and GARCH's for exchange rates. Yu (2002) ranks SV top for forecasting New Zealand stock market volatility, but the margin is very small, partly because the evaluation is based on variance and not standard deviation. Lopez (2001) finds no difference between SV and other time series forecasts using conventional error statistics. All three papers have the 1987's crash in the in-sample period, and the impact of the 1987 crash on the result is unclear.

Three other studies, Bluhm and Yu (2000), Dunis, Laws and Chauvin (2000) and Hol and Koopman (2002) compare SV and other time series forecasts with option implied volatility forecast. Dunis et.al. (2000) find combined forecast is the best for six exchange rates so long as the SV forecast is excluded. Bluhm and Yu (2000) rank SV equal to GARCH. Both Bluhm and Yu (2000) and Hol and Koopman (2002) conclude that implied is better than SV for forecasting stock index volatility.

5.4 Recent development in long memory volatility models

Volatility forecasts based on models that exploit the long memory (LM) characteristics of volatility appear rather late in the literature. These include Andersen, Bollerslev, Diebold and Labys (2002), Vilasuso (2002), Zumbach (2002) and three other papers that compare LM forecasts with option implied volatility, viz. Li (2002), Martens and Zein (2002) and Pong, Shakleton, Taylor and Xu (2002). We have pointed out in Section 3.1.2 that other short memory models (e.g. extreme values, breaks, mixture of distribution, and regime witching) are also capable of producing long memory in second moments and each of them entails a different data generating process. At the time of writing, there is no direct contest between these and the LM models.

An earlier LM paper by Hwang and Satchell (1998) uses LM models to forecast Black-

Scholes implied volatility of equity option. This paper contains some useful insights about properties of LM models, but since we are focusing on forecasting volatility of the underlying asset, the results of Hwang and Satchell (1998) will not be discussed here.

Examples of LM models include the FIGARCH in Baillie, Bollerslev and Mikkelsen (1996) and FIEGARCH in Bollerslev and Mikkelsen (1996). In Andersen, Bollerslev, Diebold and Labys (2002) a vector autoregressive model with long distributed lags was build on realized volatility of three exchange rates, which they called the VAR-RV model. In Zumbach (2002) the weights apply to time series of realized volatility follow a power law, which he called the LM-ARCH model.

As noted before in Section 3.1.2, all fractional integrated models of volatility have a non-zero drift. In practice the estimation of fractional integrated models require an arbitrary truncation of the infinite lags and as a result, the mean will be biased. Zumbach's (2002) LM-ARCH will not have this problem because of the fixed number of lags and the way in which the weights are calculated. Hwang and Satchell's (1998) scaled-truncated log-ARFIMA model is mean adjusted to control for the bias that is due to this truncation and the log transformation.

Vilasuso (2002) finds FIGARCH produces significantly better 1- and 10-day ahead volatility forecasts for five major exchange rates. Zumbach (2002) produces only 1-day ahead forecasts and find no difference among model performance. Andersen, Bollerslev, Diebold and Labys (2002) find the realized volatility constructed VAR model, i.e. VAR-RV, produces the best 1- and 10-day ahead volatility forecasts. It is difficult to attribute this superior performance to LM model alone because the VAR structure allows a cross series linkage that is absence in all other univariate models and we also know that the more accurate realized volatility estimates would result in improved forecasting performance everything else equal.

The other three papers that compare forecasts from LM models with implied volatility forecasts generally find implied volatility forecast to produce highest explanatory power. Martiens and Zein (2002) find log-ARFIMA forecast beats implied in S&P 500 futures but not in ¥/US$ and crude oil futures. Li (2002) finds implied produces better short horizon forecast whereas the ARFIMA provides better forecast for a six-month horizon. However, when regression coefficients are constrained to be $\alpha = 0$ and $\beta = 1$, the regression R^2 becomes negative at long horizon. From our discussion in Section 4.3, this suggests that volatility at the six-month horizon might be better forecast using the unconditional variance instead of model based forecasts.

As all LM papers in this group were written very recently and after the publication of Andersen and Bollerslev (1998), the realized volatilities are constructed from intraday high frequency data. When comparison is made with implied forecast, the implied

volatility is usually extracted from daily closing option prices however. Despite the lower data frequency, implied appears to outperform forecasts from LM models.

5.5 Regime switching models

It has long been argued that financial market reacts to large and small shocks differently and the rate of mean reversion is faster for large shocks. Friedman and Laibson (1989), Jones, Lamont and Lumsdaine (1998) and Ederington and Lee (2001) all provide explanations and empirical support for the conjecture that volatility adjustment in high and low volatility states follow a twin-speed process; slower adjustment and more persistent volatility in low volatility state and faster adjustment and less volatility persistence in high volatility state.

One approach for modelling changing volatility persistence is to use a Hamilton (1989) type regime switching (RS) model, which like GARCH model is strictly stationary and co-variance stationary. The TAR model used in Cao and Tsay (1992) is similar to a SV model with regime switching, and Cao and Tsay (1992) prefers TAR to EGARCH and GARCH. The earlier RS applications, such as Pagan and Schwert (1990) and Hamilton and Susmel (1994) tend to be more rigid, where conditional variance is state dependent but not time dependent. Until recently, only ARCH class conditional variance is permitted. Recent extensions by Gray (1996) and Klaassen (1998) allow GARCH type heteroskedasticity in each state and the probability of switching between states to be time dependent.

Hamilton and Susmel (1994) find regime switching ARCH with leverage effect produces better volatility forecast than asymmetry version of GARCH. Hamilton and Lin (1996) use a bivariate RS model and find stock market returns are more volatile during recession period. Gray (1996) fits a RSGARCH (1,1) model to US 1-month T-Bill rates, where the rate of mean level reversion is permitted to differ under different regimes, and find substantial improvement in forecasting performance. Klaassen (1998) also applies RSGARCH (1,1) to the foreign exchange market and find a superior, though less dramatic, performance.

It is worth noting that interest rates are different to the other assets in that interest rates exhibit "level" effect i.e. volatility depends on the level of the interest rate. It is plausible that it is this level effect that Gray (1996) is picking up that result in superior forecasting performance. This level effect also appears in some European short rates (Ferreira (1999)). There is no such level effect in exchange rates and so it is not surprising that Klaassen (1998) does not find similar dramatic improvement.[12]

---

[12] We thank a referee for this important insight.

5.6 Extreme Values and outliers

There are at least two stylized facts about volatility in the financial markets that were not captured by ARCH models: (i) The standardized residuals from ARCH models still display large kurtosis (see McCurdy and Morgan (1987), Milhoj (1987), Hsieh (1989) and Baillie and Bollerslev (1989)). Conditional heteroskedasticity alone could not account for all the tail thickness. This is true even when the Student-t distribution is used to construct the likelihood function (see Bollerslev (1987) and Hsieh (1989)). (ii) ARCH effect significantly reduced or disappeared once large shocks are controlled for (Aggarwal, Inclan and Leal (1999)). Franses and Ghijsels (1999) find forecasting performance of the GARCH model is substantially improved in 4 out of 5 stock markets studied when the additive outliers are removed.

Diebold and Pauly (1987) and Lamoureux and Lastrapes (1990) show that the high volatility persistence in the GARCH model could be due to structural changes in the variance process. A shift in unconditional variance will result in volatility persistence in GARCH that assumes covariance stationarity. Kearns and Pagan (1993) investigate the issue if volatility persistence was an artifact of extremes or outliers by symmetrically trimming the scores of the largest observations but find volatility persistence remained in Australian stock index returns. On the other hand, Aggarwal, Inclan and Leal (1999) use Inclan and Tiao (1994) method to identify and adjust for volatility level changes and find a big difference after these level shifts are controlled for. The difference between the two approaches is tenuous however. As the time between volatility level changes gets smaller, the second approach converges to the first.

At the time of writing, there is no consensus about the treatment of extreme values and outliers; whether they should simply be removed, trimmed or their impact on volatility be separately handled. The financial market literature is also rather loose in its terminology regarding outliers and extremes, which can lead to opaque discussions. To a statistician, there are two extremes in each sample, the minimum and the maximum, although this can extend to the first few ordinal statistics at each end. It follows that the number of extremes does not increase with sample size, n. The number of terms in the 5% tail of the distribution does increase with sample size, being $0.05n$, so that tails and extremes are not identical concepts; extremes lie in the tails but tails include data that are not extremes. Of course, for small samples, the two sets become almost identical, but they are quite different for large samples. Tails are part of the central distributions but for large samples true extremes can be considered to be drawn from an extreme-value distribution. In the far tail the two could be the same. "Outliers" could be drawn from a quite different distribution when the market goes into a different mode, in which case the observed distribution is a

mixture of the two. An obvious problem is that there are rather few observations from the outlier distributions and so its estimation is difficult.

The volume-volatility literature has documented a strong link between contemporaneous trading volume and conditional volatility. It is plausible that residuals that are scaled by trading volume might be approximately Gaussian. No doubt, the Christmas wish list of volatility forecasters must be for a method of forecasting crashes, even if it is for a short period ahead. It may be possible to find such a method using options or very high frequency data, but a great deal of further exploration is required.

5.7 Getting the right conditional variance and forecast with the "wrong" models

Many of the time series volatility models including the GARCH models can be thought of as approximating a deeper time-varying volatility construction, possibly involving several important economic explanatory variables. Since time series models involve only lagged returns but have considered many forms of specification, it seems likely that they will provide an adequate, possibly even a very good approximation to actuality for long periods but not at all times. This means that they will forecast well on some occasions, but less well on others, depending on fluctuations in the underlying driving variables.

Nelson (1992) proves that if the true process is a diffusion or near-diffusion model with no jumps, then even when misspecified, appropriately defined sequences of ARCH terms with a large number of lagged residuals may still serve as consistent estimators for the volatility of the true underlying diffusion, in the sense that the difference between the true instantaneous volatility and the ARCH estimates converges to zero in probability as the length of the sampling frequency diminishes. Nelson (1992) shows that such ARCH models may misspecify both the conditional mean and the dynamic of the conditional variance; in fact the misspecification may be so severe that the models make no sense as data generating processes, they could still produce consistent one-step-ahead conditional variance estimates and short term forecasts.

Nelson and Foster (1995) provide further conditions for such misspecified ARCH models to produce consistent forecasts over medium- and long-term. They show that forecasts of the process and its volatility generated by these misspecified models will converge in probability to the forecast generated by the true diffusion or near diffusion process provided that all unobservable state variables are consistently estimated and that the conditional mean and conditional covariances of all state variables are correctly specified. An example of a true diffusion process given by Nelson and Foster (1995) is the stochastic volatility model described in Section 3.2.2.

These important theoretical results confirm our empirical observations that under normal circumstances, i.e. no big jumps in prices, there may be little practical difference in choosing between volatility models provided that the sampling frequency is small and that whichever model one has chosen, it must contains long enough lagged residuals. This might be an explanation for the success of high frequency and long memory volatility models (e.g. Blair, Poon and Taylor (2001) and Andersen, Bollerslev, Diebold and Labys (2002)).

6 Volatility forecasting based on option ISD

In contrast to time series volatility forecasting models described in Section 6, the use of option ISD (Implied Standard Deviation) as a volatility forecast involves some extra complexities. A test on the forecasting power of option ISD is a joint test of option market efficiency and a correct option pricing model. Since trading frictions differ across assets, some options are easier to replicate and hedge than the others. It is therefore reasonable to expect different levels of efficiency and different forecasting power for options written on different assets. We will focus on this aspect of the forecasting contest in Section 6.1 and compare implied and time series volatility forecasts within each asset class.

While each historical price constitutes an observation in the sample used in calculating volatility forecast, each option price constitutes a volatility forecast over the option maturity, and there can be many option prices at any one time. As mentioned in Section 3.2, there are are also the problem of volatility smile and volatility skew. Options of different strike prices produce different Black-Scholes implied volatility estimates. Section 6.2 discusses the information content of ISD across strikes and the effectiveness of different weighting schemes used to produce an implied volatility composite for forecasting.

The issue of a correct option pricing model is more fundamental in finance. Option pricing has a long history and various extensions have been made since Black-Scholes to cope with dividend payments, early exercise and stochastic volatility. However, none of the option pricing models (except Heston (1993)) that appeared in the volatility forecasting literature allows for a premium for bearing volatility risk. In the presence of a volatility risk premium, we expect the option price to be higher which means implied volatility derived using an option pricing model that assumes zero volatility risk premium (such as the Black-Scholes model) will also be higher, and hence automatically be more biased. Section 6.3 examines the issue of biasness of ISD forecasts and evaluates the extent to which implied biasness is due to the omission of volatility risk premium.

6.1 Predictability across different assets

As noted in Section 3.2.3, early studies that test forecasting power of option ISD are fraught with many estimation deficiencies. Despite these complexities, option ISD has been found empirically to contain a significant amount of information about future volatility and it often beats volatility forecasts produced by sophisticated time series models. Such a superior performance appears to be common across assets.

6.1.1 Individual stocks

Latane and Rendleman (1976) were the first to discover the forecasting capability of option ISD. They find actual volatilities of 24 stocks calculated from in-sample period and extended partially into the future are more closely related to implied than historical volatility. Chiras and Manaster (1978) and Beckers (1981) find prediction from implied can explain a large amount of the cross-sectional variations of individual stock volatilities. Chiras and Manaster (1978) document an R^2 of 34-70% for a large sample of stock options traded on CBOE whereas Beckers (1981) reports R^2 of 13-50% for a sample that varies from 62 to 116 U.S. stocks over the sample period. Gemmill (1986) produces R^2 of 12-40% for a sample of 13 U.K. stocks. Schmalensee and Trippi (1978) find implied volatility rises when stock price falls and that implied volatilities of different stocks tend to move together. From a time series perspective, Lamoureux and Lastrapes (1993) and Vasilellis and Meade (1996) find implied could also predict time series variations of equity volatility better than forecasts produced from time series models.

The forecast horizons of this group of studies are usually quite long ranging from 3 months to 3 years. Studies that examine incremental information content of time series forecasts find volatility historical average provides significant incremental information in both cross-sectional (Beckers (1981), Chiras and Manaster (1978), Gemmill (1986)) and time series settings (Lamoureux and Lapstrapes (1993)) and that combining GARCH and implied produces the best forecast (Vasilellis and Meade (1996)). These findings have been interpreted as an evidence of stock option market inefficiency since option implied does not subsume all information. In general, stock option implied exhibits instability and suffers most from measurement errors and bid-ask spread because of the lower liquidity.

6.1.2 Stock market index

There are 22 studies that use index option ISD to forecast stock index volatility; seven of these forecast volatility of S&P 100, ten forecast volatility of S&P 500 and the remaining five forecast index volatility of smaller stock markets. The S&P 100 and S&P 500 forecasting results make an interesting contrast as almost all studies that forecast S&P 500

volatility use S&P 500 futures options which is more liquid and less prone to measurement errors than the OEX stock index option written on S&P 100. We will return to the issue of measurement errors when we discuss biasness in Section 6.3.

All but one study (viz. Canina and Figlewski (1993)) conclude that implied contains useful information about future volatility. Blair, Poon and Taylor (2001) and Poteshman (2000) record the highest R^2 for S&P 100 and S&P 500 respectively. About 50% of index volatility is predictable up to a 4-week horizon when actual volatility is estimated more accurately using very high frequency intra-day returns.

Similar, but less marked, forecasting performance emerged from the smaller stock markets, which include the German, Australian, Canadian and Swedish markets. For a small market such as the Swedish market, Frennberg and Hanssan (1996) find seasonality to be prominent and that implied forecast cannot beat simple historical models such as the autoregressive model and random walk. Very erratic and unstable forecasting results were reported in Brace and Hodgson (1991) for the Australian market. Doidge and Wei (1998) find the Canadian Toronto index is best forecast with GARCH and implied combined, whereas Bluhm and Yu (2000) find VDAX, the German version of VIX, produces the best forecast for the German stock index volatility.

A range of forecast horizons were tested among this group of studies though the most popular choice is 1 month. There is evidence that the S&P implied contains more information after the 1987 crash (see Christensen and Prabhala (1998) for S&P 100 and Ederington and Guan (2002) for S&P 500). Some described this as the "awakening" of the S&P option markets.

About half of the papers in this group test if there is incremental information contained in time series forecasts. Day and Lewis (1992), Ederington and Guan (1999, 2002), and Martin and Zein (2002) find ARCH class models and volatility historical average add a few percentage points to the R^2, whereas Blair, Poon and Taylor (2001), Christensen and Prabhala (1998), Fleming (1998), Fleming, Ostdiek and Whaley (1995), Hol and Koopman (2002) and Szamany, Ors, Kim and Davidson (2002) all find option implied dominates time series forecasts.

6.1.3 Exchange rate

The strong forecasting power of implied is again confirmed in the currency markets. Sixteen papers study currency options for a number of major currencies, the most popular of which are DM/US$ and ¥/US$. Most studies find implied to contain information about future volatility for short horizon up to 3 months. Li (2002) and Scott and Tucker (1989) find implied forecast well for up to a 6- to 9-months horizon. Both studies register the highest R^2 in the region of 40-50%.

