

The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks

Stephen J. Taylor, Pradeep K. Yadav, and Yuanyuan Zhang*

Department of Accounting and Finance, Management School, Lancaster University, Lancaster LA1 4YX, United Kingdom

December 2007

Abstract

The volatility information content of stock options for individual firms is measured using option prices for 149 U.S. firms during the period from January 1996 to December 1999. Volatility forecasts defined by historical stock returns, at-the-money (ATM) implied volatilities and model-free (MF) volatility expectations are compared for each firm. The recently developed model-free volatility expectation incorporates information across all strike prices, and it does not require the specification of an option pricing model.

Our analysis of ARCH models shows that, for one-day-ahead estimation, historical estimates of conditional variances outperform both the ATM and the MF volatility estimates extracted from option prices for more than one-third of the firms. This result contrasts with the consensus about the informational efficiency of options written on stock indices; several recent studies find that option prices are more informative than daily stock returns when estimating and predicting index volatility. However, for the firms with the most actively traded options, we do find that the option forecasts are nearly always more informative than historical stock returns.

* *E-mail addresses:* s.taylor@lancaster.ac.uk (S.J. Taylor), p.yadav@lancaster.ac.uk (P.K. Yadav), y.zhang18@lancaster.ac.uk (Y.Zhang).

When the prediction horizon extends until the expiry date of the options, our regression results show that the option forecasts are more informative than forecasts defined by historical returns for a substantial majority (86%) of the firms. Although the model-free (MF) volatility expectation is theoretically more appealing than alternative volatility estimates and has been demonstrated to be the most accurate predictor of realized volatility by Jiang and Tian (2005) for the S&P 500 index, the results for our firms show that the MF expectation only outperforms both the ATM implied volatility and the historical volatility for about one-third of the firms. The firms for which the MF expectation is best are not associated with a relatively high level of trading in away-from-the-money options.

JEL classifications: C22; C25; G13; G14

Keywords: Volatility; Stock options; Information content; Implied volatility; Model-free volatility expectations; ARCH models

1 Introduction

The information content of option implied volatilities have been widely examined by the literature since the introduction of the option pricing model by Black and Scholes (1973). Their option pricing formula provides an easy and fast way to transfer the option price into a market forecast of the average volatility of underlying asset returns over the option's life. When option market participants price options, they incorporate both the historical information and some other information about the future events. So the implied volatilities embedded in the option prices are expected to contain more information of the future volatility than historical predictors.

The ability of option implied volatility to provide better estimates than historical forecasts has been established for stock indices. Jiang and Tian (2005) contribute to the literature by adding the model-free volatility expectation, developed by Britten-Jones and Neuberger (2000), into the comparison and confirm the superior informational efficiency of it when predicting the realized volatility of the S&P500 index.

However, no study has ever comprehensively investigated which is the most informative measure to forecast the future volatility of individual stocks. Volatility estimation and forecasting is well known to be an important area in Finance, especially for risk management and option pricing,. The main purpose of this paper is to help understand more about the informational efficiency of the volatility estimates implicit by the options written on the U.S. individual stocks.

1.1 Prior literature

Numerous empirical studies show the interest to the relative importance of implied and historical volatility predictors [Poon and Granger (2003)]. Day and Lewis (1992) test the information content of the S&P100 index options by incorporating option implied volatility into GARCH and EGARCH models. However, they do not find conclusive evidence that option implied volatilities could subsume all the information provided by the conditional volatility from GARCH or EGARCH models. Lamoureux and Lastrapes (1993) employ the same GARCH specifications that add implied volatility as a factor in the conditional variance equation. Their studies on 10 U.S. individual stocks and with forecast horizon equal to one day reach similar results as Day and Lewis (1992). However, both studies have been criticized for the mismatching between forecast horizons and option time to maturity in later literature.

Another early empirical study by Canina and Figlewski (1993) challenge the usefulness of implied volatility as a guide to the future variability of the S&P100 index returns. They find that implied volatility does not have a statistically significant correlation with realized volatility for most of their subsamples, where subsamples are defined according to option moneyness levels and time to maturities. Even a simple measure of historical volatility outperforms the implied volatility for their data.

The negative conclusions might be caused by a lack of data, mis-measurement of implied volatilities, or inappropriate statistical inference. After correcting various methodological errors, later studies of the S&P 100 index provide a consensus that the

ATM option implied volatility or the near-the-money volatilities are more efficient estimates of the subsequent realized volatility than estimates based solely on historical information. Directly related to Canina and Figlewski (1993), Christensen and Prabhala (1998) re-examine the information content of the S&P 100 index options by adopting non-overlapping monthly volatility observations. Their regression analysis shows that the implied volatility does predict future realized volatility and outperforms the historical forecasts. The same conclusions are drawn by the contemporaneous study also on the S&P 100 index options, by Fleming (1998). Studies that focus on exchange rate options also favour the conclusion that option implied volatility is an efficient estimate of the future realized volatility [Kroner, Keafsey and Claessens (1995), Jorion (1995), Xu and Taylor (1995) and Guo (1996)].

Ederington and Guan (2002) summarize and discuss the regressions results of several earlier studies. Their regressions for the S&P500 futures options conclude that option implied volatility is an informationally efficient estimate of the future realized volatility but it is not an unbiased estimate. In the comparison with the historical forecasts, option implied volatility is more efficient and, for some assets during some sample periods, it subsumes all information contained in its competitor.

More recent studies start to use intraday returns to produce volatility forecast [Andersen, Bollerslev, Diebold and Laybys (2003)], which is proved to be more informative than low-frequency data. Concentrating on the S&P 100 index, Blair, Poon and Taylor (2001) use the CBOE volatility index (VIX)¹ as the measure of option implied volatility and daily and intra-day returns sources to measure historical

¹ Blair, Poon and Taylor (2001) adopt the old VIX index listed on the CBOE. The old VIX is renamed as VOX after September 2003.

volatility. They find no evidence for incremental information contained in daily index returns beyond that provided by the VIX. The implied volatilities in their sample outperform intraday returns for both in-sample estimation and out-of-sample forecasting. Studies using high-frequency foreign exchange rate data include Taylor and Xu (1997), Li (2002) , Pong et al (2004) and Martens and Zein (2004).

The most accurate forecasts of stock index volatility were often provided by functions of option implied volatilities in many studies until the introduction of the model-free volatility expectation². Constructive theoretical relationships between volatility and option prices have been developed by Carr and Madan (1998) and Demeterfi, Derman, Kamal and Zou (1999a, 1999b). Building on the pioneering work of Breeden and Lizenberger (1978), they show that the fair value of a variance swap rate, which is the risk-neutral forecast of the subsequent realized variance, can be replicated by taking a static position in options of all strike prices. Likewise, Britten-Jones and Neuberger (2000) show that a complete set of call options can be used to infer the risk-neutral expectation of the integrated variance until the option expiry date, which is a natural input into volatility forecasting. Jiang and Tian (2007) prove that the risk-neutral variance expectation derived by Britten-Jones and Neuberger (2000) is theoretically identical to the equilibrium variance swap rate in Demeterfi, Derman, Kamal and Zou (1999a, 1999b).

This newly-developed volatility measure also extracts information from option prices and thus is potentially superior to historical predictors. It is also expected to be more

² Previous studies, like Britten-Jones and Neuberger (2000) and Jiang and Tian (2005, 2007), define the square root of the risk-neutral variance expectation as the *model-free implied volatility*. We name the same concept as *model-free volatility expectation* in this study, as it is not a volatility measure “implicit” in a single option price but is an estimate of risk-neutral expectation.

efficient than the Black-Scholes implied volatility because, firstly, it does not depend on any option pricing formula; and secondly, it is estimated from all available option prices while Black-Scholes implied volatilities rely on only one or a few near-the-money options. In September 2003, the model-free volatility expectation was adopted by the CBOE for calculating the components of its volatility index (VIX). The definition of the new VIX index and the theory that underpinned both the old and the new are discussed and compared by Carr and Wu (2006).

When comparing the information content and forecasting ability of the model-free volatility expectation with the other volatility estimates, Jiang and Tian (2005) adopt regression analysis on the S&P 500 index options and reach the strong conclusions that the model-free volatility expectation not only outperforms both the ATM implied volatility and the past realized volatility, calculated from intraday index returns, but also subsumes all information that are provided by them.

Lynch and Panigirtzoglou (2004) also examine the information content of the model-free volatility expectation in a comparison with historical volatility measured by intraday returns. Their results for various financial assets, including the S&P 500 index, the FTSE 100 index, Eurodollar and short sterling futures, show that the model-free volatility expectation is more informative than the historical high-frequency returns, but is a biased estimator of future realized volatility.

Although not intended to examine the information efficiency of the model-free volatility expectation, Carr and Wu (2004) synthesize the variance swap rates of 5 stock indices and 35 individual stocks using option prices, following the strategy

suggested by Carr and Madan (1998) and Demeterfi, Derman, Kamal and Zou (1999a, 1999b). Their estimates of the equilibrium variance swap rate, equivalent to the model-free variance expectation, are significant when explaining the time-series movements of realized variance measured by daily returns for all the indices and the majority of the individual stocks.

However, nearly all of the previous studies, whether they use the Black-Scholes implied volatility or the model-free volatility expectation, only investigate the information content of option prices that are written on stock indices or exchange rates. There are very few studies that test the information content of individual stock options. We may anticipate that the volatility information contained in the prices of stock options is less efficient when estimating and predicting volatility, compared with index option prices, since stock options are traded far less frequently. Lamoureux and Lastrapes (1993) study two years data for each of 10 U.S. firms and their results indicate that the simple GARCH(1,1) model is more informative than a model that incorporates implied volatility alone. It is inevitably difficult to draw firm conclusions from their small quantity of data.

1.2 Scope

This paper is the first to examine the volatility information content of individual stock options based on a large sample of U.S. stocks. We develop a method to implement the model-free volatility expectation for individual stock options that are less liquid than options written on stock indices or exchange rates. For each of 149 sample firms, we compare the historical information from daily stock returns, the information

contained in the ATM implied volatility and the information provided by the model-free volatility expectation.

Our empirical results show that both model-free volatility expectation and ATM implied volatility do contain relevant information about future return variations. However, in contrast to previous studies about stock index options, our research on individual stocks shows that for one-day-ahead estimation neither the ATM implied volatility nor the model-free volatility expectation is consistently superior to a simple ARCH model for all firms. Especially for firms with few traded options, it is often better to use an asymmetric ARCH model to estimate the next day's volatility. When the estimation horizon extends until the end of the option lives, it is found that both of the volatility estimates extracted from option prices outperform the historical volatility for a substantial majority of our sample firms.

Interestingly, in our results, the ATM implied volatility outperforms the model-free volatility expectation for 87 out of 149 firms when predicting volatility one-day-ahead, and for 85 firms when the forecast horizon equals the option's life. The relatively poor performance of the model-free volatility expectation compared with the ATM implied volatility for some firms is different from what we expected according to the theory and can not be explained by either selected properties of market available option observations (such as number of option observations, range of option moneyness) or the relative trading liquidity of ATM options to all options. It appears probable that overall illiquidity of the individual stock option market constitutes one explanation for our findings.

The rest of this paper is organized as follows. Section 2 describes the newly developed model-free volatility expectation and related implementation issues. Data descriptions are provided in Section 3. Section 4 describes the empirical methodologies that we choose to compare the information efficiency of different volatility estimates. Section 5 compares the volatility estimates for one-day-ahead forecasts and option life forecasts. In Section 6, 149 firms are separated into different groups according to the most informative volatility estimate for each firm. By doing so, we try to point out the properties of the firms in different groups. Conclusions are presented in Section 7.

2 Model-free volatility expectation and implementation issues

Previous studies have shown that the Black-Scholes implied volatility curve with varying strike prices for a fixed maturity appears to be a U-shape for currency options [Taylor and Xu (1994a)] and to be a “smirk” for equity index options [Rubinstein (1994)]. When using Black-Scholes implied volatilities to define a volatility forecast, we must select one or more of the available option prices and thus lose the information contained in the other prices. This section introduces a new method to construct the risk-neutral expectation of the future integrated variance. This new method also uses the information provided by market option prices but, in contrast to implied volatility forecasts, it incorporates a complete set of option prices and thus is model-independent.

2.1 Theoretical formula

The concept of the model-free variance expectation appeared in Carr and Madan (1998), Demeterfi, Derman, Kamal and Zou (1999a, 1999b) and Carr and Wu (2006) in connection with the development of volatility and variance swap contracts. It was refined and related to volatility forecasting by Britten-Jones and Neuberger (2000).

At time 0 it is supposed that there is a complete set of European option prices for an expiry time T . For a general strike price K , these option prices are denoted by $c(K, T)$ and $p(K, T)$ respectively. For a risk-neutral measure Q , the price of the underlying asset S_t is assumed to satisfy the equation $dS = (r - q)Sdt + \sigma SdW$, where r is the risk-free rate, q is the dividend yield, W_t is a Wiener process and σ_t is the risk-neutral stochastic volatility. The integrated squared volatility of the asset over the horizon T is defined as:

$$V_{0,T} = \int_0^T \sigma_t^2 dt.$$

Britten-Jones and Neuberger (2000) show that the risk-neutral expectation of the integrated squared volatility is given by the following function of the continuum of European out-of-the-money (hereafter OTM) option prices:

$$E^Q[V_{0,T}] = 2e^{rT} \left[\int_0^{F_{0,T}} \frac{p(K, T)}{K^2} dK + \int_{F_{0,T}}^{\infty} \frac{c(K, T)}{K^2} dK \right] \quad (1),$$

where $F_{0,T}$ is the forward price at time 0 for a transaction at the expiry time T .

Following previous literature, the quantity defined by Equation (1) will be referred to as the *model-free variance expectation* and its square root as the *model-free volatility*

expectation³. Dividing the variance expectation by the option time to maturity defines the average integrated variance.

The key assumption required to derive Equation (1) is that the stochastic processes for the underlying asset price and volatility are continuous. As the volatility expectation does not rely on a specific option pricing formula, the expectation is “model-free”, in contrast to the Black-Scholes implied volatility. Both Carr and Wu (2004) and Jiang and Tian (2005) show that Equation (1) is an excellent approximation when there are occasional jumps in stock price process. However, the real options market obviously does not satisfy the assumption of continuous option prices with strike prices ranging from zero to infinity.

2.2 Discrete formula

The CBOE employs a discrete form of the model-free volatility expectation to construct the volatility index, VIX. At time 0, when there are N available OTM options, the model-free volatility expectation with time to maturity T is computed by the equation⁴:

$$\sigma_{MF} = \sqrt{\frac{2}{T} e^{rT} \sum_{i=1}^N \frac{\Delta K_i}{K_i^2} Q(K_i, T) - \frac{1}{T} \left[\frac{F_{0,T}}{K^*} - 1 \right]^2} \quad (2),$$

where K^* is the strike price used to select call or put options, $Q(K_i, T)$ is the call price with strike price K_i when $K_i \geq K^*$ and otherwise it is the put price, and ΔK_i is

³ Previous studies have noticed that taking the squared root of the risk-neutral variance expectation as the expectation of subsequent volatility is subject to a bias, which is determined by the variance of volatility. The bias is positive unless the volatility is actually constant.

⁴ This equation is from the white paper, which is on the CBOE's website.

set equal to $\frac{K_{i+1} - K_{i-1}}{2}$ ⁵. The quantity σ_{MF} is the annualized value of the model-free volatility expectation from time 0 until time T .

The CBOE uses market available option prices directly to estimate Equation (2) and sets K^* as the available strike price just below the forward price $F_{0,T}$, which might incur various estimation errors, as shown by Jiang and Tian (2007). When we employ Equation (2), we use market available strikes to estimate a risk-neutral density and hence we can infer option prices for as many strikes as necessary. Consequently we can set $K^* = F_{0,T}$. Thus the final term in Equation (2) then disappears and $Q(K_i, T)$ always represents an OTM or ATM option price in our calculations.

2.3 Construction of implied volatility curve

As shown in the Equation (1), the model-free volatility expectation is obtained from the integrals of functions of option prices at all strikes. However, stock option prices are usually only available for a small number of strike prices. In order to obtain consistent option prices for a large number of strikes, we must estimate implied volatility curves from small sets of observed option prices. Jiang and Tian (2007) prove the importance of constructing implied volatility curves when estimating the model-free volatility expectation using discrete option data. They also show that the bias from ignoring the step is economically substantial, especially for volatility futures contracts.

⁵ ΔK for the lowest strike is the difference between the lowest strike and the next higher strike. Likewise, ΔK for the highest strike is the difference between the highest strike and the next lower strike.

We implement a variation of the practical strategy described by Malz (1997a, 1997b), who proposed estimating the implied volatility curve as a quadratic function of the Black-Scholes option's delta; previously a quadratic function of the strike price had been suggested by Shimko (1993). As stated by Malz (1997a), making implied volatility a function of delta, rather than of the strike price, has the advantage that the away-from-the-money implied volatilities are grouped more closely together than the near-the-money implied volatilities. Also, extrapolating a function of delta provides sensible limits for the magnitudes of the implied volatilities.

The quadratic specification is chosen because it is the simplest function that captures the basic properties of the volatility smile. Furthermore, there are insufficient stock option prices to estimate higher-order polynomials. Only three strike prices are required to estimate the parameters of a quadratic implied volatility function.

Delta is defined here as the first derivative of the Black-Scholes call option price with respect to the underlying forward price:

$$\Delta_i = \frac{\partial C(F_{0,T}, K_i, T)}{\partial F_{0,T}} = e^{-rT} \Phi(d_1(K_i)) \quad (3),$$

with

$$d_1(K_i) = \frac{\log(F_{0,T}/K_i) + 0.5\sigma^* T^2}{\sigma^* \sqrt{T}},$$

where $C(F_{0,T}, K_i, T)$ is the call option price with strike price K_i and maturity day T .

$\Phi(\cdot)$ is the cumulative probability distribution function. The value of i ranges from 1 to the number of strike prices that are observed for the firm's options at time zero.

Following Bliss and Panigirtzoglou (2002, 2004), σ^* is a constant that permits a convenient one-to-one mapping between delta and the strike price. In this study, σ^* is the volatility implied by the option observations whose strike price is nearest to the forward price, $F_{0,T}$. The value of the call delta, Δ , increases from zero for deep out-of-the-money (OTM) call options to e^{-rT} for deep ITM (in-the-money) call options.

The parameters of the quadratic function have been estimated by minimizing the sum of squared errors function:

$$\sum_{i=1}^N w_i (IV_i - \hat{IV}_i(\Delta_i, \theta))^2 \quad (4),$$

where N is the number of observed strike prices for the firm on the observation day, IV_i is the observed implied volatility for a strike price K_i and $\hat{IV}_i(\Delta_i, \theta)$ is the fitted implied volatility also for strike price K_i , with Δ_i given by Equation (3) and θ the vector of the three parameters defining the quadratic function. The minimization is subject to the constraint that the fitted implied volatility curve is always positive when delta, Δ , is between 0 and e^{-rT} .