A number of studies in this group find implied beats time series forecasts including volatility historical average (see Fung, Lie and Moreno (1990) and Wei and Frankel (1991)) and ARCH class models (see Guo (1996a, 1996b), Jorion (1995, 1996), Martens and Zein (2002), Pong, Shackleton, Taylor and Xu (2002), Szamany, Ors, Kim and Davidson (2002) and Xu and Taylor (1995)). Some studies find combined forecast is the best choice (see Dunis, Law and Chauvin (2000) and Taylor and Xu (1997)).

Two studies find high frequency intra day data can produce more accurate time series forecast than implied. Fung and Hsieh (1991) find 1-day ahead time series forecast from a long-lag autoregressive model fitted to 15-minutes returns is better than implied. Li (2002) finds ARFIMA model outperformed implied in long horizon forecasts while implied dominates over shorter horizons. Implied forecasts were found to produce higher R^2 than other long memory models, such as the Log-ARFIMA model in Martens and Zein (2002) and Pong, Shackleton, Taylor and Xu (2002). All these long memory forecasting models are more recent and are built on volatility compiled from high frequency intra-day returns while the implied volatility remains to be constructed from less frequent daily option prices.

6.1.4   Other assets

The forecasting power of implied from interest rate options was tested in Edey and Elliot (1992), Fung and Hsieh (1991) and Amin and Ng (1997). Interest rate option models are very different other option pricing models because of the need to price the whole term structure of interest rate derivatives consistently all at the same time in order to rule out arbitrage opportunities. Trading in interest rate instruments is highly liquid as trading friction and execution cost are negligible. Practitioners are more concerned about the term structure fit than the time series fit, as million of pounds of arbitrage profits could change hands instantly if there is any inconsistency in contemporaneous prices.

Earlier studies such as Edey and Elliot (1992) and Fung and Hsieh (1991) use the Black model (a modified version of Black-Scholes) that prices each interest rate option without cross referencing to prices of other interest rate derivatives. The single factor Heath-Jarrow-Morton model used in Amin and Ng (1997) and fitted to short rate only works in the same way although the authors have added different constraints to the short rate dynamics as the main focus of their paper is to compare different variants of short rate dynamics. Despite the complications, all three studies find significant forecasting power in implied of interest rate (futures) options. Amin and Ng (1997) in particular report R^2 of 21% for 20-day ahead volatility forecasts, and volatility historical average adds only a few percentage points to the R^2.

Implied volatilities from options written on non-financial assets were examined in

Day and Lewis (1993, crude oil), Kroner, Kneafsey and Claessens (1995, agriculture and metals), Martens and Zein (2002, crude oil) and a recent study (Szamany, Ors, Kim and Davidson (2002)) that covers 35 futures options contracts across 9 markets including S&P500, interest rates, currency, energy, metals, agriculture and livestock futures. All four studies find implied dominates time series forecasts although Kroner, Kneafsey and Claessens (1995) find combining GARCH and implied produces the best forecast.

6.2 ATM (at-the-money) or weighted implied?

Since options of different strikes have been known to produce different implied volatilities, a decision has to be made as to which of these implied volatilities should be used, or which weighting scheme should be adopted, that will produce a forecast that is most superior. The most common strategy is to choose the implied derived from ATM option based on the argument that ATM option is the most liquid and hence ATM implied is least prone to measurement errors. ATM implied is also theoretically most sound. Feinstein (1989a) shows that for the stochastic volatility process described in Hull and White (1987), implied volatility from ATM and near expiration option provides the closest approximation to the average volatility over the life of the option provided that volatility risk premium is either zero or a constant. This means that if volatility is stochastic, ATM implied is least prone to bias as well compared with implied at other strikes.

If ATM implied is not available, then NTM (nearest-to-the-money) option is used instead. Sometimes, to reduce measurement errors and the effect of bid-ask bounce, an average is taken from a group of NTM implied volatilities. VIX described in Section 3.2.5, for example, is an average of eight implied volatilities derived from four calls and four puts and a weighting scheme that aim to produce a composite implied that is of a constant maturity of 28 calendar days and approximately ATM. Other weighting schemes that also give greater weight to ATM implied are vega (i.e. the partial derivative of option price w.r.t. volatility) weighted or trading volume weighted, weighted least square (WLS) and some multiplicative versions of the these three. The WLS method, first appeared in Whaley (1982), aims to minimize the sum of squared errors between the market and the theoretical prices of a group of options. Since ATM option has the highest trading volume and ATM option price is the most sensitive to volatility input, all three weighting schemes (and the combinations thereof) have the effect of placing the greatest weight on ATM implied. Other less popular weighting schemes include equally weighted, and weight based on the elasticity of option price to volatility.

The forecasting power of individual and composite implied volatilities has been tested in Ederington and Guan (2000), Fung, Lie and Moreno (1990), Gemmill (1986), Kroner, Kneafsey and Classens (1995), Scott and Tucker (1989) and Vasilellis and Meade (1996).

The general consensus is that among the weighted implied volatilities, those that have a VIX style composite weight seem to be the best, followed by schemes that favour ATM option such as the WLS and the vega weighted implied. The worst performing ones are equally weighted and elasticity weighted implied using options across all strikes. Different findings emerged as to whether an individual implied forecasts better than a composite implied. Becker (1981) Feinstein (1989b), Fung, Lie and Moreno (1990) and Gemmill (1986) find evidence to support individual implied although they all prefer a different implied (viz. ATM, Just-OTM, OTM and ITM respectively for the four studies). Kroner, Kneafsey and Clasessens find composite implied forecasts better than ATM implied. On the other hand, Scott and Tucker (1989) conclude that when emphasis is placed on ATM implied, which weighting scheme one chooses does not really matter.

As mentioned in the Section 6.1.1, implied volatility, especially that of stock option, can be quite unstable across time. Beckers (1981) finds taking a 5-day average improves forecasting power of stock option implied. Hamid (1998) finds such an intertemporal averaging is also useful for stock index option during very turbulent periods. On a slightly different note, Xu and Taylor (1995) find implied estimated from sophisticated volatility term structure model produces similar forecasting performance as implied from the shortest maturity option.

A series of studies by Ederington and Guan have reported some interesting findings. Ederington and Guan (1999) report that information content of implied volatility of S&P 500 futures options exhibits a frown shape across strikes with options that are NTM and have moderately high strike (i.e. OTM calls and ITM puts) possess the largest information content with R^2 equal to 17% for calls and 36% for puts. In a follow-on paper, Ederington and Guan (2000) find that using regression coefficients that are produced from in-sample regression of forecast against realised volatility is very effective in correcting implied forecasting bias. They also find that after such a bias correction, there is little to be gained from averaging implied across strikes. This means that ATM implied together with a bias correction scheme could be the simplest, and yet the best, way forward.

Findings in Ederington and Guan (1999, 2000) raise a very profound issue in finance. So far the volatility forecasting literature has been relying heavily on ATM implied. Implied from option that are far away from ATM were found to be bad forecasts and are persistently biased. This reflects a crucial fact that we have not yet found an option pricing model that is capable of pricing far-from-the-money option accurately and consistently. Although this is a tremendously important issue in finance, it is a research question different from the one we are addressing here, and we shall leave it as a challenge for future research.

6.3 Implied biasness

Usually, forecast unbiasedness is not an overriding issue in any forecasting exercise. Forecast bias can be estimated and corrected if the degree of bias remains stable through time. However, biasness in implied volatility can have a more serious undertone since it means option might be over- or under-priced which can only be a result of an incorrect option pricing model or an inefficient option market. Both deficiencies have important implications in finance.

As mentioned in Section 4.3, testing for biasness is usually carried using regression equation (11), where $\widehat{X}_i = \widehat{X}_t$ is the implied forecast of period t volatility. For a forecast to be unbiased, one would require $\alpha = 0$ and $\beta = 1$. Implied forecast is upwardly biased if $\alpha > 0$ and $\beta = 1$, or $\alpha = 0$ and $\beta > 1$. In the case where $\alpha > 0$ and $\beta < 1$, which is the most common scenario, implied under-forecasts low volatility and over-forecasts high volatility.

It has been argued that implied bias will persist only if it is difficult to perform arbitrage trades that may remove the mispricing. This is more likely in the case of stock index option and less likely for futures options. Stocks and stock options are traded in different markets. Since trading of a basket of stocks is cumbersome, arbitrage trades in relation to a mispriced stock index option may have to be done indirectly via index futures. On the other hand, futures and futures options are traded alongside each other. Trading in these two contracts are highly liquid. Despite these differences in trading friction, implied biasness is reported in both the S&P 100 OEX market (Canina and Figlewski (1993), Christensen and Prabhala (1998), Fleming, Ostdiek & Whaley (1995) and Fleming (1998)) and the S&P 500 futures options market (Feinstein (1989b) and Ederington and Guan (1999, 2002)).

Biasness is equally widespread among implied volatilities of currency options (see Guo (1996b), Jorion (1995), Li (2002), Scott and Tucker (1989) and Wei and Frankel (1991)). The only exception is Jorion (1996) who cannot reject the null hypothesis that the 1-day ahead forecasts from implied are unbiased. The five studies listed earlier use implied to forecast exchange rate volatility over a much longer horizon from one to nine months.

Unbiasness of implied forecast was not rejected in the Swedish market (Frennberg and Hansson (1996)) though we already know from Section 6.1.2 that the forecasting power of implied is not very strong anyway in this small stock market. Unbiasness of implied forecast was rejected for UK stock options (Gemmill (1986)), US stock options (Lamoureux and Lastrapes (1993)), options and futures options across a range of assets in Australia (Edey and Elliot (1992)) and for 35 futures options contracts traded over 9 markets ranging from interest rate to livestock futures (Szakmary, Ors, Kim and Davidson

37

(2002)). On the other hand, Amin and Ng (1997) find the hypothesis that $\alpha = 0$ and $\beta = 1$ cannot be rejected for the Eurodollar futures options market.

Where unbiasness was rejected, the bias in all but two cases is due to $\alpha > 0$ and $\beta < 1$. These two exceptions are Fleming (1998) who reports $\alpha = 0$ and $\beta < 1$ for S&P 100 OEX options, and Day and Lewis (1993) who find $\alpha > 0$ and $\beta = 1$ for distant term oil futures options contracts.

Christensen and Prabhala (1998) argue that implied is biased because of error-in-variable caused by measurement errors described in Section 3.2.3. Using last period implied and last period historical volatility as instrumental variables to correct for these measurement errors, Christensen and Prabhala (1998) find unbiasness cannot be rejected for implied of S&P 100 OEX option. Ederington and Guan (1999, 2002) find bias in S&P 500 futures options implied also disappeared when similar instrument variables were used.

It has been suggested to us that implied biasness could not have been caused by model misspecification or measurement errors because this has relatively small effects for ATM options, which is used in most of the studies that report implied biasness. In addition, the clientele effect cannot explain the bias either because it only affects OTM options. Research now turns to volatility risk premium as an explanation.[13]

Poteshman (2000) finds half of the bias in S&P 500 futures options implied was removed when actual volatility was estimated with a more efficient volatility estimator based on intra-day 5-minutes returns. The other half of the bias was almost completely removed when a more sophisticated and less restrictive option pricing model, i.e. the Heston (1993) model, was used. The Heston model allows volatility to be stochastic similar to the Hull-While model used in Guo (1996b) and Lamoureux and Lastrapes (1993)) who both report implied biasness. But unlike the Hull-White model, the Heston model also allows the market price of risk to be non-zero.

Further research on option volatility risk premium is currently underway in Benzoni (2001) and Chernov (2001). Chernov (2001) finds, similar to Poteshman (2000), when implied volatility is discounted by a volatility risk premium and when the errors-in-variables problems in measures of historical and realized volatility are removed, the unbiasness of VIX cannot be rejected over the sample period from 1986 to 2000. The volatility risk premium debate continues if we are able to predict the magnitude and the variations of the volatility premium and if implied from an option pricing model that permits a non-zero market price of risk will outperform time series models when all forecasts (including forecasts of volatility risk premium) are made in an *ex ante* manner.

---

[13]We thank a referee for this important insight.

7 Where next?

The volatility forecasting literature is still very active. Many more new results are expected in the near future. There are several areas where future research could seek to improve upon. First is the issue about forecast evaluation and combining forecasts of different models. It will be useful if statistical tests were conducted to test if the forecast errors from Model A are significantly smaller, in some sense, than those from Model B, and so on for all pairs. Even if Model A is found to be better than all the other models, the conclusion is NOT one should henceforth forecast volatility with Model A and ignore the other models as it is very likely that a linear combination of all the forecasts will be superior. To find the weights one can either run a regression of empirical volatility (the quantity being forecast) on the individual forecasts, or as approximation just use equal weights. Testing the effectiveness of a composite forecast is just as important as testing the superiority of the individual models, but this has not been done more often and across different data sets.

A mere plot of any measure of volatility against time will show the familiar "volatility clustering" which indicates some degree of forecastibility. The biggest challenge lies in predicting changes in volatility. If implied volatility is agreed to be the best performing forecast, on average, this is in agreement with general forecast theory, which emphasizes the use of a wider information set than just the past of the process being forecast. Implied volatility uses option prices and so potentially the information set is richer. What needs further consideration is if all of its information is now being extracted and if it could still be widened to further improve forecast accuracy especially that of long horizon forecast. To achieve this we need to understand better the cause of volatility (both historical and implied). Such an understanding will help to improve time series methods, which are the only viable methods when options, or market based forecast, are not available.

Closely related to the above is to understand the source of volatility persistence and the volume-volatility research appears to be promising in providing a framework in which volatility persistence may be closely scrutinized. The Mixture of Distribution Hypothesis (MDH) proposed by Clark (1973), the link between volume-volatility and market trading mechanism in Tauchen and Pitts (1983), and the empirical findings of the volume-volatility relationship surveyed in Karpoff (1987) are useful starting points. Given that Lamoureux and Lastrapes (1990) find volume to be strongly significant when it is inserted into the ARCH variance process, while returns shocks become insignificant, and that Gallant, Rossi and Tauchen (1993) find conditioning on lagged volume substantially attenuates the "leverage" effect, the volume-volatility research may lead to a new and better way for modelling returns distributions. To this end, Andersen (1996) puts forward a generalised

framework for the MDH where the joint dynamics of returns and volume are estimated, and reports a significant reduction in the estimated volatility persistence. Such a model may be useful for analyzing the economic factors behind the observed volatility clustering in returns but such a line of research has not yet been pursued vigorously.

There are many old issues that have been around for a long time. These include consistent forecasts of interest rate volatilities that satisfies the no-arbitrage relationship across all interest rate instruments, more tests on the use of absolute returns models in comparison with squared returns models in forecasting volatility, a multivariate approach to volatility forecasting where cross correlation and volatility spillover may be accommodated, etc.

There are many new adventures that are currently underway as well.[14] These include the realized volatility approach noticeably driven by Andersen, Bollerslev, Diebold and various co-authors, the estimation and forecast of volatility risk premium described in Section 6.3, the use of spot and option price data simultaneously (e.g., Chernov and Ghysels (2000)), and the use of Bayesian and other methods to estimate stochastic volatility models (e.g. Jones (2001)), etc.

It is difficult to envisage in which direction volatility forecasting research will flourish in the next five years. If, within the next five years, we can cut the forecast error by half and remove the option pricing bias in *ex ante* forecast, this will be a very good achievement indeed. Producing by then forecasts of large events will also be worthwhile.

8   Summary and conclusion

This survey has concentrated on two questions: is volatility forecastable? If it is, which method will provide the best forecasts? To consider these questions, a number of basic methodological viewpoints need to be discussed, mostly about the evaluation of forecasts. What exactly is being forecast? Does the time interval (the observation interval) matter? Are the results similar for different speculative markets? How does one measure predictive performance?

Volatility forecasts are classified in this section as belonging in one of the following four categories:

- HISVOL; for historical volatility, which include random walk, historical averages of squared returns, or absolute returns. Also included in this category are time series models based on historical volatility using moving averages, exponential weights, autoregressive models, or even fractionally integrated autoregressive absolute returns,

---

[14]We thank a referee for these suggestions.

for example. Note that HISVOL models can be highly sophisticated. The multivariate VAR realised volatility model in Andersen, Bollerslev, Diebold and Labys (2002) is classified here as a 'HISVOL' model. All models in this group model volatility directly omitting the goodness of fit of the returns distribution or any other variables such as option prices.

- GARCH; any member of the ARCH, GARCH, EGARCH, and so forth family are included.

- ISD; for option implied standard deviation, based on the Black-Scholes model and various generalizations.

- SV; for stochastic volatility model forecasts.

The survey of papers includes 93 studies, but 25 of them did not involve comparisons between methods from at least two of these groups, and so were not helpful for comparison purposes.