The squared errors of the fitted implied volatilities are weighted by $w_i = \Delta_i(1 - \Delta_i)$, to ensure that the most weight is given to near-the-money options. Far-from-the-money options are given low weights because their contracts are less liquid and hence their prices are the most susceptible to non-synchronicity errors. Introducing weights when fitting the quadratic function reduces the impact from any outliers of far-from-the-money options.

After the implied volatility function is fitted, we extract 1000 equally spaced values of delta that cover the range from 0 to e^{-rT} . Equation (3) is used again for transferring delta values back to strike prices and then Black-Scholes option prices are calculated for OTM options. If either the least call price or the least put price exceeds 0.001 cents then we extend the range of strike prices⁶, to eliminate any error caused by truncating integrals. The OTM prices are then used to evaluate the daily values of model-free volatility expectations.

3 Data

The number of U.S. firms with option trading on the CBOE increased dramatically from 1996 to 1999. This provides some motivation for studying the individual stock options market. However, although the market as a whole is developing fast, the trading for each individual firm's options is far less active than the stock index options or exchange rate options. Therefore, research on individual stock options requires a selection of firms that includes those with sufficient option trading activities. This section introduces the data sources and the method to select firms from those listed on the CBOE.

⁶ The extrapolation in either tail occurs with an equal spacing of 0.01 in moneyness, defined as $K/F_{0,T}$, with the implied volatility equal to the appropriate end-point of the quadratic function. It continues until the OTM prices are less than 0.001 cents.

3.1 Sources

The option data used in the following studies are from the IvyDB Database of OptionMetrics, which contains the prices for all U.S. listed equities and market indices and all U.S. listed index and equity options, based on daily closing quotes at the CBOE. Daily stock price data are from CRSP. The dataset also includes interest rate curves, dividend projections. Our sample starts in January 1996 and ends in December 1999. There are 1009 trading days during this period.

We use the implied volatilities provided by IvyDB Database directly in our study, as do Carr and Wu (2004) and Xing, Zhang and Zhao (2007) who choose the same database. Each implied volatility provided is computed from the midpoint of the highest closing bid price and the lowest closing offer price across all exchanges on which the option trades. As all individual stock options on the CBOE are American-style options, the implied volatilities are calculated based on the binomial tree model of Cox, Ross and Rubinstein (1979), which takes account of the early exercise premium and dividends. Whenever call and put implieds are both available, for the same firm, trading day, expiry date and strike price, the average of the two implied volatilities is used. Doing so reduces any measurement errors from nonsynchronous asset and option prices, because the call error is normally negatively correlated with the put error. Options with less than seven days to maturity are excluded in order to avoid any liquidity and market microstructure effects around expiry.

The interest rate that corresponds to each option's expiration is obtained by linearly interpolating between the two closest continuously compounded zero-coupon rates that are derived from BBA LIBOR rates and included by the zero-curve file provided in the IvyDB Database. We calculate the forward stock price $F_{0,T}$ that has the same expiry date, T , as options. It is defined as the future value of the difference between the current spot price and the present value of all future dividend distributions between times 0 and T inclusive. The dividend distribution information is also included in the IvyDB Database. Daily stock returns, taking account of dividends, have been obtained from CRSP for each firm. They are transformed into continuously compounded returns, such that $r_t = \log(1 + r_t^*)$, where r_t^* is the CRSP stock return.

3.2 Selection of firms

All firms with sufficient option trading activity are included in our study. Two criteria are used to select firms from the database. Firstly, only firms that have options written on them throughout the whole sample period are included. So every selected firm has option observations for 1009 trading days from January 1996 to December 1999. Consequently, the comparison of different volatility measures for each firm will not be influenced by either the firms' sample period or sample size.

Secondly, a firm must have sufficient option trading activity, where sufficient is defined by us as enough to construct implied volatility curves for at least 989 (i.e. 98%) of the 1009 trading days. If the firm has too many days of missing data, the

firm's options are considered to be illiquid and then the information content of option prices is expected to reduce.

Our method for constructing the implied volatility curve is described in the next section and requires at least three strike prices and their corresponding implied volatilities to estimate a quadratic function. Whenever there are less than three available strike prices on a trading day, we are unable to construct the implied volatility curve and thus unable to calculate the model-free volatility expectation.

The options with the nearest time-to-maturity are usually chosen. When there are less than three available strike prices for the nearest time-to-maturity, we switch to the second nearest time-to-maturity, which is usually in the month after the trading day. However, when it is impossible to estimate the implied volatility curve for the two nearest-to-maturity sets of option contracts, we assume the estimates from option prices are missing data for that trading day and instead remain unchanged from the previous trading day.

A total of 149 firms pass both filters. The number of market option observations during the sample period varies from firm to firm and from period to period. There are less option observations in 1996, compared with later periods. The maximum number of daily option observations occurs in 1999 for 58 firms.

Figure 1 shows the distribution of the average number of daily available strike prices for the 149 selected firms. The average number for firm i , \bar{N}_i , equals the total number of available strike prices for firm i during the sample period divided by the

number of trading days, which is 1009 for all firms; for those trading days when it is impossible to construct an implied volatility curve because of a lack of observations, the number of available strike prices is set to zero. The minimum, median and maximum values of \bar{N}_i across firms are 3.7, 5.1 and 12.9 respectively. More than a half of the averages \bar{N}_i are between 4 and 6. As shown by Figure 1, it is obvious that the stock options in our sample have far less observations than the stock index options studied in previous literature.

4 Empirical methodology

We investigate the information efficiency of the model-free volatility expectation and the ATM implied volatility, firstly when the forecast horizon is one day and secondly when it is matched with the option's days to maturity. This section firstly introduces the ARCH specifications for one-day-ahead forecasts and the OLS regression models for option life forecasts. The second subsection describes how we compute the volatility estimates under consideration. Finally, the summary statistics of the volatility estimates are presented.

4.1 Econometric specifications

Both ARCH and regression models have been estimated in many previous comparisons of the information content of different volatility estimates. ARCH models can be estimated from daily returns, while regression models are estimated for

a data-frequency that is determined by the expiration dates of the option contracts. The primary advantages of ARCH models are the availability firstly of more observations and secondly of maximum likelihood estimates of the model parameters. A disadvantage of ARCH models, however, is that the data-frequency is usually very different to the forecasting horizon that is implicit in option prices, namely the remaining time until expiry. This fact may weaken the relative performance in an ARCH context of volatility estimates extracted from option prices. To learn as much as we can about the volatility estimates drawn from the option prices, our study therefore evaluates both ARCH specifications for the one-day-ahead forecasts and regressions that employ a forecast horizon equal to the option's time to maturity.

ARCH specifications for one-day-ahead forecasts

To compare the performance of historical daily returns, ATM implied volatilities and model-free volatility expectations, when estimating future volatility, three different ARCH specifications that incorporate different daily information sets are estimated for daily stock returns r_t from January 1996 to December 1999. The specifications include an MA(1) term in the conditional mean equation to capture any first-order autocorrelation in stock returns.

The general specification is as follows:

$$\begin{aligned}
 r_t &= \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \\
 \varepsilon_t &= \sqrt{h_t} z_t, \quad z_t \sim i.i.d.(0,1), \\
 h_t &= \frac{\omega + \alpha\varepsilon_{t-1}^2 + \alpha^- s_{t-1} \varepsilon_{t-1}^2}{1 - \beta L} + \frac{\gamma\sigma_{MF,t-1}^2}{1 - \beta_\gamma L} + \frac{\delta\sigma_{ATM,t-1}^2}{1 - \beta_\delta L}.
 \end{aligned} \tag{5}$$

Here L is the lag operator, h_t is the conditional variance of the return in period t and s_{t-1} is 1 if $\varepsilon_{t-1} < 0$ and it is 0 otherwise. The terms $\sigma_{MF,t-1}$ and $\sigma_{ATM,t-1}$ are respectively the daily estimates of model-free volatility expectation and the ATM implied volatility, computed at time $t-1$ by dividing the annualized values by $\sqrt{252}$.

By placing restrictions on selected parameters in the conditional variance equation, three different volatility models based upon different information sets are obtained:

- (1) The GJR(1,1)-MA(1) model, as developed by Glosten, Jagannathan and Runkle (1993): $\gamma = \beta_\gamma = \delta = \beta_\delta = 0$.
- (2) The model that uses the information provided by model-free volatility expectations alone: $\alpha = \alpha^- = \beta = \delta = \beta_\delta = 0$.
- (3) The model that uses information provided by ATM implied volatilities alone: $\alpha = \alpha^- = \beta = \gamma = \beta_\gamma = 0$.

The parameters are estimated by maximising the quasi-log-likelihood function, defined by assuming that the standardized returns z_t have a Normal distribution. To ensure that the conditional variances of all models remain positive, constraints such as $\omega > 0$, $\alpha \geq 0$, $\alpha + \alpha^- \geq 0$, $\beta \geq 0$, $\beta_\gamma \geq 0$ and $\beta_\delta \geq 0$ are placed on the parameters. Inferences are made through t -ratios, constructed from the robust standard errors of Bollerslev and Wooldridge (1992). The three special cases listed above are ranked by comparing their log-likelihood values; a higher value indicates that the information provides a better description of the conditional distributions of daily stock returns.

Similar models with implied volatility added into ARCH models are also estimated by Day and Lewis (1992) and Blair, Poon and Taylor (2001) on the S&P100 index, Lamoureux and Lastrapes (1993) on individual stocks and Xu and Taylor (1995) on foreign exchange rates. The GJR(1,1) model is adopted here because asymmetric volatility effects have been found for individual U.S. firms by previous studies such as Cheung and Ng (1992) and Duffee (1995). It should be noted that as ARCH conditional volatility is always for the next day or week, to enhance the specification in Equation (5) the implied volatilities obtained from option prices should be chosen to have short times to maturity.

OLS regressions for option life forecasts

Univariate and encompassing regressions are estimated for each firm, as in the index studies by Canina and Figlewski (1993), Christensen and Prabhala (1998) and Jiang and Tian (2005). While a univariate regression can assess the information content of one volatility estimation method, the encompassing regression addresses the relative importance of competing volatility estimates.

The most general regression equation is specified as follows:

$$\sigma_{RE,t,T} = \beta_0 + \beta_{His}\sigma_{His,t,T} + \beta_{MF}\sigma_{MF,t,T} + \beta_{ATM}\sigma_{ATM,t,T} + \varepsilon_{t,T} \quad (6),$$

where $\sigma_{RE,t,T}$ is the measure of the realized volatility from time t to time T , and $\sigma_{His,t,T}$ is a historical volatility forecast calculated from the GJR(1,1)-MA(1) model using the information up to time t . The terms $\sigma_{MF,t,T}$ and $\sigma_{ATM,t,T}$ are non-overlapping measures of the model-free volatility expectation and the ATM implied volatility. Inferences are made using the robust standard errors of White (1980),

which take account of heteroscedasticity in the residual terms $\varepsilon_{t,T}$. Regression results are also obtained when the volatility variables are replaced either by variances or by the natural logarithms of 100 multiplied by volatilities.

4.2 Estimating volatility

Option measures

We use daily estimates of the ATM implied volatility and the model-free volatility expectation in the estimation of the ARCH models for one-day-ahead forecasts. The ATM implied volatility is the implied volatility corresponding to the available strike price closest to the forward price. It is taken from the IvyDB Database directly and is calculated through the binomial tree model as introduced in Section 3.1. The model-free volatility expectation of Equation (2) is calculated by extracting a large number of OTM option prices from the fitted implied volatility curve described in Section 2.3.

Figure 3 shows two time series of daily option measures of volatility, for General Electric and IBM during our sample period. The dark line represents the model-free volatility expectation and the dotted line the ATM implied volatility. It is seen that these two volatility measures move closely with each other and that the ATM implied volatility tends to be slightly lower than the model-free volatility expectation.

To implement regression analysis for option life forecasts, both the model-free volatility expectation and the ATM implied volatility on the trading date that follows the previous maturity date are selected, so that non-overlapping samples of volatility

expectations are obtained. We are able to use sets of 49 monthly observations, with maturity days from January 1996 to January 2000, for 129 of the 149 firms. For each of the remaining 20 firms, the number of observations is 46, 47 or 48 because of the occasional illiquidity of option trading for some firms. To match the horizon of all the variables in the OLS regressions with the one-month horizon of the options information, realized volatility measures and historical volatility forecasts are required for the remaining lives of the options.

Realized volatility

The annualized realized volatility from a day t until the option's maturity date T is calculated by applying the well-known formula of Parkinson (1980) to daily high and low stock prices, such that:

$$\sigma_{RE,t,T} = \sqrt{\frac{252}{H} \sum_{i=1}^H \frac{[\log(h_{t+i}) - \log(l_{t+i})]^2}{4 \log(2)}} \quad (7),$$

where h_t and l_t are, respectively, the highest and lowest stock price for day t , and H is the number of days until the options expire. Parkinson (1980)'s estimator is expected to be a more accurate measure of realized volatility than the sum of daily stock returns, because intraday prices theoretically contain more volatility information than daily closing prices.

Historical volatility

In regression analysis, historical forecasts of volatility are evaluated using the GJR(1,1)-MA(1) model. The historical information up to the observation day t

provides the conditional variance h_{t+1} for day $t+1$. The forecast of the aggregate variance until the expiry time T , whose square root represents the historical volatility forecast, σ_{His} , in the regressions, is given by:

$$h_{t+1} + \sum_{j=2}^H E[h_{t+j} | I_t] \quad (8),$$

where H is the forecast horizon and I_t is the historical information up to day t . For the GJR(1,1) model, the annualized historical volatility forecast simplifies to:

$$\sigma_{His,t,T} = \sqrt{\frac{252}{H} \left[H\sigma^2 + \frac{1-\phi^H}{1-\phi} (h_{t+1} - \sigma^2) \right]} \quad (9),$$

where $\phi = \alpha + \frac{1}{2}\alpha^- + \beta$ and $\sigma^2 = \frac{\omega}{1-\phi}$ are respectively equal to the persistence parameter and the unconditional variance of the returns.

The parameters of the ARCH models used to define the historical forecasts are estimated by maximizing the log-likelihood of a set of n returns that do not go beyond time t . Ninety of the 149 firms have continuous price histories from January 1988 until January 2000. For these firms, we initially use $n = 2024$ returns for the trading days between 4 January 1988 and 4 January 1996, as our first forecasts are made on 4 January 1996; the subsequent forecasts use parameters estimated from the 2024 most recent returns. For each of the other firms, whose histories commence after January 1988, we use all the daily returns until the observation day t (although we stop adding to the historical sample if n reaches 2024).

4.3 Descriptive Statistics

Table 1 presents the summary statistics for daily estimates of model-free volatility expectation, ATM implied volatility and the difference between them. Statistics, including mean, standard deviation, maximum and minimum, are first obtained for each firm from time series of volatility estimates. Then the cross-sectional mean, lower quartile, median and upper quartile values of each statistic, across the 149 firms, are calculated and displayed in Table 1.

On average the model-free volatility expectation is higher than the ATM implied volatility although occasionally the latter is higher than the former. This also occurs in the study by Jiang and Tian (2005) on S&P 500 index options. We therefore conclude that the squared ATM implied volatility tends to be a downward biased measure of the risk-neutral expected variance. The null hypothesis, that the ATM implied volatility is an unbiased estimate of the model-free volatility expectation, is rejected for each of the 149 firms, at the 1% significance level using Student's t statistic.

The numbers in Panel D are the cross-sectional statistics of the correlations between daily σ_{MF} and σ_{ATM} for 149 firms. The two volatility estimates are highly correlated, with the mean and median respectively equal to 0.926 and 0.940. The high values reflect the similar information that is used to price ATM and OTM options.

Table 2 shows the summary statistics for non-overlapping volatility estimates used in the option life forecasts. The numbers, again, are the cross-sectional statistics including mean, lower quartile, median and upper quartile, calculated from time-series statistics. The patterns of model-free volatility expectation and ATM implied volatility are overall similar to those of daily estimates presented in Table 1. We also

show the summary statistics of variance estimates in Panel B, because the model-free volatility expectation obtained as the square root of risk-neutral variance expectation is theoretically higher than the risk-neutral expectation of volatility, according to Jensen's inequality. Regressions involving volatility thus might be affected if the upward bias is large.

The realized volatility measured by daily high and low prices is lower than the other volatility estimates. Parkinson (1980)'s measure of realized volatility is derived under the assumption of continuous trading. Discrete trading both in time and in price can reduce the efficiency of the volatility estimate and induce a downward bias, stated in Garman and Klass (1980). In addition, the reported skewness and excess kurtosis statistics in Table 2 reveal that the natural logarithm of volatility is closest to a Normal distribution. Thus regressions that explain the logarithm of volatility may be statistically more reliable than those that explain volatility or variance.

Table 3 presents the cross-sectional mean, lower quartile, median and upper quartile values of the correlation coefficients, which are calculated using time-series of non-overlapping volatility estimates for each firm. Overall the ATM implied volatility provides the highest correlation values with the realized volatility and the historical volatility the lowest. The correlations of model-free volatility expectation with realized volatility are slightly lower than, but very close to, those of ATM implied volatility. The highest correlations statistics are between model-free estimates and ATM estimates in all panels.

5 Results

This section provides the results of both one-day-ahead forecasts and option life forecasts for 149 sample firms. For each forecast horizon, estimated coefficients for each model, as well as comparisons across different sets of information, are discussed.

5.1 One-day-ahead forecasts from ARCH specifications

Estimates of parameters

Table 4 provides the summary statistics of the sets of 149 point estimates (their mean, standard deviation, lower quartile, median, and upper quartile) from the three ARCH specifications defined by Equation (5). Results are shown in Panel A for the GJR(1,1)-MA(1) model, in Panel B for the model that only uses the information provided by the model-free volatility expectation, and in Panel C for the model that uses the information provided by the ATM implied volatility alone. The column named “10%/5%/1%” contains the percentages of the estimates out of 149 that are significantly different from zero at the 10%, 5% and 1% levels.

The first model is the standard GJR(1,1)-MA(1) model, which uses previous stock returns to calculate the conditional variance. The value of α measures the symmetric impact of new information (defined by ε_t) on volatility while the value of α^-

measures the additional impact of negative information (when $\varepsilon_t < 0$) on volatility.

Approximately 75% of all firms have a value of $\alpha + \alpha^-$ that is more than twice the estimate of α , indicating a substantial asymmetric effect for individual stocks.

For the majority of firms, the estimates of α and α^- are not significantly different from zero at the 5% level. This is probably a consequence of the relatively short sample period. The volatility persistence parameter, assuming returns are symmetrically distributed, is $\alpha + 0.5\alpha^- + \beta$. The median estimate of persistence across 149 firms equals 0.94.

The second model includes only the information contained in the time series of model-free volatility expectations, $\sigma_{VE,t-1}$. The series $\sigma_{VE,t-1}$ is filtered by the function $\gamma/(1 - \beta_\gamma L)$. For half of the firms, the estimates of γ are between 0.48 and 0.85; also, 50.3% of the estimates are significantly different from zero at the 5% level. In contrast, most of the estimates of β_γ are near zero. This suggests that a conditional variance calculated from the model-free volatility expectation given by the latest option prices can not be improved much by using older option prices.