The following table involves just pair-wise comparisons. Of the 66 studies that were relevant, some compared just one pair of forecasting techniques, other compared several. For those involving both HISVOL and GARCH models, 22 found HISVOL better at forecasting than GARCH (56% of the total), and 17 found GARCH superior to HISVOL (44%). The full table is:

	Number of Studies	Studies Percentage
HISVOL > GARCH	22	56%
GARCH > HISVOL	17	44%
HISVOL > ISD	8	24%
ISD > HISVOL	26	76%
GARCH > ISD	1	6
ISD > GARCH	17	94
SV > HISVOL	3	
SV > GARCH	3	
GARCH > SV	1	
ISD > SV	1	

The combination of forecasts has a mixed picture. Two studies find it to be helpful but another does not.

The overall ranking suggests that ISD provides the best forecasting with HISVOL and GARCH roughly equal, although possibly HISVOL does somewhat better in the comparisons. The success of the implied volatility should not be surprising as these forecasts use a larger, and more relevant, information set than the alternative methods as they use option prices. They are also less practical, not being available for all assets.

Among the 93 papers, 17 studies compared alternative version of GARCH. It is clear that GARCH dominates ARCH. In general, models that incorporate volatility asymmetry such as EGARCH and GJR-GARCH, perform better than GARCH. But certain specialized specifications, such as fractionally integrated GARCH (FIGARCH) and regime switching GARCH (RSGARCH) do better in some studies. However, it seems clear that one form of study that is included is conducted just to support a viewpoint that a particular method is useful. It might not have been submitted for publication if the required result had not been reached. This is one of the obvious weaknesses of a comparison such as this; the papers being reported are being prepared for different reasons, use different data sets, many kinds of assets, various intervals between readings, and a variety of evaluation techniques. Rarely discussed is if one method is significantly better than another. Thus, although a suggestion can be made that a particular method of forecasting volatility is the best, no statement is available about the cost-benefit from using it rather than something simpler or how far ahead the benefits will occur.

Financial market volatility is clearly forecastable. The debate is on how far ahead could one accurately forecast and to what extent could volatility changes be predicted. This conclusion does not violate market efficiency since accurate volatility forecast is not in conflict with underlying asset and option prices being correct. The option implied volatility being a market based volatility forecast has been shown to contain most information about future volatility. The supremacy among historical time series models depends on the type of asset being modelled. But, as a rule of thumb, historical volatility methods work equally well compared with more sophisticated ARCH class and SV models. Better reward could be gained by making sure that actual volatility is measured accurately. These are broad brush conclusions omitting the fine details which we outline in this documents. Because of the complex issues involved and the importance of volatility measure, volatility forecasting will continue to remain as a specialist subject and be studied vigorously.

References

[1] Aggarwal R., C. Inclan and R. Leal (1999) Volatility in emerging stock markets, *Journal of Financial and Quantitative Analysis*, 34, 1, March, 33-55.

[2] Akgiray V. (1989) Conditional heteroskedasticity in time series of stock returns: evidence and forecasts, *Journal of Business*, 62, 55-80.

[3] Alford A.W. and J.R. Boatsman (1995) Predicting long-term stock return volatility: Implications for accounting and valuation of equity derivatives, *Accounting Review*, 70, 4, 599-618.

[4] Amin K. and V. Ng (1997) Inferring future volatility from the information in implied volatility in Eurodollar options: A new approach, *Review of Financial Studies*, 10, 333-367.

[5] Andersen T.G. (1996) Return volatility and trading volume: An information flow interpretation of stochastic volatility, *Journal of Finance*, 51, 1, 169-204, March.

[6] Andersen, T., and Bollerslev, T. (1998) Answering the skeptics: yes, standard volatility models do provide accurate forecasts, *International Economic Review*, 39, 4, 885-905.

[7] Andersen T.G., T. Bollerslev and S. Lange (1999) Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon, *Journal of Empirical Finance*, 6, 5, 457-477.

[8] Andersen T.G., T. Bollerslev, F.X. Diebold and H. Ebens (2001) The distribution of realized stock return volatility, *Journal of Financial Economics*, 61, 1, 43-76.

[9] Andersen T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001) The distribution of realized exchange rate volatility, *Journal of American Statistical Association*, 96, 453, 42-57, March.

[10] Andersen T.G., T. Bollerslev, F.X. Diebold and P. Labys (2002) Modeling and forecasting realized volatility, Working paper, North Western University, Duke University and University of Pennsylvania.

[11] Andersen T.G. and B.E. Sorensen (1997) GMM and QML asymptotic standard deviations in stochastic volatility models, *Journal of Econometrics*, 76, 397-403.

[12] Aydemir A.B. (1998) Volatility modelling in finance, in Knight J. and S. Satchell ed. Forecasting Volatility in the Financial Markets, Chapter 1, pp.1-46, Butterworth.

[13] Baillie R.T. and T. Bollerslev (1989) The message in daily exchange rates: a conditonal-variance tale, *Journal of Business and Economic Statistics*, 7, 3, 297-305.

[14] Baillie R.T., T. Bollerslev and H.O. Mikkelsen (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 74, 1, 3-30.

[15] Bali T.G. (2000) Testing the empirical performance of stochastic volatility models of the short-term interest rate, *Journal of Financial and Quantitative Analysis*, 35, 2, 191-215.

[16] Ball C.A. and W.N. Torous (1984) The Maximum Likelihood Estimation of Security Price Volatility: Theory, Evidence and Application to Option Pricing, *Journal of Business*, 57, 1, 97-113.

[17] Bates D.S. (1996) Testing option pricing models, in Maddala G.S. and C.R. Rao eds. Handbook of Statistics, v.14: Statistical methods in Finance, Elsevier, Amsterdam: North Holland, pp.567-611.

[18] Beckers S. (1981) Standard deviations implied in option prices as predictors of future stock price variability, *Journal of Banking and Finance*, 5, 363-382.

[19] Benzoni L. (2001) Pricing options under stochastic volatility: An empirical investigation, Working paper, Carlson School of Management, Minneapolis.

[20] Bera A.K. and M.L. Higgins (1993) ARCH models: properties, estimation and testing, *Journal of Economic Surveys*, 7, 4, 305-365.

[21] Bera, A., and M. Higgins (1997) ARCH and bilinearity as competing models for nonlinear dependence, *Journal of Business and Economic Statistics*, 15, 1, 43-50.

[22] Black F. and M. Scholes (1973) The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637-654, May/June.

[23] Blair B., S-H Poon and S.J. Taylor (2001) Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency index returns, *Journal of Econometrics*, 105, 5-26.

[24] Bluhm H.H.W. and J. Yu (2000) Forecasting volatility: Evidence from the German stock market, Working paper, University of Auckland.

[25] Bollerslev T. (1986) Generalized autoregressive conditional heteroskedasticity, *Journal of Econometricss*, 31, 307-328.

44

[26] Bollerslev T. (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics*, 69, 3, 542-547.

[27] Bollerslev T., R.Y. Chou and K.P. Kroner (1992) ARCH modeling in Finance: a review of the theory and empirical evidence, *Journal of Econometrics*, 52, 5-59.

[28] Bollerslev T., R.F. Engle and D.B. Nelson (1994) ARCH models, in Engle R.F. and D.L. McFadden ed. Handbook of Econemetrics, Volume IV, Amsterdam: North Holland, pp.2959-3038.

[29] Bollerslev T. and E. Ghysels (1996) Periodic autoregressive conditional heteroskedasticity, *Journal of Business and Economic Statistics*, 14, 2, 139-151.

[30] Bollerslev T. and H.O. Mikkelsen (1996) Modeling and pricing long memory in stock market volatility, *Journal of Econometrics*, 73, 1, 151-184.

[31] Boudoukh J., M. Richardson and R.F. Whitelaw (1997) Investigation of a class of volatility estimators, *Journal of Derivatives*, 4, 3, 63-71.

[32] Brace A. and A. Hodgson (1991) Index futures options in Australia - An empirical focus on volatility, *Accounting and Finance*, 31, 2, 13-31.

[33] Brailsford T.J. and R.W. Faff (1996) An evaluation of volatility forecasting techniques, *Journal of Banking and Finance*, 20, 3, 419-438.

[34] Brooks C. (1998) Predicting stock market volatility: Can market volume help? *Journal of Forecasting*, 17, 1, 59-80.

[35] Brown S. (1990) Estimating volatility, in Figlewski S., W.L. Silber, and M.G. Subrahmanyam ed. Financial options: from theory to practice, Chicago: Irwin.

[36] Canina, L. and S. Figlewski (1993) The informational content of implied volatility, *Review of Financial Studies*, 6, 3, 659-681.

[37] Cao C.Q. and R.S. Tsay (1992) Nonlinear time-series analysis of stock volatilities, *Journal of Applied Econometrics*, December, Supplement, 1, S,165-185.

[38] Chernov M. (2001) Implied volatilities as forecasts of future volatility, the market risk premia, and returns variability, Working paper, Columbia Business School.

[39] Chernov M. and E. Ghysels (2000) A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purposes of options valuation, *Journal of Financial Economics*, 56, 3, 407-458.

[40] Chiras D. and S. Manaster (1978) The information content of option prices and a test of market efficiency, *Journal of Financial Economics*, 6, 213-234.

[41] Cho D.C. and Frees E.W. (1988) Estimating the volatility of discrete stock prices, *Journal of Finance*, 43, 2, 451-466.

[42] Chong Y.Y. and D.F. Hendry, (1986) "Econometric evaluation of linear macroeconomic models" *Review of Economics Studies*, 53, 671-690.

[43] Christensen B.J. and N.R. Prabhala (1998) The relation between implied and realized volatility, *Journal of Financial Economics*, 50, 2, 125-150.

[44] Christodoulakis G.A. and S.E. Satchell (1998) Hashing GARCH: a re-assessment of volatility forecast and perfomance, Chapter 6, 168-192, in Knight J. and S. Satchell ed. Forecasting Volatility in the Financial Markets, Butterworth.

[45] Christoffersen P.F. and F.X. Diebold (2000) How relevant is volatility forecasting for risk management? *Review of Economics and Statistics*, 82, 1, 12-22.

[46] Clark P. (1973) A subordinated stochastic process model with finite variance for speculative prices, *Econometrica*, 41, 135-156.

[47] Cox J.C. and M. Rubinstein (1985) Options Markets, Englewood Cliffs, NJ: Prentice Hall.

[48] Cumby R., S. Figlewski and J. Hasbrouck (1993) Forecasting volatilities and correlations with EGARCH models, *Journal of Derivatives*, 1, 51-63.

[49] Danielsson J. (1994) Stochastic volatility in asset prices: Estimation with simulated maximum likelihood, *Journal of Econometrics*, 64, 375-400.

[50] Davidian M. and R.J. Carroll (1987) Variance function estimation, *Journal of American Statistical Association*, 82, 1079-1091.

[51] 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.

[52] Day T.E. and C.M. Lewis (1993) Forecasting futures market volatility, *Journal of Derivatives*, 1, 33-50.

[53] Diebold F.X. (1988) Empirical modeling of exchange rate dynamics, Springer Verlag, New York.

[54] Diebold F.X., A. Hickman, A. Inoue and T. Schuermann (1998) Scale Models, *RISK* Magazine, 11, 104-107.

[55] Diebold F.X. and A. Inoue (2001) Long memory and regime switching, *Journal of Econometrics*, 105, 1, 131-159.

[56] Diebold F.X. and J.A. Lopez (1995) Modelling volatility dynamics, in Hoover K. ed. Macroeconomics: developments, tensions and prospects, pp.427-466.

[57] Diebold F.X. and R.S. Mariano (1995) Comparing predictive accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

[58] Diebold F.X. and P. Pauly (1987) Structural-change and the combination of forecasts, *Journal of Forecasting*, 6, 1, 21-40.

[59] Dimson E. and P. Marsh (1990) Volatility forecasting without data-snooping, *Journal of Banking and Finance*, 14, 2-3, 399-421.

[60] Ding Z., C.W.J. Granger and R.F. Engle (1993) A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 1, 83-106.

[61] Doidge C. and J.Z. Wei (1998) Volatility forecasting and the efficiency of the Toronto 35 index options market, *Canadian Journal of Administrative Science*, 15, 1, 28-38.

[62] Drost F.C. and T.E. Nijman (1993) Temporal aggregation of GARCH process, *Econometrica*, 61, 4, 909-927.

[63] Duffie D. and K.J. Singleton (1993) Simulated moments estimation of Markov models of asset prices, *Econometrica*, 61, 929-952.

[64] Dunis C. L., Laws J. and S. Chauvin (2000) The Use of Market Data and Model Combination to Improve Forecast Accuracy, Working paper, Liverpool Business School.

[65] Durbin J. and S.J. Koopman (2000) Time series analysis of non-Gaussian observations based on state space models from both classical and Baysian perpectives, *Journal of Royal Statistical Society Series*, 62, 1, 3-56.

[66] Ederington L.H. and W. Guan (1999) The information frown in option prices, Working paper, University of Oklahoma.

[67] Ederington L.H. and W. Guan (2000a) Forecasting volatility, Working paper, University of Oklahoma.

[68] Ederington L.H. and W. Guan (2000b) Measuring implied volatility: Is an average better? Working paper, University of Oklahoma.

[69] Ederington L.H. and W. Guan (2002) Is implied volatility an informationally efficient and effective predictor of future volatility? *Journal of Risk*, 4, 3.

[70] Ederington L.H. and J.H. Lee (2001) Intraday volatility in interest rate and foreign exchange markets: ARCH, announcement, and seasonality effects, *Journal of Futures Markets*, 21, 6, 517-552.

[71] Edey, M., and G. Elliot (1992) Some evidence on option prices as predictors of volatility, *Oxford Bulletin of Economics & Statistics*, 54, 4, 567-578.

[72] Engle R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 4, 987-1007.

[73] Engle R.F. (1993) Statistical models for financial volatility, *Financial Analysts Journal*, January/February, 49, 1, 72-78.

[74] Engle R.F. and T. Bollerslev (1986) Modelling the persistence of conditional variances, *Econometric Reviews*, 5, 1-50.

[75] Engle R. F., V. Ng and M. Rothschild (1990) Asset pricing with a factor-ARCH covariance structure: empirical estimates for Treasury Bills, *Journal of Econometrics*, 45, 213-239.

[76] Engle R.F. and J.R. Russell (1998) Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica*, 66, 5, 1127-1162.

[77] Fair R.C. and R.J. Shiller (1989) The informational content of ex ante forecasts, *Review of Economics and Statistics,* May, 71, 2, 325-332.

[78] Fair R.C. and R.J. Shiller (1990) Comparing information in forecasts from econometric models, *American Economic Review,* June, 80, 3, 375-380.

[79] Feinstein Steven P. (1989a) The Black-Scholes formula is nearly linear in sigma for at-the-money options; Therefore implied volatilities from at-the-money options are virtually unbiased, Working paper, Federal Reserve Bank of Atlanta.

[80] Feinstein S.P. (1989b) Forecasting stock market volatility using options on index futures, *Economic Review*, Federal Reserve Bank of Atlanta, 74, 3, 12-30.

[81] Ferreira M.A. (1999) Forecasting interest rate volatility from the information in historical data, Working paper, Department of Finance, University of Wisconsin-Madison.

[82] Figlewski S. (1997) Forecasting volatility, *Financial Markets, Institutions and Instruments*, New York University Salomon Center, 6, 1, 1-88.

[83] Figlewski S. and T.C. Green (1999) Market risk and model risk for a financial institution writing options, *Journal of Finance*, 54, 4, 1465-1999.

[84] Fleming J. (1998) The quality of market voltility forecasts implied by S&P 100 index option prices, *Journal of Empirical Finance*, 5, 4, 317-345.

[85] Fleming J., C. Kirby and B. Ostdiek (2000) The economic value of volatility timing *Journal of Finance*, 56, 1.

[86] Fleming J., C. Kirby and B. Ostdiek (2002) The economic value of volatility timing using realized volatility, *Journal of Financial Economics*, Forthcoming.

[87] Fleming J., B. Ostdiek and R.E. Whaley (1995) Predicting stock market volatility: A new measure, *Journal of Futures Market*, 15, 3, 265-302.

[88] Franke G., Stapleton R.C., M.G. Subrahmanyam (1998) Who Buys and Who Sells Options: The Role of Options in an Economy with Background Risk, *Journal of Economic Theory*, 82, 1, 89-109(21).

[89] Franses P.H. and D. Van Dijk (1996) Forecasting stock market volatility using (non-linear) Garch models, *Journal of Forecasting*, 15, 3, 229-235.

[90] Franses P.H. and H. Ghijsels (1999) Additive outliers, GARCH and forecasting volatility, *International Journal of Forecasting*, 15, 1-9.

[91] Franses P.H. and D. van Dijk (2000) Non-Linear Time Series Models in Empirical Finance, Cambridge University Press.

[92] French K.R., Schwert G.W. and Stambaugh R.F. (1987) Expected stock returns and volatility, *Journal of Financial Economics*, 19, 1, 3-30.

[93] Frennberg P. and B. Hansson (1996) An evaluation of alternative models for predicting stock volatility, *Journal of International Financial Markets, Institutions and Money*, 5, 117-134.

[94] Fridman M. and L. Harris (1998) A maximum likelihood approach for non Gaussian stochastic volatility models, *Journal of Business and Economic Statistics*, 16, 284-291.

[95] Friedman B.M. and D.I. Laibson (1989) Economic implications of extraordinary movements in stock prices, *Brooking papers on Economic Activity*, 2, 137-189.

[96] Fung H-G, C-J Lie and A. Moreno (1990) The forecasting performance of the implied standard deviation in currency options, *Managerial Finance*, 16, 3, 24-29.

[97] Fung W.K.H. and D.A. Hsieh (1991) Empirical Analysis of Implied Volatility: Stocks, Bonds and Currencies, Working paper, Department of Finance, Fuqua School of Business.

[98] Gallant A.R., P.E. Rossi and G. Tauchen (1993) Nonlinear dynamic structures, *Econometrica*, 61, 4, 871-907.

[99] Garman M.B. and M.J. Klass (1980) On the estimation of security price volatilities from historical data, *Journal of Business*, 53, 1, 67-78.

[100] Gemmill G. (1986) The forecasting performance of stock options on the London Traded Options Markets, *Journal of Business Finance and Accounting*, 13, 4, 535-546.

[101] Ghysels E., A. Harvey and E. Renault (1996) Stochastic volatility, pp.119-191, in Maddala G.S. and C.R. Rao eds. Handbook of Statistics: Statistical Methods in Finance, Vol.14, Amsterdam: Elsevier Science.

[102] 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.

[103] Granger C.W.R. (1999) Empirical modeling in Economics. Specification and Evaluation. Cambridge University Press.

[104] Granger C.W.J. (2001) Long memory processes - An Economist's Viewpoint, Working paper, University of California, San Diego.

[105] Granger C.W.J. and N. Hyung (2000) Occasional structural breaks and long memory, Working paper, 99-14, University of California, San Diego.

[106] Granger C.W.J., Z. Ding and S. Spear (2000) Stylized facts on the temporal and distributional properties of absolute returns: An update, Working paper, University of California, San Diego.

[107] Gray S.F. (1996) Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics*, 42, 1, 27-62.

[108] Guo D. (1996) The predictive power of implied stochastic variance from currency options, *Journal of Futures Markets*, 16, 8, 915-942.

[109] Guo D. (1996) The information content of implied stochastic volatility from currency options, *Canadian Journal of Economics*, 29, S, 559-561, April.

[110] Hamid S. (1998) Efficient consolidation of implied volatilities and a test of intertemporal averaging, *Derivatives Quarterly*, 4, 3, 35-49.

[111] Hamilton J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57, 357-384.

[112] Hamilton J.D. and G. Lin (1996) Stock market volatility and the business cycle, *Journal of Applied Econometrics*, 11, 5, Sep/Oct, 573-593.

[113] Hamilton J.D. and R. Susmel (1994) Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics*, 64, 1-2, 307-333.

[114] Hansen L.P. and R.J. Hodrick (1980) Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *Journal of Political Economy*, 88, 829-853.

[115] Harvey A.C. (1998) Long memory in stochastic volatility, Chapter 12, 307-320, in Knight J. and S. Satchell ed. Forecasting Volatility in the Financial Markets, Butterworth.

[116] Harvey A.C., E. Ruiz and N. Shephard (1994) Multivariate stochastic variance models, *Review of Economic Studies*, 61, 247-264.

[117] Harvey C.R. and R.E. Whaley (1992) Market volatility prediction and the efficiency of the S&P100 Index option market, *Journal of Financial Economics*, 31, 1, 43-74.

[118] Heston S.L (1993) A Closed Solution For Options With Stochastic Volatility, With Application to Bond and Currency Options, *Review of Financial Studies*, 6, 2, 327-343.

[119] Heynen R.C. (1995) Essays on Derivatives Pricing Theory, Thesis Publishers, Amersterdam.

[120] Heynen R.C. and H.M. Kat (1994) Volatility prediction: A comparison of stochastic volatility, GARCH(1,1) and EGARCH(1,1) models, *Journal of Derivatives*, 50-65.

[121] Hol E. and S.J. Koopman (2001) Forecasting the variability of stock index returns with stochastic volatility models and implied volatility, working paper, Free University Amsterdam.

[122] Hsieh D.A. (1989) Modeling Heteroscedasticity in daily foreign ecxchange rates, *Journal of Business and Economic Statistics*, 7, 3, 307-317.

[123] Hu M.Y. and C. Tsoukalas (1999) Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates, *Journal of International Financial Markets, Institutions and Money*, 9, 407-422.

[124] Hull J. and A. White (1987) The pricing of options on assets with stochastic volatilities, *Journal of Finance*, 42, 2, 281-300.

[125] Hull J. and A. White (1988) An analysis of the bias in option pricing caused by a stochastic volatility, *Advances in Futures and Options Research*, 3, 27-61.

[126] Hwang S. and S. Satchell (1998) Implied volatility forecasting: a comparison of different procedures including fractionally integrated models with applications to UK equity options, Ch7, 193-225, in Knight J. and S. Satchell ed. Forecasting Volatility in the Financial Markets, Butterworth.

[127] Inclan C. and G.C. Tiao (1994) Use of cumulative sum of squares for retrospective detection of change of variance, *Journal of American Statistical Association*, 89, 913-923.

[128] Jacquier E., N.G. Polson and P.E. Rossi (1994) Bayesian analysis of stochastic volatility models: reply, *Journal of Business and Economic Statistics*, 12, 4, 413-417.

[129] Jones C.S. (2001) The dynamics of stochastic volatility: Evidence from underlying and options markets, Working paper, Simon School of Business, University of Rochester.

[130] Jones C., O. Lamont and R. Lumsdaine (1998) Macroeconomic news and bond market volatility, *Journal of Financial Economics*, 47, 315-337.

[131] Jorion, P. (1995) Predicting volatility in the foreign exchange market, *Journal of Finance*, 50, 2, 507-528.

[132] Jorion P. (1996) Risk and turnover in the foreign exchange market, in Frankel J.A., G. Galli and A. Giovannini ed. The Microstructure of Foreign Exchange Markets, Chicago: The University of Chicago Press.

[133] Karatzas I. and S.E. Shreve (1988) Brownian Motion and Stochastic Calculus, New York: Springer Verlag.

[134] Karolyi G.A. (1993) A Bayesian approach to modeling stock return volatility and option valuation, *Journal of Financial and Quantitative Analysis*, 28, 4, 579-595.

[135] Karpoff J.M. (1987) The relation between price changes and trading volume: a survey, *Journal of Financial and Quantitative Analysis*, 22, 1, 109-126.

[136] Kearns P. and A.R. Pagan (1993) Australian Stock Market Volatility, 1875-1987, *Economic Record*, 69, 163-178.

[137] Kim S, N. Shephard and S. Chib (1998) Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies*, 65, 224, 361-393.

[138] Klaassen F. (1998) Improving GARCH volatility forecasts, Forthcoming in *Empirical Economics*.

[139] Knight J. and S. Satchell (1998) ed. Forecasting Volatility in the Financial Markets, Butterworth.

[140] Kroner K.F. (1996) Creating and using volatility forecasts, *Derivatives Quarterly*, 3, 2, 39-53.

[141] Kroner K., Kneafsey K. and S. Claessens (1995) Forecasting volatility in commodity markets, *Journal of Forecasting*, 14, 77-95.

[142] Lamoureux C.G. and W.D. Lastrapes (1990) Persistence in variance, structural change and the GARCH model, *Journal of Business and Economic Statistics*, 8, 2, 225-234.

[143] Lamoureux C., and W. Lastrapes (1993) Forecasting stock-return variance: toward an understanding of stochastic implied volatilities, *Review of Financial Studies*, 6, 2, 293-326.

[144] Latane H. and R.J. Rendleman (1976) Standard deviations of stock price ratios implied in option prices, *Journal of Finance*, 31, 2, 369-381.

[145] Lee K Y (1991) Are the GARCH models best in out-of-sample performance? *Economics Letters*, 37, 3, 305-308.

[146] Li K. (2002) Long-memory versus option-implied volatility prediction, *Journal of Derivatives*, 9, 3, 9-25.

[147] Lopez J.A. (2001) Evaluating the predictive accuracy of volatility models, *Journal of Forecasting*, 20, 2, 87-109.

[148] Loudon G.F., W.H. Watt and P.K. Yadav (2000) An empirical analysis of alternative parametric ARCH models, *Journal of Applied Econometrics*, 15, 117-136.

[149] Markowitz H.M. (1991) Portfolio Selection, Blackwell Publisher.

[150] Martens M. and J. Zein (2002) Predicting financial volatility: high-frequency time-series forecasts vis-à-vis implied volatility, Working paper, Erasmus University.

[151] Mayhew S. (1995) Implied volatility, *Financial Analyst Journal*, 51, 8-20.

[152] McCurdy T.H. and I. Morgan (1987) Tests of the martingale hypothesis for foreign currency futures with time varying volatility, *International Journal of Forecasting*, 3, 131-148.

[153] McKenzie M.D. (1999) Power transformation and forecasting the magnitude of exchange rate changes, *International Journal of Forecasting*, 15, 49-55.

[154] McMillan D.G., A.H. Speight and O.A.P. Gwilym (2000) Forecasting UK stock market volatility, *Journal of Applied Economics*, 10, 435-448.

[155] Milhoj A. (1987) A conditional variance model for daily observations of an exchange rate, *Journal of Business and Economic Statistics*, 5, 99-103.

[156] Nasar S. (1992) For Fed, a new set of tea leaves, New York Times.

[157] Nelson D.B. (1991) Conditional heteroskedasticity in asset returns: A new approach, *Econometrica*, 59, 2, 347-370, March.

[158] Nelson D.B. (1992) Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model, *Journal of Econometrics*, 52, 61-90.

[159] Nelson D.B. and D.P. Foster (1995) Filtering and forecasting with misspecified ARCH models II: Making the right forecast with the wrong model, *Journal of Econometrics*, 67, 2, 303-335.

[160] Noh J. R.F. Engle and A. Kane (1994) Forecasting volatility and option prices of the S&P 500 index, *Journal of Derivatives*, 2, 17-30.

[161] Pagan A.R. and G.W. Schwert (1990) Alternative models for conditional models for conditional stock volatility, *Journal of Econometrics*, 45, 1-2, 267-290.

[162] Parkinson M. (1980) The extreme value method for estimating the variance of the rate of return, *Journal of Business*, 53, 61-65.

[163] Pitt M.J. and N. Shephard (1997) Likelihood analysis of non-Gaussian measurement time series, *Biometrika*, 84, 653-667.

[164] Perry P.R. (1982) The time-variance relationship of security returns: implications for the return-generating stochastic process, *Journal of Finance*, 37, 857-870.

[165] Pong S., M.B. Shackleton, S.J. Taylor and X. Xu (2002) Forecasting Sterling/Dollar volatility: A comparison of implied volatilities and AR(FI)MA models, Working paper, Lancaster University.

[166] Poon S. and S.J. Taylor (1992) Stock returns and stock market volatilities, *Journal of Banking and Finance*, 16, 37-59.

[167] Poteshman A.M. (2000) Forecasting future volatility from option prices, Working paper, University of Illinois at Urbana-Champaign.

[168] Randolph W.L. and M. Najand (1991) A test of two models in forecasting stock index futures price volatility, *Journal of Futures Markets*, 11, 2, 179-190.

[169] Schmalensee R. and R.R. Trippi (1978) Common stock volatility expectations implied by option premia, *Journal of Finance*, 33, 1, 129-147.

[170] Scott E. and Tucker A.L. (1989) Predicting currency return volatility, *Journal of Banking and Finance*, 13, 6, 839-851.

[171] Sill D.K. (1993) Predicting stock-market volatility, Jan/Feb, Business Review, Federal Reserve Bank of Philadelphia, 15-27.

[172] Singleton K. (2001) Estimation of affine asset pricing models using the empirical characteristic function, *Journal of Econometrics*, 102, 111-141.

[173] Szakmary A., E. Ors, J.K. Kim and W.D. Davidson III (2002) The predictive power of implied volatility: Evidence from 35 futures markets, Working paper, Southern Illinois University.

[174] Tauchen G. and M. Pitts (1983) The price variability-volume relationship on speculative markets, *Econometrica*, 51, 485-505.

[175] Taylor J.W. (2001) Volatility forecasting with smooth transition exponential smoothing, Working paper, Oxford University.

[176] Taylor S.J. (1986) Modelling Financial Time Series, Wiley.

[177] Taylor S.J. (1987) Forecasting of the volatility of currency exchange rates, *International Journal of Forecasting*, 3, 159-170.

[178] Taylor S.J. and G.X. Xu (1997) The incremental volatility information in one million foreign exchange quotations, *Journal of Empirical Finance*, 4, 4, 317-340.

[179] Theil H. (1966) Applied Economic Forecasting, North-Holland, Amsterdam.

[180] Tse Y.K. (1991) Stock return volatility in the Tokyo Stock Exchange, *Japan and the World Economy*, 3, 285-298.

[181] Tse T.Y.K. and S.H. Tung (1992) Forecasting volatility in the Singapore stock market, Asia Pacific *Journal of Management*, 9, 1, 1-13.

[182] Vasilellis G.A. and N. Meade (1996) Forecasting volatility for portfolio selection, *Journal of Business Finance and Accounting*, 23, 1, 125-143.

[183] Vilasuso J. (2002) Forecasting exchange rate volatility, *Economics Letters*, 76, 59-64.

[184] Walsh D.M. and G.Y-G Tsou (1998) Forecasting index volatility: Sampling integral and non-trading effects, *Applied Financial Economics*, 8, 5, 477-485.

[185] Wei S.J. and J.A. Frankel (1991) Are option-implied forecasts of exchange rate volatility excessively variable? NBER working paper No.3910.

[186] West K.D. (1996) Asymptotic inference about predictive ability, *Econometrica*, 64, 1067-1084

[187] West K.D. and D. Cho (1995) The predictive ability of several models of exchange rate volatility, *Journal of Econometrics*, 69, 2, 367-391.

[188] West K.D., H.J. Edison and D. Cho (1993) A utility based comparison of some methods of exchange rate volatility, *Journal of International Economics*, 35, 1-2, 23-45.

[189] West K.D. and M. McCracken (1998) Regression based tests of predictive ability, *International Economic Review*, 39, 817-840.

[190] Whaley R.E. (1982) Valuation of American call options on dividend-paying stocks, *Journal of Financial Economics*, 10, 29-58, Vanderbilt University.

[191] Whaley R.E. (1993) Derivatives on market volatility: Hedging tools long overdue, *Journal of Derivatives*, Fall, 71-84.

[192] Wiggins J.B. (1987) Option values under stochastic volatility: Theory and empirical estimates, *Journal of Financial Economics*, 19, 351-372.

[193] Wiggins J.B. (1992) Estimating the volatility of S&P 500 futures prices using the Extreme-Value Method, *Journal of Futures Markets*, June, 12, 3, 256-273.

[194] Xu X. and S.J. Taylor (1995) Conditional volatility and the informational efficiency of the PHLX currency options market, *Journal of Banking and Finance*, 19, 5, 803-821.

[195] Yu J. (2002) Forecasting volatility in the New Zealand Stock Market, *Applied Financial Economics*, 12, 193-202.

[196] Zumbach G. (2002) Volatility processes and volatility forecast with long memory, Working paper, Olsen Associates.

[197] Zumbach G., O.V. Pictet and O. Masutti (2001) Genetic programming with syntactic restrictions applied to financial volatility forecasting, Working paper, Olsen Associates.

Appendix A: Historical Price Volatility Models

A.1 Prediction models build on sample standard deviations

Volatility, σ_t, in this section is the sample standard deviation of period t returns, and $\widehat{\sigma}_t$ is the forecast of σ_t. If t is a month, then σ_t is often calculated as the sample standard deviation of all daily returns in the month. For a long time, σ_t is proxied by daily squared return if t is a day. More recently and with the availability of high frequency data, daily σ_t is derived from the cumulation of intra-day returns.

Random Walk (RW)

$$\widehat{\sigma}_t = \sigma_{t-1} \tag{14}$$

Historical Average (HA)

$$\widehat{\sigma}_t = \left(\sigma_{t-1} + \sigma_{t-2} + \cdots + \sigma_1 \right) / \left(t - 1 \right) \tag{15}$$

Moving Average (MA)

$$\widehat{\sigma}_t = \left(\sigma_{t-1} + \sigma_{t-2} + \cdots + \sigma_{t-\tau} \right) / \left(\tau \right) \tag{16}$$

Exponential Smoothing (ES)

$$\widehat{\sigma}_t = \left(1 - \beta \right) \sigma_{t-1} + \beta \widehat{\sigma}_{t-1} \quad \text{and} \quad 0 \leq \beta \leq 1 \tag{17}$$

Exponentially Weighted Moving Average (EWMA)

$$\widehat{\sigma}_t = \sum_{i=1}^{\tau} \beta^i \, \sigma_{t-i} \left/ \sum_{i=1}^{\tau} \beta^i \right. \quad . \tag{18}$$

EWMA is a truncated version of ES with a finite τ.