The third model uses only the information contained in the ATM implied volatility series, $\sigma_{ATM,t-1}$, to calculate the conditional variances. The interquartile range for δ is from 0.62 to 1.01 and 42.3% of the estimates are significantly different from zero at the 5% level. More than half of the estimates of the lag coefficient, β_δ , are zero and few of them are far from zero. On average, δ exceeds γ and β_δ is less than β_γ .

The total weight in the conditional variance equation given to the model-free volatility expectations and the ATM implied volatilities are respectively defined by the quantities $\gamma/(1-\beta_\gamma)$ and $\delta/(1-\beta_\delta)$. A higher value of these quantities may imply the information provided is more relevant to the conditional variance movements, or it may also indicate a lower level of the volatility estimates. Figure 3 is a scatter diagram of these two variables for the 149 firms. It is seen that there is a strong positive correlation between these two variables. Most points are closer to the x-axis than to the y-axis because $\delta/(1-\beta_\delta)$ is usually higher than $\gamma/(1-\beta_\gamma)$.

Comparisons of log-likelihoods

A higher log-likelihood value indicates a more accurate description of the conditional distributions of daily stock returns. We use L_{His} , L_{MF} and L_{ATM} to represent the log-likelihoods of the three models defined after Equation (5). The mean of L_{ATM} , 2144, is slightly higher than the means of L_{His} and L_{MF} . At the same time, the values of L_{ATM} are a little more dispersed across firms than the values of L_{His} and L_{MF} . Overall, there is no evidence showing that one method is systematically better than the others for all of our sample firms.

In Table 5, the frequency counts that show how often each of the three ARCH specifications has the highest log-likelihood for the observed daily returns are presented in the left block. More than a third of the firms (35.6% or 53 firms) have a log-likelihood for the GJR(1,1)–MA(1) model, L_{His} , that is higher than both L_{MF} and

L_{ATM} , which are obtained from the options' information. For the 64.4% (96 firms) whose log-likelihoods are maximized using option specifications, the ATM specification (36.9%) is the best more often than the model-free volatility expectation (27.5%). This is an evidence for the superior efficiency of ATM option implied volatilities when estimating individual stock volatility.

The high frequency for the historical specification is contrary to the studies on options written on stock indices, which reach a consensus that option prices perform much better than ARCH models estimated from daily returns. However, our results are consistent with the in-sample conclusions of Lamoureux and Lastrapes (1993), who show that the GARCH(1,1) model has a slightly higher log-likelihood than the model that incorporates ATM implied volatilities, for all the 10 U.S. firms in their sample.

There are two reasons why the GJR model performs the best for so many firms. Firstly, the key difference between our data for individual stock options and the stock index options is that the latter are much more liquid than the former. The illiquidity of individual stock options may cause the inefficiency of volatility expectations. When we select the 30 firms with the highest average option trading volume, the historical volatility performs the best for only 2 firms; model-free volatility expectation for 11 and ATM implied volatility for 17 firms. The interquartile range for γ (δ) is from 0.64 to 1.87 (from 0.81 to 1.03) and 73.3% (76.7%) of the estimates are significantly different from zero at the 5% level. The means across these 30 firms of L_{His} , L_{MF} and L_{ATM} are respectively 2203, 2213 and 2216.

Secondly, our ARCH specifications and the models in Lamoureux and Lastrapes (1993) are estimated with a horizon of one day, while volatility estimates from option prices represent the expected average daily variation until the end of the option's life. The mismatch between the estimation horizon and the option's time to maturity may reduce the relative performance of both model-free volatility expectation and ATM implied volatility when they are compared with the GJR (1,1) model.

5.2 Option life forecasts from OLS regressions

The regression results are for non-overlapping observations, defined so that the estimation horizon is matched with the option's time to maturity. Panel A, B and C of Table 6, respectively, report the results of both univariate and encompassing regressions that explain volatility, variance and the logarithm of volatility, respectively. The realized volatility is defined after Equation (7)⁷. As before, Table 6 shows the mean, median, lower and upper quartile of the point estimates across 149 firms. The two numbers in parentheses for each parameter estimate are the percentages of firms whose estimates are significantly different from zero at the 10% and 5% levels. The last three columns show the summary statistics for the regression explanatory powers, the adjusted explanatory powers and the sum of squared errors.

⁷ We have also estimated the regressions when realized volatility is measured by the sum of daily squared returns, such that: $\sigma_{RE,t,T} = \sqrt{\frac{252}{H} \sum_{i=1}^H r_{t+i}^2}$, where r_t is the return for day t , calculated using closing stock prices, and H is the number of days between t and the option expiry day. The results are similar to the results from Equation (6) but with lower average explanatory powers.

We begin our discussion with the results of the univariate regressions in Panel A of Table 6. The null hypotheses $\beta_{His} = 0$, $\beta_{MF} = 0$ and $\beta_{ATM} = 0$ are separately rejected for 71.1%, 89.9% and 92.6% of all firms at the 10% level. The values of R^2 are highest for the ATM implied volatility (mean 0.281), but the values for the model-free volatility expectation are similar (mean 0.272); the values for historical volatility, however, are much lower (mean 0.166). This evidence suggests that volatility estimates extracted from option prices are much more informative than historical daily stock returns when the estimation horizons match the lives of the options.

We next consider the encompassing regressions with two independent variables in Panel A of Table 6. The bivariate regression models that include the historical volatility variable increase the mean adjusted R^2 values slightly from the univariate levels for option specifications; from 0.265 to 0.281 for the ATM implied volatility and from 0.256 to 0.273 for the model-free volatility expectation. For these bivariate regressions, only a small number of firms reject the null hypothesis $\beta_{His} = 0$ at the 5% level (34 for σ_{His} and σ_{MF} , and 35 for σ_{His} and σ_{ATM}). Therefore for most firms it can not be rejected that the historical volatility of the underlying asset is redundant when forecasting future volatility, which may be a consequence of the informative option prices and/or the small number of forecasts that are evaluated.

The bivariate regressions involving the model-free volatility expectation and the ATM implied volatility have a mean value of adjusted R^2 equal to 0.274, which is fractionally less than that for the bivariate regressions involving the historical and the ATM volatilities. This can be explained by the very high correlation between the

model-free volatility expectation and the ATM implied volatility. For most firms, both the null hypotheses $\beta_{MF} = 0$ and $\beta_{ATM} = 0$ can not be rejected, showing that we can not decide that one option measure subsumes all the information contained in the other.

The regression involving all three volatility estimates has a mean adjusted R^2 equal to 0.286. The mean values of β_{His} , β_{MF} and β_{ATM} are respectively 0.058, 0.201 and 0.408, suggesting that the ATM forecasts are the most informative.

The volatility variables in the regressions are replaced by variance and by logarithms of volatilities in Panel B and Panel C of Table 6. The comparisons between different specifications provide the same conclusions as those deduced from volatility regressions in Panel A. The variance regressions have lower values of R^2 and, therefore, fewer firms with significant coefficient estimates. However, the regressions that explain the logarithm of volatility generate higher values of R^2 and more firms with significant coefficient estimates. The differences are expected from the summary statistics in Table 2, which shows that variance estimates have higher variations than volatility estimates. Regressions of logarithms of volatilities are less affected by extreme observations.

Comparisons of the explanatory powers of univariate regressions

We focus on the results of the volatility regressions in Panel A of Table 4.6, as the variance regressions and the logarithmic volatility regressions produce similar

conclusions. Firstly, from univariate regressions, we find that the ATM implied volatilities on average result in the highest R^2 values and the historical volatilities the lowest. Secondly, in the bivariate regressions, the statistics of the adjusted R^2 values are always higher than those in univariate regression. For 75 firms, the adjusted R^2 of the bivariate regression involving both σ_{MF} and σ_{His} is higher than that of the univariate regression with only σ_{MF} . Similarly, for 71 firms, the adjusted R^2 of the regression with both σ_{ATM} and σ_{His} is higher than that with only σ_{ATM} . Thirdly, the highest mean adjusted R^2 occurs for the encompassing regression including all three volatility estimates. Therefore, we conclude that on average every volatility estimate contains some additional information beyond that provided by the other estimates. None of historical volatility, model-free volatility expectation or ATM implied volatility can subsume all the information in the others.

The right block of Table 5 provides the frequency counts that show how often each of the three univariate forecasts has the highest value of R^2 in volatility regressions. There are important differences between the frequencies for one-day-ahead forecasts and in option life forecasts shown in Table 5. Only 15.4% (i.e. 23) of the firms have historical volatility ranking highest in the option life forecasts, compared with 35.6% in the left block. Both the model-free volatility expectation and the ATM implied volatility rank the highest more often in the right block than in the left block. The ATM implied volatility has the best regression results for 47.7% (i.e. 71) of the sample firms, while the model-free volatility expectation performs the best for 36.9% (i.e. 55 firms). Thus, only when the estimation horizon is matched do we find that the option prices are clearly more informative than the historical daily returns.

6 Cross-sectional comparisons

For both one-day-ahead forecasts and option life forecasts, we can not find any one volatility estimate, among historical volatility, model-free volatility expectation and ATM implied volatility, that is consistently more informative than the others for most of our sample firms. The liquidity of option trading can partially explain why some firms appear to have more informative option prices (relative to historical forecasts) while the others do not. However, it is hard to discover why ATM implied volatility outperforms the model-free volatility expectation for nearly half of the sample firms; this is true even for the firms having the more actively traded options.

Theoretically, the model-free volatility expectation is superior to Black-Scholes ATM implied volatility, as it is defined to contain the information from a complete set of option prices and is model-independent. In this section, we select some firm specific variables and try to discover the properties of the firms for which one prediction method is better than the others.