Smooth Transition Exponential Smoothing (STES)

$$
\begin{aligned}
\widehat{\sigma}_t &= \alpha_{t-1} \varepsilon_{t-1}^2 + \left(1 - \alpha_{t-1} \right) \widehat{\sigma}_{t-1}^2 \\
\alpha_{t-1} &= \frac{1}{1 + \exp\left(\beta + \gamma V_{t-1} \right)}
\end{aligned}
\tag{19}
$$

where V_{t-1} is the transition variable; $V_{t-1} = \varepsilon_{t-1}$ for STES-E, $V_{t-1} = |\varepsilon_{t-1}|$ for STES-AE and V_{t-1} is a function of both ε_{t-1} and $|\varepsilon_{t-1}|$ for STES-EAE.

Simple Regression (SR)

$$\widehat{\sigma}_t = \gamma_{1,\,t-1} \, \sigma_{t-1} + \gamma_{2,\,t-1} \, \sigma_{t-2} + \cdots \tag{20}$$

Threshold autoregressive (TAR)

$$\widehat{\sigma}_t = \phi_0^{(i)} + \phi_1^{(i)} \sigma_{t-1} + \cdots + \phi_p^{(i)} \sigma_{t-p}, \quad i = 1, 2, \ldots, k \tag{21}$$

A.2 ARCH class conditional volatility models

For all models described in this section, returns, r_t, has the following process

$$
\begin{aligned}
r_t &= \mu + \varepsilon_t \\
\varepsilon_t &= \sqrt{h_t} z_t
\end{aligned}
$$

and h_t follows one of the following ARCH class models.

ARCH (q)

$$
h_t = \omega + \sum_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2 \tag{22}
$$

where $\omega > 0$ and $\alpha_k \geq 0$.

GARCH (p, q)

$$
h_t = \omega + \sum_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{23}
$$

where $\omega > 0$. (See Nelson and Cao (1992) for constraint on α_k and β_j.) For finite variance, $\sum \alpha_k + \sum \beta_j < 1$.

EGARCH (p, q)

$$
\begin{aligned}
\ln h_t &= \alpha_0 + \sum_{j=1}^{q} \beta_j \ln h_{t-j} \\
&+ \sum_{k=1}^{p} \alpha_k \left[\theta \psi_{t-k} + \gamma \left(|\psi_{t-k}| - \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \right) \right] \\
\psi_t &= \mu_t \Big/ \sqrt{h_t}
\end{aligned} \tag{24}
$$

TGARCH (1,1) or GJR-GARCH (1,1)

$$
\begin{aligned}
h_t &= \omega + \alpha \varepsilon_{t-1}^2 + \delta D_{t-1} \varepsilon_{t-1}^2 + \beta h_{t-1} \\
D_{t-1} &= \begin{cases} 1 & \text{if } \varepsilon_{t-1} < 0 \\ 0 & \text{if } \varepsilon_{t-1} \geq 0 \end{cases}
\end{aligned} \tag{25}
$$

QGARCH (1,1)

$$
h_t = \omega + \alpha \left(\varepsilon_{t-1} - \gamma \right)^2 + \beta h_{t-1} \tag{26}
$$

STGARCH (*Smooth Transition* GARCH)

$$
h_t = \omega + (1 - F(\varepsilon_{t-1})) \alpha \varepsilon_{t-1}^2 + F(\varepsilon_{t-1}) \delta \varepsilon_{t-1}^2 + \beta h_{t-1} \tag{27}
$$

where

$$
F(\varepsilon_{t-1}) = \frac{1}{1 + \exp(-\theta \varepsilon_{t-1})} \quad \text{for logistic STGARCH,}
$$

$$F\left(\varepsilon_{t-1}\right) = 1 + \exp\left(-\theta\varepsilon_{t-1}^2\right) \quad \text{for exponential STGARCH,}$$

GARCH(1,1) *regime switching*

$$h_{t,\ S_{t-1}} = \omega_{S_{t-1}} + \alpha_{S_{t-1}}\varepsilon_{t-1}^2 + \beta_{S_{t-1}}h_{t-1,\ S_{t-1}} \tag{28}$$

where S_t indicate the state of regime at time t.

CGARCH(1,1) (*Component* GARCH)

$$
\begin{aligned}
h_t &= \omega_t + \alpha_1\left(\varepsilon_{t-1}^2 - \omega_{t-1}\right) + \beta_1\left(h_{t-1} - \omega_{t-1}\right) \\
\omega_t &= \omega + \rho\omega_{t-1} + \xi\left(\varepsilon_{t-1}^2 - h_{t-1}\right)
\end{aligned}
\tag{29}
$$

where ω_t represents a time-varying trend or permanent component in volatility which is driven by volatility prediction error $\left(\varepsilon_{t-1}^2 - h_{t-1}\right)$ and is integrated if $\rho = 1$.

A.3 Stochastic volatility model

Stochastic Volatility (SV)

$$
\begin{aligned}
r_t &= \mu + \varepsilon_t \\
\varepsilon_t &= z_t \exp\left(0.5h_t\right) \\
h_t &= \omega + \beta h_{t-1}^2 + v_t
\end{aligned}
\tag{30}
$$

v_t may or may not be independent of z_t.

	Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
1.	Akigray (1989)	CRSP VW & EW indices	Jan63-Dec86 (Pre-crash) Split into 4 subperiods of 6 years each.	D	GARCH(1,1) ARCH(2) EWMA HIS (ranked)	20 days ahead estimated from rolling 4 years data. Daily returns used to construct "actual vol"; adjusted for serial correlation.	ME, RMSE, MAE, MAPE	GARCH is least biased and produced best forecast especial in periods of high volatility and when changes in volatility persist. Heteroskedasticity is less string in low frequency data and monthly returns are approximately Normal.
2.	Alford and Boatman (1995)	6879 stocks listed in NYSE/ASE & NASDAQ	12/66 – 6/87	W, M	"Shrinkage" forecast (HIS adjusted towards comparable firms) HIS Median HIS vol of "comparable" firm (ranked)	5 years starting from 6 months after firm's fiscal year	MedE, MedAE	To predict 5-year monthly volatility one should use 5 years worth of weekly or monthly data. Adjusting historical forecast using industry and size produced best forecast.
3.	Amin and Ng (1997)	3M Eurodollar futures & futures options	1/1/88 – 1-/11/92	D	Implied $_{American\ All\ Call+Put}$ (WLS, 5 variants of the HJM model) HIS (ranked)	20 days ahead (1 day ahead forecast produced from in-sample with lag implied in GARCH/GJR not discussed here.)	R^2 is 21% for implied and 24% for combined. $H_0: a_{implied}=0$, $b_{implied}=1$ cannot be rejected with robust SE.	Interest rate models that incorporate volatility term structure (e.g. Vasicek) perform best. Interaction term capturing rate level and volatility contribute additional forecasting power.
4.	Andersen and Bollerslev (1998)	DM/$, ¥/$	In: 1/10/87–30/9/92 Out: 1/10/92-30/9/93	D (5 min)	GARCH(1,1)	1 day ahead, use 5-min returns to construct "actual vol"	R^2 is 5 to 10% for daily squared returns, 50% for 5-min square returns.	R^2 increases monotonically with sample frequency.
5.	Andersen, Bollerslev, Diebold and Labys (2002)	¥/US$, DM/US$ Reuters FXFX quotes	1/12/86 – 30/6/99 In: 1/12/86-1/12/96, 10 years Out: 2/12/96 – 30/6/99, 2.5 years	Tick (30min)	VAR-RV, AR-RV, FIEGARCH-RV GARCH-D, RM-D, FIEGARCH-D VAR-ABS (ranked)	1 and 10 days ahead. "Actual vol" derived from 30-min returns.	1-day ahead R^2 ranges between 27-40% (1-day ahead) and 20-33% (10-day ahead).	RV is realised volatility, D is daily return, and ABS is daily absolute return. VAR allows all series to share the same fractional integrated order and cross series linkages. Forecast improvement is largely due to the use of high frequency data (and realised volatility) instead of the model(s).

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
6. Andersen, Bollerslev and Lange (1999)	DM/US$ Reuters quotes	1/12/86 – 30/11/96 In: 1/10/87-30/9/92	5 min	GARCH(1,1) at 5-min, 10-min, 1-hr, 8-hr, 1-day, 5-day, 20-day interval.	1, 5 and 20 days ahead, use 5-min returns to construct "actual vol"	RMSE, MAE, HRMSE, HMAE, LL	HRMSE and HMAE are heteroskedasticity adjusted error statistics; LL is the logarithmic loss function. High frequency returns and high frequency GARCH(1,1) models improve forecast accuracy. But, for sampling frequencies shorter than 1 hour, the theoretical results and forecast improvement break down.
7. Bali (2000)	3-, 6-, 12-month T-Bill rates	8/1/54 – 25/12/98	W	NGARCH GJR, TGARCH AGARCH, QGARCH TSGARCH GARCG VGARCH Constant vol (CKLS) (ranked, forecast both level and volatility)	1 week ahead. Use weekly interest rate absolute change to proxy "actual vol".	R^2 increases from 2 % to 60% by allowing for asymmetries, level effect and changing volatility.	CKLS: Chan, Karolyi, Longstaff and Sanders (1992).
8. Beckers (1981)	62 to 116 Stocks options	28/4/75 – 21/10/77	D	FBSD Implied $_{ATM\ call,\ 5\ days\ ave}$ Implied $_{vega\ call,\ 5\ days\ ave}$ RW $_{last\ quarter}$ (ranked, both implieds are 5-day average because of large variations in daily stock implied.)	Over option's maturity (3 months), 10 non-overlapping cycles. Use sample SD of daily returns over option maturity to proxy "actual vol".	MPE, MAPE. Cross sectional R^2 ranges between 34-70% across models and expiry cycles. FBSD appears to be least biased with a =0, b =1. a >0, b <1 for the other two implieds.	FBSD: Fisher Black's option pricing service takes into account stock vol tend to move together, mean revert, leverage effect and implied can predict future. ATM, based on vega WLS, outperforms vega weighted implied, and is not sensitive to *ad hoc* dividend adjustment. Incremental information from all measures suggests option market inefficiency. Most forecasts are upwardly biased as actual vol was on a decreasing trend.
	50 stock option	4 dates: 18/10/76, 24/1/77, 18/4/77, 18/7/77	D (from Tick)	TISD$_{vega}$ Implied $_{ATM\ call,\ 1\ days\ ave}$ (ranked)	ditto	Cross sectional R^2 ranges between 27-72% across models and expiry cycles.	TISD: Single intra-day transaction data that has the highest vega. The superiority of TISD over implied of closing option prices suggest significant non-simultaneity and bid-ask spread problems.
9. Bera and Higgins (1997)	Daily SP500, Weekly $/£, Monthly US Ind Prod	SP 1/1/88-28/5/93 $/£ 12/12/85-28/2/91 IndProd 1/60–3/93	D W M	GARCH Bilinear model (ranked)	One step ahead. Reserve 90% of data for estimation	Cox MLE RMSE (LE: logarithmic error)	Consider if heteroskedasticity is due to bilinear in level. Forecasting results show strong preference for GARCH.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
10. Blair, Poon & Taylor (2001)	S&P 100 (VIX)	2/1/87 to 31/12/99 Out: 4/1/93-31/2/99	Tick	Implied $_{VIX}$ GJR HIS$_{100}$ (ranked)	1, 5, 10 and 20 days ahead estimated using a rolling sample of 1,000 days. Daily actual volatility is calculated from 5-min returns.	1-day ahead R^2 is 45% for VIX, and 50% for combined. VIX is downward biased in out-of-sample period.	Using squared returns reduces R^2 to 36% for both VIX and combined. Implied volatility has its own persistence structure. GJR has no incremental information though integrated HIS vol can almost match IV forecasting power.
11. Bluhm and Yu (2000)	German DAX stock index and VDAX the DAX volatility index	In 1/1/88-28/6/96 Out 1/7/66-30/6/99	D	Implied $_{VDAX}$ GARCH(-M), SV EWMA, EGARCH, GJR, HIS (approx ranked)	45 calendar days, 1, 10 and 180 trading days. "Actual" is the sum of daily squared returns.	MAPE, LINEX	Ranking varies a lot depend on forecast horizons and performance measures.
12. Boudoukh, Richardson and Whitelaw (1997)	3-month US T-bill	1983-1992	D	EWMA, MDE GARCH(1,1), HIS (ranked)	1 day ahead based on 150-day rolling period estimation. Realized volatility is the daily squared changes averaged across t+1 to t+5.	MSE & regression. MDE has the highest R^2 while EWMA has the smallest MSE	MDE is multivariate density estimation where volatility weights depend on interest rate level and term spread. EWMA and MDE have comparable performance and are better than HIS and GARCH.
13. Brace and Hodgson (1991)	Futures option on Australian Stock Index (Marking to market is needed for this options)	1986-87	D	HIS $_{5, 20, 65\ days}$ Implied $_{NTM\ call,\ 20-75\ days}$ (ranked)	20 days ahead. Use daily returns to calculate standard deviations.	Adj R^2 are 20% (HIS), 17% (HIS+ implied). All a >0 & sig. some uni regr coeff are sig negative (for both HIS & implied).	Large fluctuations of R^2 from month to month. Results could be due to the difficulty in valuing futures style options.
14. Brailsford and Faff (1996)	Australian Statex-Actuaries Accumulation Index for top 56	1/1/74 – 30/6/93 In: Jan74-Dec85 Out: Jan86-Dec93 (include 87's crash period)	D	GJR, Regr, HIS, GARCH, MA, EWMA, RW, ES (rank sensitive to error statistics)	1 month ahead. Models estimated from a rolling 12-year window.	ME, MAE, RMSE, MAPE, and a collection of asymmetric loss functions	Though the ranks are sensitive, some models dominate others; MA12 >MA5 and Regr>MA> EWMA>ES. GJR came out quite well but is the only model that always underpredict.
15. Brooks (1998)	DJ Composite	17/11/78–30/12/88 Out: 17/10/86-30/12/88	D	RW, HIS, MA, ES, EWMA, AR, GARCH, EGARCH, GJR, Neural network (all about the same)	1 day ahead squared returns using rolling 2,000 observations for estimation.	MSE, MAE of variance, % over-predict. R^2 is around 4% increases to 24% for pre-crash data	Similar performance across models especially when 87's crash is excluded. Sophisticated models such as GARCH and neural net did not dominate. Volume did not help in forecasting volatility.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
16. Canina and Figlewski (1993)	S&P 100 (OEX)	15/3/83 – 28/3/87 (Pre-crash)	D	HIS $_{60\ calendar\ days}$ Implied $_{Binomial\ Call}$ (ranked) Implied in 4 maturity gp, each subdivided into 8 intrinsic gp.	7 to 127 calendar days matching option maturity, overlapping forecasts with Hansen std error. Use sample SD of daily returns to proxy "actual vol".	Combined R^2 is 17% with little contribution from implied. All a $_{implied}$>0, b $_{implied}$<1 with robust SE.	Implied has no correlation with future volatility and does not incorporate info contained in recently observed volatility. Results appear to be peculiar for pre-crash period. Time horizon of "actual vol" changes day to day. Different level of implied aggregation produces similar results.
17. Cao and Tsay (1992)	Excess returns for S&P, VW EW indices	1928-1989	M	TAR EGARCH(1,0) ARMA(1,1) GARCH(1,1) (ranked)	1 to 30 months. Estimation period ranges from 684 to 743 months	MSE, MAE Daily returns used to construct "actual vol".	TAR provides best forecasts for large stocks. EGARCH gives best long-horizon forecasts for small stocks (may be due to Leverage effect). Difference in MAE can be as large as 38%.
18. Chiras and Manaster (1978)	All stock options from CBOE	23 months from Jun73 – Apr75	M	Implied (weighted by price elasticity) HIS$_{20\ monthss}$ (ranked)	20 month ahead. Use SD of 20 monthly returns to proxy "actual vol".	Cross sectional R^2 of implied ranges 13-50% across 23 months. HIS adds 0-15% to R^2.	Implied outperformed HIS especially in the last 14 months. Find implied increases and better behave after dividend adjustments and evidence of mispricing possibly due to the use European pricing model on American style options.
19. Christensen and Prabhala (1998)	S&P 100 (OEX) Monthly expiry cycle	Nov83 - May95	M	Implied $_{BS\ ATM\ 1-month\ Call}$ HIS$_{18\ days}$ (ranked)	Non-overlapping 24 calendar (or 18 trading) days. Use SD of daily returns to proxy "actual vol".	R^2 of log var are 39% (implied), 32% (HIS) and 41% (combined). a <0 (because of log), b <1 with robust SE. Implied is more biased before the crash.	Not adj for dividend and early exercise. Implied dominates HIS. HIS has no additional information in subperiod analysis. Proved that results in Canina & Figlewski (1993) is due to pre-crash characteristics and high degree of data overlap relative to time series length. Implied is unbiased after controlling for measurement errors using implied$_{t-1}$ and HIS$_{t-1}$.
20. Christoffersen and Diebold (2000)	4 stk indices 4 ex rates US 10 year T-Bond	1/1/73 - 1/5/97	D	No model. (No rank; Evaluate volatility forecastibility (or persistence) by checking interval forecasts.)	1 to 20 days	Run tests and Markov transition matrix eigenvalues (which is basically 1st-order serial coefficient of the hit sequence in the run test).	Equity & FX: forecastibility decrease rapidly from 1 to 10 days. Bond: may extend as long as 15 to 20 days. Estimate bond returns from bond yields by assuming coupon equal to yield.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
21. Cumby, Figlewski and Hasbrouck (1993)	¥/$, stocks(¥, $), bonds (¥, $)	7/77 to 9/90	W	EGARCH HIS (ranked)	1 week ahead, estimation period ranges from 299 to 689 weeks.	R^2 varies from 0.3% to 10.6%.	EGARCH is better than naïve in forecasting volatility though R-square is low. Forecasting correlation is less successful.
22. Day and Lewis (1992)	S&P 100 OEX option Reconstructed S&P 100	Out: 11/11/83 – 31/12/89 In: 2/1/76 – 11/11/83	W	Implied $_{BS Call}$ (shortest but > 7 days, volume WLS) HIS$_{1 week}$ GARCH EGARCH (ranked)	1 week ahead estimated from a rolling sample of 410 observations. Use sample variance of daily returns to proxy weekly "actual vol".	R^2 of variance regr are 2.6% (implied) & 3.8% (encomp). All forecasts add marginal info. H_0: $a_{implied}$=0, $b_{implied}$=1 cannot be rejected with robust SE.	Omit early exercise. Effect of 87's crash is unclear. When weekly squared returns were used to proxy "actual vol", R^2 increase and was max for HIS contrary to expectation (9% compared with 3.7% for implied).
23. Day and Lewis (1993)	Crude oil futures options Crude oil futures	14/11/86 – 18/3/91 8/4/83 – 18/3/91 coincide with Kuwait invasion by Iraq in second half of sample.	D	Implied $_{Binomial ATM Call}$ HIS $_{forecast horizon}$ GARCH-M EGARCH-AR(1) (ranked)	Option maturity of 4 nearby contracts, (average 13.9, 32.5, 50.4 & 68 trading days to maturity). Estimated from rolling 500 observations.	ME, RMSE, MAE. R^2 of variance regr are 72% (short mat) and 49% (long maturity). With robust SE a >0 for short and a =0 for long, b =1 for all maturity.	Implied performed extremely well. Performance of HIS and GARCH are similar. EGARCH much inferior. Bias adjusted and combined forecasts do not perform as well as unadjusted implied. GARCH has no incremental information. Result likely to be driven by Kuwait invasion by Iraq.
24. Dimson and Marsh (1990)	UK FT All Share	1955-89	Q	ES, Regression RW, HA, MA (ranked)	Next quarter. Use daily returns to construct "actual vol".	MSE, RMSE, MAE, RMAE	Recommend exponential smoothing and regression model using fixed weights. Find ex ante time-varying optimization of weights does not work well ex post.
25. Doidge and Wei (1998)	Toronto 35 stock index & European options	In 2/8/88-31/12/91 Out: 1/92-7/95	D	Combine3 Combine2 GARCH EGARCH HIS$_{100 days}$ Combine1 Implied $_{BS Call+Put}$(All maturities>7days, volume WLS) (ranked)	1 month ahead from rolling sample estimation. No mention on how "actual vol" was derived.	MAE, MAPE, RMSE	Combine1 equal weight for GARCH and implied forecasts. Combine2 weighs GARCH and implied based on their recent forecast accuracy. Combine3 puts implied in GARCH conditional variance. Combine3 was estimated using full sample due to convergence problem; so not really out-of-sample forecast..

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
26. Dunis, Laws and Chauvin (2000)	DM/¥, £/DM, £/$, $/CHF, $/DM, $/¥	In: 2/1/91-27/2/98 Out: 2/3/98-31/12/98	D	GARCH(1,1) AR(10)-Sq returns AR(10)-Abs returns SV(1) in log form HIS $_{21 \text{ or } 63 \text{ trading days}}$ 1- & 3-M forward Implied $_{\text{ATM quotes}}$ Combine Combine (except SV) (rank changes across currencies and forecast horizons)	1 and 3 months (21 & 63 trading days) with rolling estimation. Actual volatility is calculated as the average absolute return over the forecast horizon.	RMSE, MAE, MAPE, Theil-U, CDC (Correct Directional Change index)	No single model dominates though SV is consistently worst, and implied always improves forecast accuracy. Recommend equal weight combined forecast excluding SV.
27. Ederington and Guan (1999) "Frown"	S&P500 futures options	1Jan88-30Apr98	D	Implied $_{\text{BK 16Calls 16 Puts}}$ HIS $_{40 \text{ days}}$ (ranked)	Overlapping 10 to 35 days matching maturity of nearest to expiry option. Use SD of daily returns to proxy "actual vol".	Panel R^2 19% and individual R^2 ranges 6-17% (calls) and 15-36% (puts). Implied is biased & inefficient, $a_{\text{implied}}>0$ and $b_{\text{implied}}<1$ with robust SE.	Information content of implied across strikes exhibit a frown shape with options that are NTM and have moderately high strikes possess largest information content. HIS typically adds 2-3% to the R^2 and nonlinear implied terms add another 2-3%. Implied is unbiased & efficient when measurement error is controlled using Implied $_{t-1}$ and HIS $_{t-1}$.
28. Ederington and Guan (2000) "Forecasting Volatility"	5 DJ Stocks S&P 500 3m Euro$ rate 10y T-Bond yield DM/$	2/7/62-30/12/94 2/7/62-29/12/95 1/1/73-20/6/97 2/1/62-13/6/97 1/1/71-30/6/97	D	GW $_{\text{MAD}}$ GW $_{\text{STD}}$, GARCH, EGARCH AGARCH HIS $_{\text{MAD,n}}$, HIS $_{\text{STD,n}}$ (ranked, error statistics are close; GW $_{\text{MAD}}$ leads consistently though with only small margin.)	n=10, 20, 40, 80 & 120 days ahead estimated from a 1260-day rolling window; parameters re-estimated every 40 days. Use daily squared deviation to proxy "actual" vol.	RMSE, MAE	GW: geometric weight, MAD: mean absolute deviation, STD: standard deviation. Volatility aggregated over a longer period produces a better forecast. Absolute returns models generally perform better than square returns models (except GARCH>AGARCH). As horizon lengthens, no procedure dominates. GARCH & EGARCH estimations were unstable at times.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
29. Ederington and Guan (2000) "Averaging"	S&P500 futures options	In 4/1/88-31/12/91 Out: 2/1/92-31/12/92	D	Implied[*]: 99%, VIX $HIS_{40 \text{ trading days}}$ Implied: VIX>Eq4 Implied: WLS>vega> Eq32>elasticity (ranked)	Overlapping 10 to 35 days matching maturity of nearest to expiry option. Use SD of daily returns to proxy "actual vol".	RMSE, MAE, MAPE '*' indicates individual implieds were corrected for biasness first before averaging using in-sample regr on realised.	VIX: 2calls+2puts, NTM weighted to get ATM. Eq4/32: calls+puts equally weighted. WLS, vega and elasticity are other weighting scheme. 99% means 1% of regr error used in weighting all implieds. Once the biasness has been corrected using regr, little is to be gained by any averaging in such a highly liquid S&P500 futures market.
30. Ederington and Guan (2002) "Efficient Predictor"	S&P500 futures options	1/1/83 – 14/9/95	D	Implied $_{\text{Black 4NTM}}$ GARCH, $HIS_{40 \text{ days}}$ (ranked)	Overlapping option maturity 7-90, 91-180, 181-365 and 7-365 days ahead. Use sample SD over forecast horizon to proxy "actual vol".	R^2 ranges 22-12% from short to long horizon. Post 87's crash R^2 nearly doubled. Implied is efficient biased; $a_{\text{implied}}>0$ and $b_{\text{implied}}<1$ with robust SE.	GARCH parameters were estimated using whole sample. GARCH and HIS add little to 7-90 day R^2. When 87's crash was excluded HIS add sig explanatory power to 181-365 day forecast. When measurement errors were controlled using implied $_{t-5}$ and implied $_{t+5}$ as instrument variables implied becomes unbiased for the whole period but remains biased when crash period was excluded.
31. Edey and Elliot (1992)	Futures options on A\$ 90d-Bill, 10yr bond, Stock index A\$/US\$ options	Futures options: inception to 12/88 A\$/US\$ option: 12/84 – 12/87	W W	Implied $_{\text{BK NTM,call}}$ Implied $_{\text{BK NTM,put}}$ (No rank, 1 call and 1 put, selected based on highest trading volume)	Option maturity up to 3M. Use sum of (return square plus Implied $_{t+1}$) as "actual vol" Constant 1M. Use sum of weekly squared returns to proxy "actual vol".	Regression (see comment). In most cases $a_{\text{implied}}>0$ and $b_{\text{implied}}<1$ with robust SE. For stock index option $b_{\text{implied}}=1$ cannot be rejected using robust SE.	R^2 cannot be compared with other studies because of the way "actual" is derived and lagged squares returns were added to the RHS.
32. Engle, Ng and Rothschild (1990)	1 to 12 months T-Bill returns, VW index of NYSE & AMSE stocks	Aug 64 – Nov 85	M	1-Factor ARCH Univariate ARCH-M (ranked)	1 month ahead volatility and risk premium of 2 to 12 months T-Bills	Model fit	Equally weighted bill portfolio is effective in predicting (i.e. in an expectation model) volatility and risk premia of individual maturities.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
33. Feinstein (1989)	S&P 500 futures options (CME)	Jun83 – Dec88	Option expiry cycle	Implied: Alanta > average > vega >elasticity Just-OTM$_{Call > P+C > Put}$ HIS$_{20\ days}$ (ranked, note pre-crash rank is very different and erratic)	23 non-overlapping forecasts of 57, 38 and 19 days ahead. Use sample SD of daily returns over the option maturity to proxy "actual vol".	MSE, MAE, ME. T-test indicates all ME>0 (except HIS) in the post crash period which means implied was upwardly biased.	Alanta: 5-day average of Just-OTM call implied using exponential weights. In general Just-OTM Implied$_{call}$ is the best.
34. Ferreira (1999)	French & German interbank 1M mid rate	In: Jan81 – Dec89 Out: Jan90– Dec97 (ERM crises: Sep92-Sep93)	W	ES, HIS$_{26,\ 52,\ all}$ GARCH(-L) (E)GJR(-L) (rank varies between French & German rates, sampling method and error statistics.)	1 week ahead. Use daily squared rate changes to proxy weekly volatility.	Regression, MPE, MAPE, RMSPE. R^2 is 41% for France and 3% for Germany.	L: interest rate level, E: exponential. French rate was very volatile during ERM crises. German rate was extremely stable in contrast. Although there are lots of differences between the two rates, best models are non-parametric; ES (French) and simple level effect (German). Suggest a different approach is needed for forecasting interest rate volatility.
35. Figlewski (1997)	S&P 500 3M US T-Bill 20Y T-Bond DM/$	1/47 – 12/95 1/47 – 12/95 1/50 – 7/93 1/71 – 11/95	M	HIS$_{6,\ 12,\ 24,\ 36,\ 48,\ 60m}$ GARCH(1,1) for S&P and bond yield. (ranked)	6, 12, 24, 36, 48, 60 months. Use daily returns to compute "actual vol".	RMSE	Forecast of volatility of the longest horizon is the most accurate. HIS uses the longest estimation period is the best except for short rate.
	S&P 500 3M US T-Bill 20Y T-Bond DM/$	2/7/62 – 29/12/95 2/1/62 – 29/12/95 2/1/62 – 29/12/95 4/1/71 – 30/11/95	D	GARCH(1,1) HIS$_{1,\ 3,\ 6,\ 12,\ 24,\ 60\ months}$ (S&P's rank, reverse for the others)	1, 3, 6, 12, 24 months	RMSE	GARCH is best for S&P but gave worst performance in all the other markets. In general, as out of sample horizon increases, the in sample length should also increase.
36. Figlewski & Green (1999)	S&P 500 US LIBOR 10 yr T-Bond yield DM/$	1/4/71-12/31/96 Out: From Feb96	D	His$_{3,\ 12,\ 60\ months}$ ES (rank varies)	1, 3, 12 months for daily data.	RMSE	ES works best for S&P (1-3 month) and short rate (all three horizon). HIS works best for bond yield, exchange rate and long horizon S&P forecast. The longer the forecast horizon, the longer th estimation period.
		1/4/71-12/31/96 Out: From Jan92	M	His$_{26,\ 60,\ all\ months}$ ES (ranked)	24 & 60 months for monthly data.		For S&P, bond yield and DM/$, it is best to use all available "monthly" data. 5 years worth of data works best for short rate.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
37. Fleming (1998)	S&P 100 (OEX)	10/85 - 4/92 (All observations that overlap with 87's crash were removed)	D	Implied $_{FW\ ATM\ calls}$ Implied $_{FW\ ATM\ puts}$ (Both implieds are WLS using all ATM options in the last 10 minutes before market close) ARCH/GARCH HIS$_{H-L\ 28\ days}$ (ranked)	Option maturity (shortest but > 15days, average 30 calendar days), 1 and 28 days ahead. Use daily square return deviations to proxy "actual vol".	R^2 is 29% for monthly forecast and 6% for daily forecast. All $a_{implied}$=0, $b_{implied}$<1 with robust SE for the last two fixed horizon forecasts.	Implied dominates. All other variables related to volatility such as stock returns, interest rate and parameters of GARCH do not possess information incremental to that contained in implied.
38. Fleming, Kirby and Ostdiek (2000)	S&P 500, T-Bond and gold futures	3/1/83 – 31/12/97	D	Exponentially weighted var-cov matrix	Daily rebalanced portfolio	Sharpe ratio (portfolio return over risk)	Efficient frontier of volatility timing strategy plotted above that of fixed weight portfolio.
39. Fleming, Ostdiek & Whaley (1995)	S&P 100 (VIX)	Jan86-Dec92	D, W	Implied $_{VIX}$ HIS$_{20\ days}$ (ranked)	28 calendar (or 20 trading) day. Use sample SD of daily returns to proxy "actual vol".	R^2 increased from 15% to 45% when crash is excluded. a_{VIX}=0, b_{VIX}<1 with robust SE.	VIX dominates HIS, but is biased upward up to 580 basis point. Orthogonality test rejects HIS when VIX is included. Adjust VIX forecasts with average forecast errors of the last 253 days helps to correct for biasness while retaining implied's explanatory power.
40. Franses & Ghijsels (1999)	Dutch, German, Spanish and Italian stock market returns	1983-94	W	AO-GARCH (GARCH adjusted for additive outliers using the 'less-one' method) GARCH GARCH-t (ranked)	1 week ahead estimated from previous 4 years. Use weekly squared deviations to proxy "actual vol".	MSE & MedSE	Forecasting performance significantly improved when parameter estimates are not influenced by 'outliers'. Performance of GARCH-t is consistently much worse. Same results for all four stock markets.
41. Franses and Van Dijk (1996)	Stock indices (Germany, Netherlands, Spain, Italy, Sweden)	1986-94	W	QGARCH RW GARCH GJR (ranked)	1 week ahead estimated from rolling 4 years. Use weekly squared deviations to proxy "actual vol".	MedSE	QGARCH is best if data has no extremes. RW is best when 87's crash is included. GJR cannot be recommended. Results are likely to be influenced by MedSE that penalize nonsymmetry. Brailsford and Faff (1996) support GJR as best model although it underpredicts over 70% of the time.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
42. Frennberg and Hansson (1996)	VW Swedish stock market returns Index option (European style)	In: 1919-1976 Out: 1977-82, 1983-90 Jan87-Dec90	M	AR_{12} (ABS)-S RW, Implied $_{BS\ ATM\ Call}$ (option maturity closest to 1 month) GARCH-S, ARCH-S (ranked) Models that are not adj for seasonality did not perform as well.	1 month ahead estimated from recursively re-estimated expanding sample. Use daily ret to compile monthly vol, adjusted for autocorrelation.	MAPE, R^2 is 2-7% in first period and 11-24% in second, more volatile period. H_0: $a_{implied}$=0 and $b_{implied}$=1 cannot be rejected wust SE.	S: seasonality adjusted. RW model seems to perform remarkably well in such a small stock market where returns exhibit strong seasonality. Option was introduced in 86 and covered 87's crash; outperformed by RW. ARCH/GARCH did not perform as well in the more volatile second period.
43. Fung, Lie and Moreno (1990)	£/$, C$/$, FFr/$, DM/$, ¥/$ & SrFr/$ options on PHLX	1/84 - 2/87 (Pre crash)	D	Implied $_{OTM>ATM}$ Implied $_{vega,\ elasticity}$ Implied $_{equal\ weight}$ $HIS_{40\ days}$, Implied $_{ITM}$ (ranked, all implied are from calls.)	Option maturity; overlapping periods. Use sample SD of daily returns over option maturity to proxy "actual vol".	RMSE, MAE of overlapping forecasts.	Each day, 5 options were studied; 1 ATM, 2 just in and 2 just out. Define ATM as S=X, OTM marginally outperformed ATM. Mixed together implied of different contract months.
44. Fung and Hsieh (1991)	S&P 500, DM/$ US T-bond Futures and futures options	3/83 - 7/89 (DM/$ futures from 26 Feb 85)	D (15min)	RV-AR(n) Implied $_{BAW\ NTM\ Call/\ Put}$ RV, RW(C-t-C) HL (ranked, some of the differences are small)	1 day ahead. Use 15-min data to construct "actual vol".	RMSE and MAE of log s	RV: Realised vol from 15-min returns. AR(n): autoregressive lags of order n. RW(C-t-C): random walk forecast based on close to close returns. HL: Parkinson's daily high-low method. Impact of 1987 crash does not appear to be drastic possibly due to taking log. In general, high frequency data improves forecasting power greatly.
45. Gemmill (1986)	13 UK stocks LTOM options. Stock price	May78 - Jul83 Jan78-Nov83	M D	Implied $_{ITM}$ Implied $_{ATM,\ vega\ WLS}$ Implied $_{equal,\ OTM,\ elasticity}$ $HIS_{20\ weeks}$ (ranked, all implied are from calls.)	13-21 non-overlapping option maturity (each average 19 weeks). Use sample SD of weekly returns over option maturity to proxy "actual vol".	ME, RMSE, MAE aggregated across stocks and time. R^2 are 6-12% (pooled) and 40% (panel with firm specific intercepts). All a >0, b <1.	Adding HIS increases R^2 from 12% to 15%. But *ex ante* combined forecast from HIS and Implied $_{ITM}$ turned out to be worst then individual forecasts. Suffered small sample and nonsynchroneity problems and omitted dividends.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
46. Gray (1996)	US 1M T-Bill	1/70 – 4/94	W	RSGARCH with time varying probability GARCH Constant Variance (ranked)	1 week ahead (model not re-estimated). Use weekly squared deviation to proxy volatility.	R^2 calculated without constant term, is 4 to 8% for RSGARCH, negative for some CV and GARCH. Comparable RMSE and MAE between GARCH & RSGARCH.	Volatility follows GARCH and CIR square root process. Interest rate rise increases probability of switching into high volatility regime. Low volatility persistence and strong rate level mean reversion at high volatility state. At low volatility state, rate appears random walk and volatility is highly persistence.
47. Guo (1996a)	PHLX US$/¥ options	Jan91 – Mar93	D	Implied $_{Heston}$ Implied $_{HW}$ Implied $_{BS}$ GARCH HIS$_{60}$ (ranked)	Information not available.	Regression with robust SE. No information on R^2 and forecast biasness.	Use mid of bid-ask option price to limit 'bounce' effect. Eliminate 'nonsynchroneity' by using simultaneous exchange rate and option price. HIS and GARCH contain no incremental information. Implied $_{Heston}$ and Implied $_{HW}$ are comparable and are marginally better than Implied $_{BS}$. *Only have access to abstract.*
48. Guo (1996b)	PHLX US$/¥, US$/DM options Spot rate	Jan86 – Feb93	Tick D	Implied $_{HW}$ (WLS, 0.8 <S/X< 1.2, 20<T<60 days) GARCH(1,1) HIS $_{60\ days}$ (ranked)	60 days ahead. Use sample variance of daily returns to proxy actual volatility.	US$/DM R^2 is 4, 3, 1% for the three methods. (9, 4, 1% for US$/¥:) All forecasts are biased a >0, b <1 with robust SE.	Conclusion same as Guo(1996a). Use Barone-Adesi/Whaley approximation for American options. No risk premium for volatility variance risk. GARCH has no incremental information. Visual inspection of figures suggests implied forecasts lagged actual.
49. Hamid (1998)	S&P 500 futures options	3/83 – 6/93	D	13 schemes (including HIS, Implied cross strike average and intertemporal averages) (ranked, see comment)	Non-overlapping 15, 35 and 55 days ahead	RMSE, MAE	Implied is better than historical and cross strike averaging is better than intertemporal averaging (except during very turbulent periods).
50. Hamilton and Lin (1996)	Excess stock returns (S&P 500 minus T-bill) & Ind Production	1/65 – 6/93	M	Bivariate RSARCH Uivariate RSARCH GARCH+L ARCH+L AR(1) (ranked)	1 month ahead. Use squared monthly residual returns to proxy volatility.	MAE	Found economic recessions drive fluctuations in stock returns volatility. "L" denotes leverage effect. RS model outperformed ARCH/GARCH+L.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
51. Hamilton and Susmel (1994)	NYSE VW stock index	3/7/62 - 29/12/87	W	RSARCH+L GARCH+L ARCH+L (ranked)	1, 4 and 8 weeks ahead. Use squared weekly residual returns to proxy volatility.	MSE, MAE, MSLE, MALE. Errors calculated from variance and log variance.	Allowing up to 4 regimes with t distribution. RSARCH with leverage (L) provides best forecast. Student-t is preferred to GED and Gaussian.
52. Harvey and Whaley (1992)	S&P 100 (OEX)	Oct85 - Jul89	D	Implied $_{ATM\ calls+puts}$ (American binomial, shortest maturity>15 days) (Predict changes in implied.)	1 day ahead implied for use in pricing next day option.	R^2 is 15% for calls and 4% for puts (excluding 1987 crash)	Implied volatility changes are statistically predictable, but market was efficient, as simulated transactions (NTM call and put and delta hedged using futures) did not produce profit.
53. Heynen and Kat (1994)	7stock indices and 5 exchange rates	1/1/80-31/12/92 In: 80-87 Out: 88-92 (87's crash included in in-sample)	D	SV(?) EGARCH GARCH RW (ranked, see also comment)	Non-overlapping 5, 10, 15, 20, 25, 50, 75, 100 days horizon with constant update of parameters estimates. Use sample standard deviations of daily returns to proxy "actual vol".	MedSE	SV appears to dominate in index but produces errors that are 10 times larger than (E)GARCH in exchange rate. The impact of 87's crash is unclear. Conclude that volatility model forecasting performance depends on the asset class.
54. Hol and Koopman (2002)	S&P 100 (VIX)	2/1/86 – 29/6/2001 Out: Jan97 – Jun01	D	SIV SVX$^+$ SV (ranked)	1,2, 5, 10, 15 and 20 days ahead. Use 10-min returns to construct "actual vol".	R^2 ranges between17 to 33%, MSE, MedSE, MAE. a and b not reported. All forecasts underestimate actuals.	SVX is SV with implied $_{VIX}$ as an exogenous variable while SVX$^+$ is SVX with persistence adjustment. SIV is stochastic implied with persistence parameter set equal to zero.
55. Hwang and Satchell (1998)	LIFFE stock options	23/3/92 – 7/10/96 240 daily out-of-sample forecasts.	D	Log-ARFIMA-RV Scaled truncated Detrended unscaled truncated MA$_{opt\ n=20}$-IV Adj MA$_{opt\ n=20}$-RV GARCH-RV (ranked, forecast implied)	1, 5, 10, 20, …, 90, 100, 120 days ahead IV estimated from a rolling sample of 778 daily observations. Different estimation intervals were tested for robustness.	MAE, MFE.	Forecast implied $_{ATM\ BS}$ of shortest maturity option (with at 15 trading days to maturity). Build MA in IV and ARIMA on log (IV). Error statistics for all forecasts are close except those for GARCH forecasts. The scaling in Log-ARFIMA-RV is to adjust for Jensen inequality.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
56. Jorion (1995)	DM/\$, ¥/\$, SrFr/\$ futures options on CME	1/85 – 2/92 7/86 – 2/92 3/85 – 2/92	D	Implied $_{\text{ATM BS call+put}}$ GARCH (1,1), MA$_{20}$ (ranked)	1 day ahead & option maturity. Use squared returns and aggregate of square returns to proxy actual volatility.	R^2 is 5% (1-day) or 10-15% (option maturity). With robust SE, a $_{\text{implied}}$ >0 and b $_{\text{implied}}$<1 for long horizon and is unbiased for 1-day forecasts.	Implied is superior to the historical methods and least biased. MA and GARCH provide only marginal incremental information.
57. Jorion (1996)	DM/\$ futures options on CME	Jan85-Feb92	D	Implied$_{\text{Black, ATM}}$ GARCH(1,1) (ranked)	1 day ahead, use daily squared to proxy actual volatility.	R^2 about 5%. H$_0$: a $_{\text{implied}}$=0, b $_{\text{implied}}$=1 cannot be rejected with robust SE.	R^2 increases from 5% to 19% when unexpected trading volume is included. Implied volatility subsumed information in GARCH forecast, expected futures trading volume and bid-ask spread.
58. Karolyi (1993)	74 stock options	13/1/84 - 11/12/85	M	Bayesian Implied $_{\text{Call}}$ Implied $_{\text{Call}}$ HIS$_{20,60}$ (Predict option price not "actual vol".)	20 days ahead volatility.	MSE	Bayesian adjustment to implied to incorporate cross sectional information such as firm size, leverage and trading volume useful in predicting next period option price.
59. Klaassen (1998)	US\$/£, US\$/DM and US\$/¥	3/1/78 – 23/7/97 Out: 20/10/87-23/7/97	D	RSGARCH RSARCH GARCH(1,1) (ranked)	1 and 10 days ahead. Use mean adjusted 1- and 10-day return squares to proxy actual volatility.	MSE of variance, regression though R^2 is not reported.	GARCH(1,1) forecasts are more variable than RS models. RS provides statistically significant improvement in forecasting volatility for US\$/DM but not the other exchange rates.
60. Kroner, Kneafsey and Claessens (1995)	Futures options on Cocoa, cotton, corn, gold, silver, sugar, wheat Futures prices	Jan87 - Dec90 (Kept last 40 observations for out of sample forecast) Jan87 – Jul91	D	GR>COMB Implied $_{\text{BAW Call}}$ (WLS>AVG>ATM) HIS $_{\text{7 weeks}}$ > GARCH (ranked))	225 calendar day (160 working days) ahead, which is longer than average.	MSE, ME	GR: Granger and Ramanathan (1984)'s regression weighted combined forecast, COMB: lag implied in GARCH conditional variance equation. Combined method is best suggests option market inefficiency.
61. Lamoureux and Lastrapes (1993)	Stock options for 10 non-dividend paying stocks (CBOE)	19/4/82-31/3/84	D	Implied $_{\text{Hull-White NTM Call}}$ (intermediate term to maturity, WLS) HIS$_{\text{updated expanding estimate}}$ GARCH (ranked, based on regression result)	90 to 180 days matching option maturity estimated using rolling 300 observations and expanding sample. Use sample variance of daily returns to proxy "actual vol".	ME, MAE, RMSE. Average implied is lower than actual for all stocks. R^2 on variance varies between 3-84% across stocks and models.	Implied volatility is best but biased. HIS provides incremental info to implied and has the lowest RMSE. When all three forecasts are included; a >0, 1> b $_{\text{implied}}$>0, b $_{\text{GARCH}}$=0, b $_{\text{HIS}}$<0 with robust SE. Plausible explanations include option traders overreact to recent volatility shocks, and volatility risk premium is non-zero and time varying.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
62. Latane and Rendleman (1976)	24 stock options from CBOE	5/10/73 – 28/6/74	W	Implied $_{vega\ weighted}$ HIS$_{4\ years}$ (ranked)	In-sample forecast and forecast that extend partially into the future. Use weekly and monthly returns to calculate actual volatility of various horizons.	Cross-section correlation between volatility estimates for 38 weeks and a 2-year period.	Used European model on American options and omitted dividends. "Actual" is more correlated (0.686) with "Implied" than HIS volatility (0.463) Highest correlation is that between implied and actual standard deviations which were calculated partially into the future.
63. Lee (1991)	$/DM, $/£, $/¥, $/FFr, $/C$ (Fed Res Bulletin)	7/3/73-4/10/89 Out: 21Oct81 – 11Oct89.	W (Wed, 12pm)	Kernel (Gaussian, Truncated) Index (combining ARMA and GARCH) EGARCH (1,1) GARCH (1,1) IGARCH with trend (rank changes see comment for general assessment)	1 week ahead (451 observations in sample and 414 observations out-of-sample)	RMSE, MAE. It is not clear how actual volatility was estimated.	Nonlinear models are, in general, better than linear GARCH. Kernel method is best with MAE. But most of the RMSE and MAE are very close. Over 30 kernel models were fitted, but only those with smallest RMSE and MAE were reported. It is not clear how the non-linear equivalence was constructed. Multi-step forecast results were mentioned but not shown.
64. Li (2002)	$/DM, $/£, $/¥ OTC ATM options $/£, $/¥ $/DM	3/12/86-30/12/99 In: 12/8/86- 11/5/95 19/6/94-13/6/99 19/6/94-30/12/98	Tick (5 min) D D	Implied $_{GK\ OTC\ ATM}$ ARFIMA $_{realised}$ (Implied better at shorter horizon and ARFIMA better at long horizon.)	1, 2, 3 and 6 months ahead. Parameters not re-estimated. Use 5-min returns to construct "actual vol".	MAE. R^2 ranges 0.3-51% (Implied), 7.3-47%(LM), 16-53% (emcompass). For both models, H$_0$:a =0, b =1 are rejected and typically b <1 with robust SE.	Both forecasts have incremental information especially at long horizon. Forcing: a =0, b =1 produce low/negative R^2 (especially for long horizon). Model realised standard deviation as ARFIMA without log transformation and with no constant, which is awkward as a theoretical model for volatility.
65. Lopez (2001)	C$/US$, DM/US$, ¥/US$, US$/£	1980-1995 In: 1980-1993 Out: 1994-1995	D	SV-AR(1)-normal GARCH-gev EWMA-normal GARCH-normal, -t EWMA-t AR(10)-Sq, -Abs Constant (approx rank, see comments)	1 day ahead and probability forecasts for four "economic events", viz. cdf of specific regions. Use daily squared residuals to proxy volatility. Use empirical distribution to derive cdf.	MSE, MAE, LL, HMSE, GMLE and QPS (quadratic probability scores).	LL is the logarithmic loss function from Pagan and Schwert (1990), HMSE is the heteroskedasticity-adj MSE from Bollerslev and Ghysels (1996) and GMLE is the Gaussian quasi-ML function from Bollerslev, Engle and Nelson (1994). Forecasts from all models are indistinguishable. QPS favours SV-n, GARCH-g and EWMA-n.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
66. Loudon, Watt and Yadav (2000)	FT All Share	Jan71 – Oct97 Sub-periods: Jan71 - Dec80 Jan81 – Dec90 Jan91 – Oct97	D	EGARCH, GJR, TS-GARCH, TGARCH NGARCH, VGARCH, GARCH, MGARCH (No clear rank, forecast GARCH vol)	Parameters estimated in period 1 (or 2) used to produce conditional variances in period 2 (or 3). Use GARCH squared residuals as "actual" volatility.	RMSE, regression on log volatility and a list of diagnostics. R^2 is about 4% in period 2 and 5% in period 3.	TS-GARCH is an absolute return version of GARCH. All GARCH specifications have comparable performance though non-linear, asymmetric versions seem to fare better. Multiplicative GARCH appears worst, followed by NGARCH and VGARCH (Engle and Ng 1993).
67. Martens and Zein (2002)	S&P500 futures, ¥/US$ futures, Crude oil futures	Jan 94 - Dec 2000 Jan 96 - Dec 2000 Jun 93 - Dec 2000	Tick	Implied $_{BAW\ VIX\ style}$ Log-ARFIMA GARCH (ranked, see comment also)	Non-overlapping 1, 5, 10, 20, 30 and 40 days ahead. 500 daily observations in in-sample which expands on each iteration.	Heteroskedasticity adjusted RMSE. R^2 ranges 25-52% (implied), 15-48% (LM) across assets and horizons. Both models provide incremental info to encompassing regr.	Scaled down one large oil price. Log-ARFIMA truncated at lag 100. Based on R^2, Implied outperforms GARCH in every case, and beats Log-ARFIMA in ¥/US$ and Crude oil. Implied has larger HRMSE than Log-ARFIMA in most cases. *Difficult to comment on implied's biasness from information presented.*
68. McKenzie (1999)	21 A$ bilateral exchange rates	Various length from 1/1/86 or 4/11/92 to 31/10/95	D	Square vs. power transformation (ARCH models with various lags. See comment for rank.)	1 day ahead absolute returns.	RMS, ME, MAE. Regressions suggest all ARCH forecasts are biased. No R^2 was reported.	The optimal power is closer to 1 suggesting squared return is not the best specification in ARCH type model for forecasting purpose.
69. McMillan, Speight and Gwilym (2000)	FTSE100 FT All Share	Jan84 – Jul96 Jan69 – Jul96 Out: 1996-1996 for both series.	D, W, M	RW, MA, ES, EWMA GARCH, TGARCH, EGARCH, CGARCH HIS, Regression, (ranked)	j=1 day, 1 week and 1 month ahead based on the three data frequencies. Use j period squared returns to proxy actual volatility.	ME, MAE, RMSE for symmetry loss function. MME(U) & MME(O), mean mixed error that penalize under/over predictions.	CGARCH is the component GARCH model. Actual volatility is proxied by mean adjusted squared returns, which is likely to be extremely noisy. Evaluation conducted on variance, hence forecast error statistics are very close for most models. RW, MA, ES dominate at low frequency and when crash is included. Performances of GARCH models are similar though not as good.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
70. Noh, Engle and Kane (1994)	S&P 500 index options	Oct85 – Feb92	D	GARCH adj for weekend & hols Implied$_{BS}$ weighted by trading volume (ranked, predict option price not "actual vol")	Option maturity. Based on 1,000 days rolling period estimation	Equate forecastibility with profitability under the assumption of an inefficient option market	Regression with call+put implieds, daily dummies and previous day returns to predict next day implied and option prices. Straddle strategy is not vega neutral even though it might be delta neutral assuming market is complete. It is possible that profit is due to now well-documented post 87's crash higher option premium.
71. Pagan and Schwert (1990)	US stock market	1834-1937 Out: 1900-1925 (low volatility), 1926-1937 (high volatility)	M	EGARCH(1,2) GARCH(1,2) 2-step conditional variance RS-AR(m) Kernel (1 lag) Fourier (1 or 2 lags) (ranked)	One month ahead. Use squared residual monthly returns to proxy actual volatility.	R^2 is 7-11% for 1900-25 and 8% for 1926-37. Compared with R^2 for variance, R^2 for log variance is smaller in 1900-25 and larger in 1926-37.	The nonparametric models fared worse than the parametric models. EGARCH came out best because of the ability to capture volatility asymmetry. Some prediction bias was documented.
72. Pong, Shackleton, Taylor and Xu (2002)	US$/£	In: Jul87-Dec93 Out: Jan94– Dec98	5-, 30-minute	Implied$_{ATM, OTC\ quote}$ (bias adj using rolling regr on last 5 years monthly data) Log-ARMA(2,1) Log-ARFIMA(1,d,1) GARCH(1,1) (ranked)	1 month and 3 month ahead at 1-month interval	ME, MSE, regression. R^2 ranges between 22-39% (1-month) and 6-21% (3-month)	Implied, ARMA and ARFIMA have similar performance. GARCH(1,1) clearly inferior. Best combination is Implied+ARMA(2,1). Log-AR(FI)MA forecasts adjusted for Jensen inequality. *Difficult to comment on implied's biasness from information presented.*
73. Poteshman (2000)	S&P 500 (SPX) & futures S&P 500	1Jun88–29Aug97 Heston estimation: 1Jun93-29Aug97 7Jun62-May93	D futures Tick M	Implied$_{Heston}$ Implied$_{BS}$ (both implieds are from WLS of all options <7 months but >6 calendar days) HIS$_{1, 2, 3,6\ months}$ (ranked)	Option maturity (about 3.5 to 4 weeks, non-overlapping). Use 5-min futures inferred index return to proxy "actual vol".	BS R^2 is over 50%. Heston implied produced similar R^2 but very close to being unbiased.	*F* test for H$_0$:a_{BS}=0, b_{BS}=1 are rejected though *t*-test supports H$_0$ on individual coefficients. Show biasness is not caused by bid-ask spread. Using ln s, high frequency realised vol, and Heston model, all help to reduce implied biasness.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
74. Randolph and Najand (1991)	S&P500 futures options ATM Calls only	2/1/1986-31/12/88 (Crash included) In: First 80 observations	Daily opening Tick	MRM $_{ATM\ HIS}$ MRM $_{ATM\ implied}$ GARCH(1,1) HIS$_{20\ day}$ Implied$_{Black}$ (ranked though the error statistics are close)	Non-overlapping 20 days ahead, re-estimated using expanding sample.	ME, RMSE, MAE, MAPE	Mean reversion model (MRM) sets drift rate of volatility to follow a mean reverting process taking implied $_{ATM}$ (or HIS) as the previous day vol. Argue that GARCH did not work as well because it tends to provide a persistent forecast, which is valid only in period when changes in vol are small.
75. Schmalensee and Trippi (1978)	6 CBOE stock options	29/4/74 – 23/5/75 56 weekly observations	W	Implied$_{BS\ call}$ (simple average of all strikes and all maturities.) (Forecast implied not actual volatility.)	1 week ahead. "Actual" proxied by weekly range and average price deviation.	Statistical tests reject the hypothesis that IV responds positively to current volatility.	Find implied rises when stock price falls, negative serial correlation in changes of IV and a tendency for IV of different stocks to move together. Argue that IV might correspond better with *future* volatility.
76. Scott and Tucker (1989)	DM/$, £/$, C$/$, ¥/$ & SrFr/$ American options on PHLX	14/3/83-13/3/87 (Pre crash)	Daily closing tick	Implied$_{GK}$ (vega, Inferred ATM, NTM) Implied$_{CEV}$ (similar rank)	Non-overlapping option maturity: 3, 6 and 9 months. Use sample SD of daily returns to proxy "actual vol".	MSE, R^2 ranges from 42 to 49%. In all cases, a >0, b <1. HIS has no incremental info content.	Simple B-S forecasts just as well as sophisticated CEV model. Claimed omission of early exercise is not important. Weighting scheme does not matter. Forecasts for different currencies were mixed together.
77. Sill (1993)	S&P 500	1959-1992	M	HIS with exo variables HIS (See comment)	1 month ahead	R^2 increase from 1% to 10% when additional variables were added.	Volatility is higher in recessions than in expansions, and the spread between commercial-paper and T-Bill rates predict stock market volatility.
78. Szakmary, Ors, Kim and Davidson (2002)	Futures options on S&P500, 9 interest rates, 5 currency, 4 energy, 3 metals, 10 agriculture 3 livestock	Various dates between Jan83-May2001	D	Implied$_{BK,\ NTM}$ $_{2Calls+2Puts\ eq\ al\ weight}$ HIS$_{30}$, GARCH (ranked)	Overlapping option maturity, shortest but >10 days. Use sample SD of daily returns over forecast horizon to proxy "actual vol".	R^2 smaller for financial (23-28%), higher for metal & agricult (30-37%), highest for livestock & energy (47,58%)	HIS$_{30}$ and GARCH have little or no incremental information content. a $_{implied}$>0 for 24 cases (or 69%), all 35 cases b $_{implied}$<1 with robust SE.
79. Taylor JW (2001)	DAX, S&P500, Hang Seng, FTSE100, Amsterdam EOE, Nikkei, Singapore All Share	6/1/88-30/8/95 (equally split between in- and out-)	W	STES (E, AE, EAE) GJR (+Smoothed variations) GARCH MA$_{20\ weeks}$, Riskmetrics (ranked)	1 week ahead using a moving window of 200 weekly returns. Use daily squared residual returns to construct weekly "actual" volatility.	ME, MAE, RMSE, R^2 (about 30% for HK and Japan and 6% for US)	Models estimated based on minimizing in-sample forecast errors instead of ML. STES-EAE (smooth transition exponential smoothing with return and absolute return as transition variables) produced consistently better performance for 1-step ahead forecasts.