To explain the different performance of volatility estimates from option prices relative to historical forecasts, as stated before, the firm's option trading liquidity is one key factor. In the IvyDB Dataset, we are provided with the option trading volume, which is the number of option contracts traded in the market, but without the exact transaction prices. We choose to use this volume data as a measure of option liquidity although it might be the results of high or low option transaction prices and thus is an

approximate measure. In addition, we also calculate the firm's average stock trading volume and the firm's average market capitalization, because firms with more liquid stock trading and with larger size tend to have more liquid option trading. Therefore, we conjecture that firms with option prices more informative than historical volatilities have higher average values of option trading volume, stock trading volume and market capitalization than the firms with historical volatilities more informative than option prices.

We consider two types of explanations for the relative performance of model-free volatility expectation and ATM implied volatility. Firstly, we define appropriate variables that summarize the available market option prices, which might influence the implementation of the model-free volatility expectation. According to the simulation tests in Jiang and Tian (2005, 2007), the estimation error of the model-free volatility expectation is higher when the range of available strike prices is small and/or when the distance between each two strike prices is large. We estimate both the range of available strike prices and the average interval between each two consecutive strike prices⁸ using the moneyness scale to measure these intervals. We also consider the average number of market available strike prices because, with more strike prices, we may extract more information and thus the constructed implied volatility curve is more reliable.

⁸ According to the contract specifications of option contracts written on individual stocks by the CBOE, the strike price intervals are generally 2.5 when the strike price is between \$5 and \$25, 5 when the strike price is between \$25 and \$200, and 10 when the strike price is over \$200. The discretization errors in the numerical integrations in our estimation of model-free volatility expectation are reduced because we use a large number of interpolated option prices with tiny intervals between each two traded. The average interval between market available strike prices included here actually assesses the reliability of the estimation of the implied volatility curve.

Secondly, the relative option liquidity of ATM or near-the-money options to OTM options may be relevant. The information efficiency of model-free volatility expectation relies on the efficiency of OTM option prices while ATM implied volatility represents only the ATM options. If most option trading is concentrated on ATM options, the measure of model-free volatility expectation might be influenced because the option prices used to derive it might not reflect the true market expectations of rational investors. Therefore we estimate the ratio between the trading volume of ATM options and all option trading volume and conjecture that the firms with ATM implied volatility outperforming model-free volatility expectation tend to have a higher average value of the ratio.

With a similar motivation, we also include the ratio between the trading volume of intermediate delta options over all option trading volume, where intermediate delta options are defined by us as the options with delta within the interquartile range of Black-Scholes delta values, which is $[0.25 e^{-rT}, 0.75 e^{-rT}]$. When the relative trading activity of these options is high, there is relatively less or no trading in the strike prices that are far away from ATM. Therefore, the model-free volatility expectation, which includes all strike prices, might then lose its advantage and thus the ATM implied volatility concentrated on only one central strike price might tend to perform better.

The eight selected firm specific variables and how we estimate them are summarized in Table 7. The means and standard deviations of each variable across 149 firms are presented in Panel A of Table 8. In Panel B of Table 8, we consider the 16 firms for which the historical volatility appear to be more informative than option prices under

both one-day-ahead forecasts and option life forecasts, in the group called “His>OP”. The symbol “OP” refers to both model-free volatility expectation and ATM implied volatility, and the group of “His>OP” includes the firms satisfying both “His>MF” and “His>ATM”. Based on similar criteria, there are 73 firms satisfying “OP>His”, which is equivalent to the criteria of “MF>His” or “ATM>His” in both one-day-ahead forecasts and option life forecasts.

First of all, as we expected, the means of variables representing or proxying option liquidity, including option trading volume, stock trading volume and firm size, in the group “His>OP” are all lower than those in the group “OP>His”. We perform the Student’s t test with the null hypothesis that the mean in group “His>OP” is equal to or higher than the mean in group “OP>His” and find that the null hypothesis is strongly rejected for each of these three variables. The t -test statistics are -5.08 (option volume), -4.03 (stock volume) and -2.75 (firm size).

Secondly, the same null hypothesis is also decisively rejected for the average number of strike prices, with a t -test value equal to -5.03 . When there are more market available strike prices, the estimation of model-free volatility expectation is likely to be improved. Moreover, the CBOE issues new strike prices and increases the number of strike prices when the underlying asset price varies beyond the existing strike prices range. So when the number of strike prices increases, options may better and more quickly capture the variation in stock prices movements, than historical forecasts that are based upon only the autocorrelation property of daily volatilities.

In Panel C of Table 8, we compare the firms satisfying “MF>ATM” and the firms satisfying “ATM>MF” for both one-day-ahead forecasts and option life forecasts. The selections of firms consider only the comparisons between model-free volatility expectation and ATM implied volatility, by ignoring the ranking of historical volatility. There are 37 firms in the group “MF>ATM” and 60 in the group “ATM>MF”, showing again that the ATM implied volatility performs better than the model-free volatility expectation for more firms under both forecast horizons.

In the one-sided hypothesis tests, the student t statistics are not highly significant for most of the variables. The p – values of average strike prices interval and firm size are slightly lower than the 5% significance level. The firms in the group “ATM>MF” have a higher average firm size than that of the firms in the group “MF>ATM”. Opposite to what we expected, the average strike price interval for the firms with the model-free volatility expectation performing better is slightly higher. This might reflect a wider strike price range for these firms. It is possible when the distribution of market available strike prices is more dispersed, that the theoretical advantages of the model-free volatility expectation will be more likely to be detected.

We can not find evidence that the relative trading volume of ATM options or the ratio of intermediate delta options to all options can explain the comparisons between ATM implied volatility and model-free volatility expectation. A high relative trading for ATM options or intermediate delta options might occur when there is illiquid trading for all options. In order to discover for which kind of firms model-free volatility expectation is more informative than ATM implied volatility, for each firm, we also estimate some other variables, including realized volatility over the sample period,

average risk-neutral skewness⁹ and Beta¹⁰. However, the average level of any of these variables for firms with model-free volatility expectations more informative than ATM implied volatilities is not significantly different from the average level for the firms that have ATM implied volatilities performing better.

From the cross-sectional analysis, it is obvious that the relative performance of option prices and historical forecasts when forecasting volatility for individual stocks is highly determined by the liquidity of option trading. As for the comparisons between model-free volatility expectation and ATM implied volatility for individual stocks, it is hard to identify the key effective factors. There are two possible explanations of the relatively poor performance of model-free volatility expectation, which can not be verified using our available data. Firstly, the individual stock options data might contain measurement errors from nonsynchronous trading of option and stocks, bid-ask spread etc, with different levels for different strike prices. The model-free volatility expectation calculated as a combination of option prices across all strikes, contains more noise in total than ATM implied volatility that is from the ATM options alone. This may also explains why the explanatory powers in our regressions for option life forecasts are lower than those in some existing literature on stock index options that contain less measurement errors.

Another explanation may be that the trading of individual stock options, especially the OTM options, is not sufficient to reflect the advantages of the model-free volatility expectation over ATM implied volatility. Even for the firms with the highest option

⁹ The measures of risk-neutral skewness are estimated according to Bakshi, Kapadia and Madan (2003). The time-series average of daily measures is used for each firm.

¹⁰ Beta for each firm is calculated as the coefficient estimate when regressing the firm's daily stock returns, from January 1996 to December 1999, on the daily returns of the S&P 100 index.

trading volume, the liquidity of their options are still far less than that of the stock indices or exchange rate options. Illiquidity might also be the reason why we can not find significant results when using the firm specific variables, such as relative option trading volume, to explain the performance of the model-free volatility expectation.

7 Conclusion

The main contribution of this chapter is that, for the first time in the existing literature, we compare the information content of three volatility estimates, namely the historical forecasts, the newly-developed model-free volatility expectation and the ATM implied volatility, for each of a large number of U.S. individual stocks.

For the majority of our sample firms, the three volatility estimates are all shown to contain relevant, but not all, information about the variation of underlying asset returns. There is a consensus from previous studies about the informational efficiency of options written on stock indices and foreign exchange rates that option prices are more informative than historical daily stock returns when estimating and predicting the volatility of underlying assets. Our analysis of 149 firms shows that, however, a different estimation conclusion applies to options for individual firms. For one-day-ahead estimation, more than a third of our firms do not have volatility estimates, extracted from option prices, that are more accurate than those provided by a simple ARCH model estimated from daily stock returns. When the prediction horizon extends until the expiry date of the options, the historical volatility becomes less informative than either the ATM implied volatility or the newly-developed model-free

volatility expectation for 126 out of 149 firms. Our results also show that both volatility estimates from option prices are more likely to outperform historical returns when the firm has higher average option trading volume.

The interest shown in the model-free volatility expectation is largely due to its property of model independence. We use a quadratic function of delta to estimate the implied volatility curve and then interpolate a large number of option prices to approximate the integral function used to estimate the model-free volatility expectation. Although the model-free volatility expectation has been demonstrated firmly to be the most accurate predictor of realized volatility by Jiang and Tian (2005) for the S&P 500 index, it only outperforms both the ATM implied volatility and the historical volatility for about one-third of our sample firms. In contrast, the ATM implied volatility is the method that most often performs the best. It is also interesting to find that for those firms with ATM implied volatility outperforming model-free volatility, the relative trading volume of ATM options are not significantly higher than that of the firms with model-free volatility expectation performing better.