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
80. Taylor SJ (1986)	15 US stocks FT30 6 metal £/$ 5 agricultural futures 4 interest rate futures	Jan66 – Dec 76 Jul75 – Aug82 Various length Nov74 – Sep 82 Various length Various length	D	EWMA Log-AR(1) ARMACH-Abs ARMACH-Sq HIS (ranked) ARMACH-Sq is similar to GARCH	1 and 10 days ahead absolute returns. 2/3 of sample used in estimation. Use daily absolute returns deviation as "actual vol".	Relative MSE	Represent one of the earliest studies in ARCH class forecasts. The issue of volatility stationarity is not important when forecast over short horizon. Non-stationary series (e.g. EWMA) has the advantage of having fewer parameter estimates and forecasts respond to variance change fairly quickly.
81. Taylor SJ (1987)	DM/$ futures	1977-83	D	High, low and closing prices (see comment)	1, 5, 10 & 20 days ahead. Estimation period, 5 years.	RMSE	Best model is a weighted average of present and past high, low and closing prices with adjustments for weekend and holiday effects.
82. Taylor SJ and Xu (1997)	DM/$ DM/$ options on PHLX	1/10/92 – 30/9/93 In: 9 months Out: 3 months	Quote D	Implied + ARCH combined Implied, ARCH HIS $_{9\ months}$ HIS $_{last\ hour\ realised\ vol}$ (ranked) See comment for details on implied & ARCH	1 hour ahead estimated from 9 months in-sample period. Use 5-min returns to proxy "actual vol".	MAE and MSE on std deviation & variance Friday macro news seasonal factors have no impact on forecast accuracy.	5-min return has information incremental to daily implied when forecasting hourly volatility. ARCH model includes with hourly and 5-min returns in the last hr plus 120 hour/day/week seasonal factors. Implied derived from NTM shortest maturity (>9 calendar days) Call+Put using BAW.
83. Tse (1991)	Topix Nikkei Stock Average	In: 1986-1987 Out: 88-89	D	EWMA HIS ARCH, GARCH (ranked)	25 days ahead estimated from rolling 300 observations	ME, RMSE, MAE, MAPE of variance of 21 non-overlapping 25-day periods.	Use dummies in mean equation to control for 1987 crash. Non-normality provides a better fit but a poorer forecast. ARCH/GARCH models are slow to react to abrupt change in volatility. EWMA adjust to changes very quickly.
84. Tse and Tung (1992)	Singapore, 5 VW market & industry indices	19/3/75 to 25/10/88	D	EWMA HIS GARCH (ranked)	25 days ahead estimated from rolling 425 observations	RMSE, MAE	EWMA is superior, GARCH worst. Absolute returns > 7% are truncated. Sign of non-stationarity. Some GARCH non-convergence

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
85. Vasilellis and Meade (1996)	Stock options 12 UK stocks (LIFFE)	In: 28/3/86 – 27/6/86 In2 (for combine forecast): 28/6/86 – 25/3/88 Out: 6/7/88 – 21/9/91	W	Combine(Implied + GARCH) Implied (various, see comment) GARCH EWMA HIS $_{3\ months}$ (ranked, results not sensitive to basis use to combine)	3 months ahead. Use sample SD of daily returns to proxy "actual vol".	RMSE	Implied: 5-day average dominates 1-day implid vol. Weighting scheme: max vega>vega weighted>elasticity weighted>max elasticity with '>' indicates better forecasting performance. Adjustment for div & early exercise: Rubinstein>Roll> constant yield. Crash period might have disadvantaged time series methods.
86. Vilasuso (2002)	C$/$, FFr/$, DM/$, ¥/$, £/$	In: 13/3/79 – 31/12/97 Out: 1/1/98 – 31/12/99	D	FIGARCH GARCH, IGARCH (ranked, GARCH marginally better than IGARCH)	1, 5 and 10 days ahead. Used daily squared returns to proxy actual volatility.	MSE, MAE, & Diebold-Mariano's test for sig. difference.	Significantly better forecasting performance from FIGARCH. Built FIARMA (with a constant term) on conditional variance without taking log. Truncated at lag 250.
87. Walsh and Tsou (1998)	Australian indices: VW20, VW50 & VW300	1Jan93-31Dec95 In: 1 year Out: 2 years	5-min to form H, D & W returns	EWMA GARCH (not for weekly returns) HIS, IEV (Improved extreme-value method) (ranked)	1 hour, 1 day and 1 week ahead estimated from a 1-year rolling sample. Use square of price changes (non cumulative) as "actual vol".	MSE, RMSE, MAE, MAPE	Index with larger number of stock is easier to forecast due to diversification, but gets harder as sampling interval becomes shorter due to problem of non-synchronous trading. None of the GARCH estimations converged for the weekly series, probably too few observations.
88. Wei and Frankel (1991)	SrFr/$, DM/$, ¥/$, £/$ options (PHLX) Spot rates	2/83–1/90	M D	Implied $_{GK\ ATM\ call}$ (shortest maturity)	Non-overlapping 1 month ahead. Use sample SD of daily exchange rate return to proxy "actual vol"	R^2 30%(£), 17% (DM), 3%(SrFr), 0%(¥). a >0, b <1 (except that for £/$, a >0, b =1) with heteroske consistent SE.	Use European formula for American style option. Also suffers from non-synchronicity problem. Other tests reveal that Implied tends to over-predict high vol and under-predict low vol. Forecast/implied could be made more accurate by placing more weight on long run average.
89. West and Cho (1995)	C$/$, FFr/$, DM/$, ¥/$, £/$	14/3/73-20/9/89 In: 14/3/73 – 17/6/81 Out: 24/6/81 – 12/4/89	W	GARCH(1,1) IGARCH(1,1) AR(12) in absolute AR(12) in squares Homoskedastic Gaussian kernel (No clear rank)	J=1, 12, 24 weeks estimated from rolling 432 weeks. Use j period squared returns to proxy actual volatility.	RMSE and regression test on variance, R^2 varies from 0.1% to 4.5%.	Some GARCH forecasts mean revert to unconditional variance in 12- to 24-weeks. It is difficult to choose between models. Nonparametric method came out worst though statistical tests for do not reject null of no significance difference in most cases..

Author(s)	Asset(s)	Data Period	Data Freq	Forecasting methods & rank	Forecasting horizon	Evaluation & R-square	Comments
90. Wiggins (1992)	S&P 500 futures	4/82–12/89	D	ARMA model with 2 types of estimators: 1. Parkinson/Garmen-Klass extreme value estimators 2. Close-to-close estimator (ranked)	1 week ahead and 1 month ahead. Compute actual volatility from daily observations.	Bias test, efficiency test, regression	Modified Parkinson approach is least biased. C-t-C estimator is three times less efficient than EV estimators. Parkinson estimator is also better than C-t-C at forecasting. 87's crash period excluded from analysis.
91. Xu and Taylor (1995)	£/$, DM/$, ¥/$ & SrFr/$ PHLX options Corresponding futures rates	In: Jan85-Oct89 Out: 18Oct89-4Feb92	D	Implied $_{BAW\ NTM\ TS\ or\ short}$ GARCH$_{Normal\ or\ GED}$ HIS $_{4\ weeks}$ (ranked)	Non-overlapping 4 weeks ahead, estimated from a rolling sample of 250 weeks daily data. Use cumulative daily squared returns to proxy "actual vol".	ME, MAE, RMSE. When $a_{implied}$ is set equal to 0, $b_{implied}=1$ cannot be rejected.	Implied works best and is unbiased. Other forecasts have no incremental information. GARCH forecast performance not sensitive to distributional assumption about returns. The choice of implied predictor (term structure, TS, or short maturity) does not affect results.
92. Yu (2002)	NZSE40	Jan80 - Dec98 In: 1980-1993 Out: 1994-1998	D	SV (of log variance) GARCH(3,2), GARCH(1,1) HIS, MA $_{5yr\ or\ 10\ yr}$ ES and EWMA (monthly revision) Regression $_{lag-1}$ ARCH(9), RW, (ranked)	1 month ahead estimated from previous 180 to 228 months of daily data. Use aggregate of daily squared returns to construct actual monthly volatility.	RMSE, MAE, Theil-U and LINEX on variance	Range of the evaluation measures for most models is very narrow. Within this narrow range, SV ranked first, performance of GARCH was sensitive to evaluation measure; regression and EWMA methods did not perform well. Worst performance from ARCH(9) and RW. Volatile periods (Oct 87 and Oct 97) included in in- and out-of samples.
93. Zumbach (2002)	USD/CHF, USD/JPY	1/1/89 – 1/7/2000	H	LM-ARCH F-GARCH GARCH And their integrated counterparts (ranked)	1 day ahead estimated from previous 5.5 years	RMSE. Realized volatility measured using hourly returns.	LM-ARCH, aggregates high frequency squared returns with a set of power law weights, is the best though difference is small. All integrated versions are more stable across time.

Ranked: models appear in the order of forecasting performance; best performing model at the top. If two weighting schemes or two forecasting models appear at both sides of ">", it means the LHS is better than the RHS in terms of forecasting performance. **SE**: Standard error.

ATM: At the money, **NTM**: Near the money, **OTM**: Out of the money, **WLS**: an implied volatility weighting scheme used in Whaley (1982) designed to minimize the pricing errors of a collection of options. In some cases the pricing errors are multiplied by trading volume or vega to give ATM implied a greater weight.

HIS: Historical volatility constructed based on past variance/standard deviation. **VIX**: Chicago Board of Option Exchange's volatility index derived from S&P 100 options. **RS**: Regime Switching,

BS: Black-Scholes, **BK**: Black model for pricing futures option. **BAW**: Barone-Adesi and Whaley American option pricing formula, **HW**: Hull and White option pricing model with stochastic volatility, **FW**: Fleming and Whaley (1994) modified binomial method that takes into account wildcard option, **GK**: Garman and Kohlhagan model for pricing European currency option. **HJM**: Heath, Jarrow and Morton (1992) forward rate model for interest rates.