The relatively poor performance of the model-free volatility expectation might be explained by the overall illiquidity of the individual stock option market. That is why we can not draw significant conclusions when using relative option trading volume to explain for which firms the model-free volatility expectation tends to outperform the ATM implied volatility. Another possible explanation is that the OTM option prices contain measurement error or are mispriced and, therefore, the model-free volatility expectation, which is a combination of option prices across all strikes, is outperformed by the information provided by ATM options alone. A third explanation

may be that option prices for all strikes can not be inferred reliably from a handful of traded strikes. So the theoretical predomination of model-free volatility expectation over ATM implied volatility is not exhibited for our individual stock options.

References

- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2003, Modeling and Forecasting Realized Volatility, *Econometrica* 71, 579-625.
- Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options, *The Review of Financial Studies* 16, 101-143.
- Black, Fischer, and Myron Scholes, 1973, The Pricing of Options and Corporate Liabilities, *The Journal of Political Economy* 81, 637-654.
- Blair, Bevan J., Ser-Huang Poon, and Stephen J. Taylor, 2001a, Modelling S&P 100 volatility: The information content of stock returns, *Journal of Banking & Finance* 25, 1665-1679.
- Blair, Bevan J., Ser-Huang Poon, and Stephen J. Taylor, 2001b, Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns, *Journal of Econometrics* 105, 5-26.
- Bliss, Robert R., and Nikolaos Panigirtzoglou, 2002, Testing the stability of implied probability density functions, *Journal of Banking & Finance* 26, 381-422.
- Bliss, Robert R., and Nikolaos Panigirtzoglou, 2004, Option-Implied Risk Aversion Estimates., *Journal of Finance* 59, 407-446.
- Bollerslev, Tim, and J.M. Woodlridge, 1992, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews* 11, 143-179.
- Breeden, Douglas T., and Robert H. Litzenberger, 1978, Prices of State-Contingent Claims Implicit in Option Prices, *The Journal of Business* 51, 621-651.

- Britten-Jones, Mark, and Anthony Neuberger, 2000, Option Prices, Implied Price Processes, and Stochastic Volatility, *The Journal of Finance* 55, 839-866.
- Canina, Linda, and Stephen Figlewski, 1993, The Informational Content of Implied Volatility, *The Review of Financial Studies* 6, 659-681.
- Carr, Peter, and Dilip Madan, 1998, Towards a Theory of Volatility Trading, in R.A.Jarrow, ed.: *Volatility* (Risk Books, London).
- Carr, Peter, and Liuren Wu, 2006, A Tale of Two Indices., *Journal of Derivatives* 13, 13-29.
- Carr, Peter., and Liuren Wu, 2004, Variance Risk Premia, *Forthcoming The Review of Financial Studies*.
- Cheung, Yin-Wong, and Lilian K. Ng, 1992, Stock Price Dynamics and Firm Size: An Empirical Investigation, *The Journal of Finance* 47, 1985-1997.
- Christensen, B. J., and N. R. Prabhala, 1998, The relation between implied and realized volatility, *Journal of Financial Economics* 50, 125-150.
- Cox, John C., Stephen A. Ross, and Mark Rubinstein, 1979, Option pricing: A simplified approach, *Journal of Financial Economics* 7, 229-263.
- Day, Theodore E., and Craig M. Lewis, 1992, Stock market volatility and the information content of stock index options, *Journal of Econometrics* 52, 267-287.
- Demeterfi, Kresimir, Emanuel Derman, Michael Kamal, and Joseph Zou, 1999a, More Than You Ever Wanted to know About Volatility Swaps, *Quantitative Strategies Research Notes, Goldman Sachs*.
- Demeterfi, Kresimir, Emanuel Derman, Michael Kamal, and Joseph Zou, 1999b, A Guide to Volatility and Variance Swaps., *Journal of Derivatives* 6, 9-32.

- Duffee, Gregory R., 1995, Stock returns and volatility a firm-level analysis, *Journal of Financial Economics* 37, 399-420.
- Ederington, Louis H, and Wei Guan, 2002a, Measuring implied volatility: Is an average better? Which average?, *The Journal of Futures Markets* 22, 811.
- Fleming, Jeff, 1998, The quality of market volatility forecasts implied by S&P 100 index option prices, *Journal of Empirical Finance* 5, 317-345.
- Garman, Mark B., and Michael J. Klass, 1980, On the Estimation of Security Price Volatilities from Historical Data, *The Journal of Business* 53, 67-78.
- Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle, 1993, On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *The Journal of Finance* 48, 1779-1801.
- Guo, Dajiang, 1996, The Predictive Power of Implied Stochastic Variance from Currency Options, *The Journal of Futures Markets (1986-1998)* 16, 915.
- Jiang, George J., and Yisong S. Tian, 2005, The Model-Free Implied Volatility and Its Information Content, *The Review of Financial Studies* 18, 1305.
- Jiang, George J., and Yisong S. Tian, 2007, Extracting Model-Free Volatility from Option Prices: An Examination of the VIX Index., *Journal of Derivatives* 14, 35-60.
- Jorion, Philippe, 1995, Predicting Volatility in the Foreign Exchange Market, *The Journal of Finance* 50, 507-528.
- Kroner, Kenneth F, Kevin P Kneafsey, and Stijn Claessens, 1995, Forecasting volatility in commodity markets, *Journal of Forecasting* 14, 77.
- Lamoureux, Christopher G., and William D. Lastrapes, 1993, Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities, *The Review of Financial Studies* 6, 293-326.

- Li, Kai, 2002, Long-Memory versus Option-Implied Volatility Predictions., *Journal of Derivatives* 9, 9-25.
- Lynch, Damien P., and Nikolaos Panigirtzoglou, 2004, Option Implied and Realized Measures of Variance, Working Paper.
- Malz, Allan M, 1997a, Estimating the probability distribution of the future exchange rate from option prices, *Journal of Derivatives* 5, 18.
- Malz, Allan M., 1997b, Option-Implied Probability Distributions and Currency Excess Returns, Working Paper.
- Parkinson, Michael, 1980, The Extreme Value Method for Estimating the Variance of the Rate of Return, *The Journal of Business* 53, 61-65.
- Pong, Shiuyan, Mark B. Shackleton, Stephen J. Taylor, and Xinzhong Xu, 2004, Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models, *Journal of Banking & Finance* 28, 2541-2563.
- Poon, Ser-Huang, and Clive W. J. Granger, 2003, Forecasting Volatility in Financial Markets: A Review, *Journal of Economic Literature* 41, 478-539.
- Rubinstein, Mark, 1994, Implied Binomial Trees., *Journal of Finance* 49, 771-818.
- Shimko, D., 1994, Bounds of Probabilty, *Risk* 6, 33-37.
- Taylor, Stephen J., and Gary X. Xu, 1994, The magnitude of implied volaitlty similes: theory and empirical evidence for exchange rates, *Review of Futures Markets* 13, 355-380.
- Taylor, Stephen J., and Xinzhong Xu, 1997, The incremental volatility information in one million foreign exchange quotations, *Journal of Empirical Finance* 4, 317-340.
- White, Halbert, 1980, A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica* 48, 817-838.

Xu, Xinzhong, and Stephen J. Taylor, 1995, Conditional volatility and the informational efficiency of the PHLX currency options market, *Journal of Banking & Finance* 19, 803-821.

Figure 1 Distribution of the average number of daily available strike prices for 149 firms

This figure shows the frequency counts of firms having different number of option observations during the sample period from January 4, 1996 to December 31, 1999. The x-axis is the average number of daily available strike prices for firm i , \bar{N}_i , defined as the total number of option observations for firm i during the sample period divided by the number of trading days. The y-axis is the number of firms.

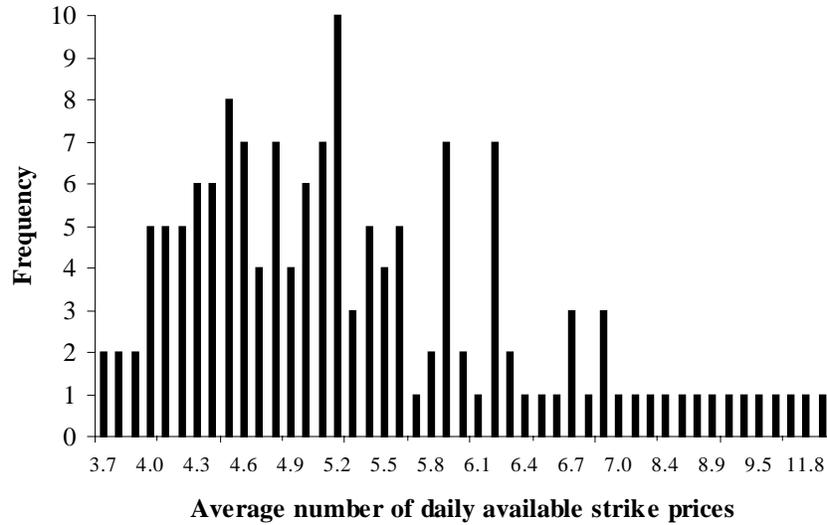
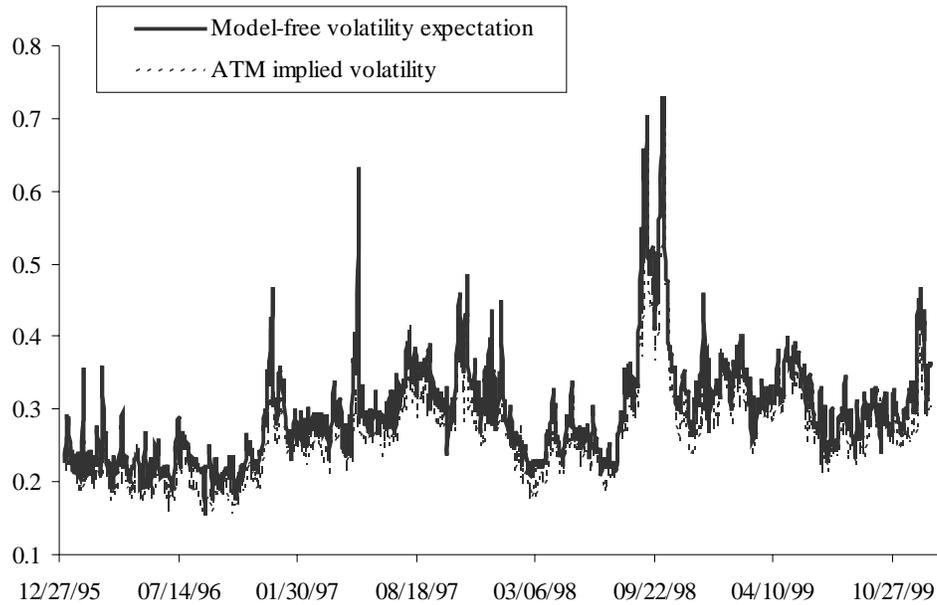


Figure 2 Examples of the model-free volatility expectation and the ATM implied volatility

The figure plots the time series of daily estimates of model-free volatility expectation and ATM implied volatility over the sample period from January 1996 to December 1999.

Panel A: Time series plot for General Electric



Panel B: Time series plot for IBM

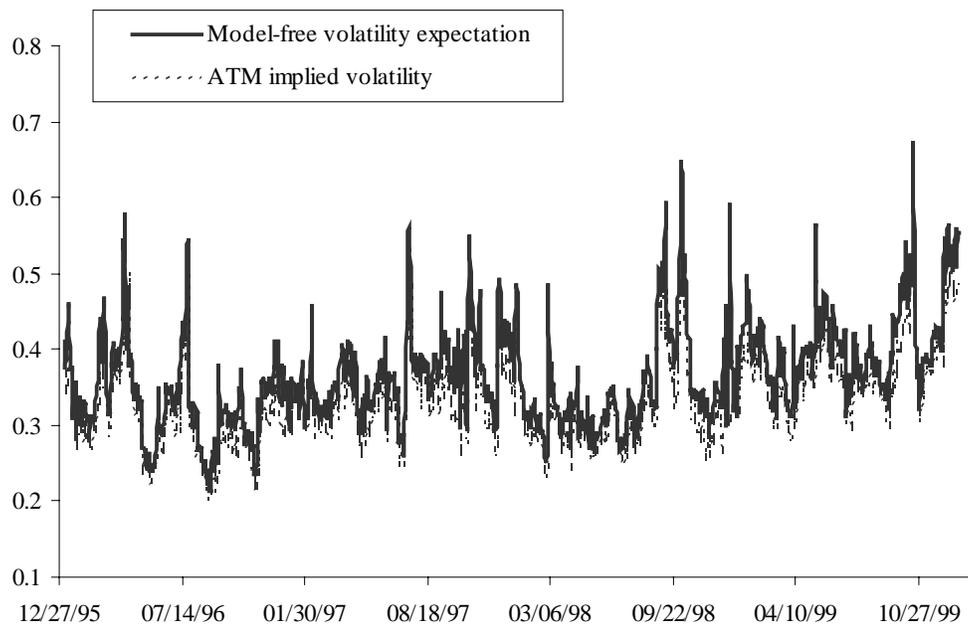


Figure 3 Comparison of the estimated values of $\frac{\gamma}{1-\beta_\gamma}$ and $\frac{\delta}{1-\beta_\delta}$ for 149 firms

For each firm, $\frac{\gamma}{1-\beta_\gamma}$ and $\frac{\delta}{1-\beta_\delta}$ are, respectively, obtained from the estimates of the ARCH specification using information provided by the model-free volatility expectation only and the estimates of the model using information provided by the ATM implied volatility only. The straight line is the 45-degree line.

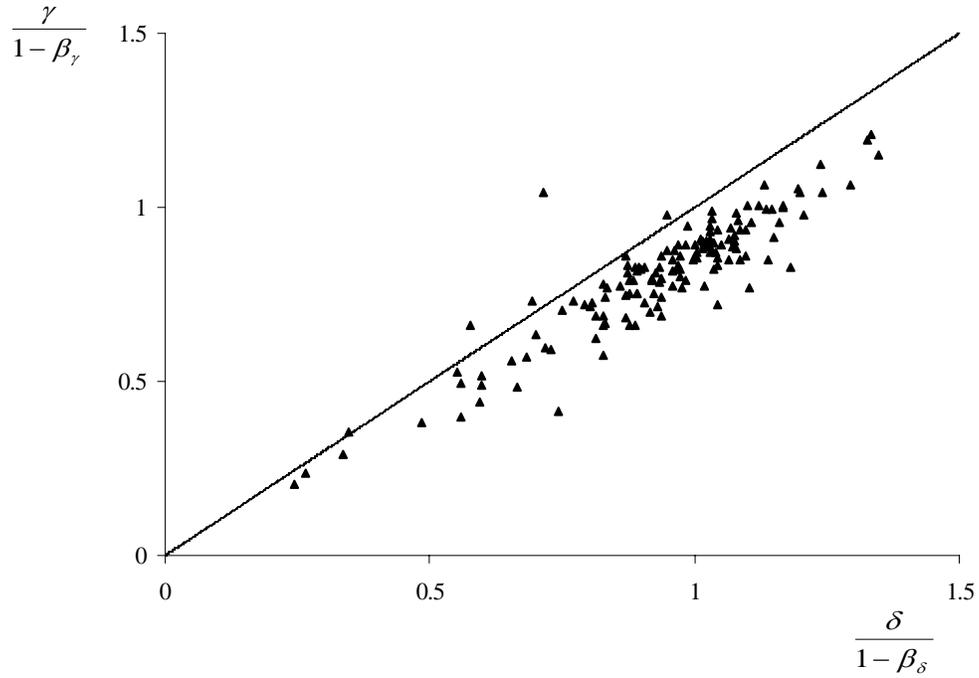


Table 1 Summary statistics for daily volatility estimates

The table contains the summary statistics for daily volatility estimates obtained from options, from January 1996 to December 1999. Numbers are cross-sectional statistics calculated from time series statistics, for a cross-section of 149 firms. The mean, L_q , Median and U_q are, respectively, the mean, lower quartile, median and upper quartile values of each statistic across 149 firms. σ_{MF} and σ_{ATM} are the daily estimates of model-free volatility expectation and ATM implied volatility. The last panel shows the cross-sectional statistics of the firms' correlation coefficients between σ_{MF} and σ_{ATM} .

	Mean	L_q	Median	U_q
<i>Panel A: σ_{MF}</i>				
Mean	0.523	0.371	0.521	0.646
Std. Dev.	0.123	0.078	0.106	0.131
Max	1.538	0.826	1.176	1.527
Min	0.316	0.222	0.311	0.403
<i>Panel B: σ_{ATM}</i>				
Mean	0.486	0.351	0.484	0.610
Std. Dev.	0.099	0.072	0.094	0.114
Max	1.093	0.755	1.023	1.278
Min	0.296	0.205	0.285	0.381
<i>Panel C: $\sigma_{MF} - \sigma_{ATM}$</i>				
Mean	0.036	0.024	0.032	0.043
Std. Dev.	0.051	0.026	0.035	0.048
Max	0.634	0.218	0.301	0.415
Min	-0.105	-0.132	-0.067	-0.042
<i>Panel D: Correlation coefficient between σ_{MF} and σ_{ATM}</i>				
Correlation	0.926	0.907	0.940	0.960

Table 2 Summary statistics for non-overlapping volatility estimates

The table contains the summary statistics for non-overlapping volatility estimates for regression analysis from January 1996 to December 1999. Numbers are cross-sectional statistics calculated from time series statistics, for a cross section of 149 firms. The L_q , Med and U_q are, respectively, the lower quartile, median and upper quartile values of each statistic across 149 firms. MF, ATM, His and RE represent respectively the annualized values of model-free volatility expectation, at-the-money implied volatility, historical forecast from GARCH model and realized volatility calculated from Parkinson (1980)'s formula.

	MF			ATM			HIS			RE		
	L_q	Med	U_q	L_q	Med	U_q	L_q	Med	U_q	L_q	Med	U_q
<i>Panel A: volatility</i>												
Mean	0.37	0.51	0.64	0.35	0.49	0.61	0.32	0.50	0.60	0.29	0.42	0.54
Std.Dev.	0.07	0.09	0.11	0.06	0.09	0.11	0.05	0.08	0.11	0.09	0.11	0.15
Skew	0.43	0.78	1.12	0.38	0.70	1.10	0.69	1.28	2.05	0.60	1.03	1.45
Exc. Kurtosis	-0.37	0.25	1.62	-0.61	0.20	1.52	0.53	2.71	5.89	0.08	1.21	2.86
Max	0.59	0.80	0.98	0.56	0.71	0.90	0.52	0.73	0.94	0.57	0.80	0.99
Min	0.25	0.36	0.45	0.23	0.34	0.43	0.22	0.34	0.46	0.16	0.22	0.31
<i>Panel B: variance</i>												
Mean	0.14	0.27	0.42	0.13	0.24	0.38	0.11	0.26	0.36	0.09	0.19	0.31
Std.Dev.	0.06	0.10	0.17	0.05	0.08	0.14	0.04	0.07	0.13	0.06	0.12	0.19
Skew	0.77	1.18	1.77	0.68	1.09	1.70	1.14	1.93	2.84	1.20	1.75	2.58
Exc. Kurtosis	0.23	1.10	4.53	-0.05	1.12	3.45	1.56	5.14	10.48	1.36	4.01	8.23
Max	0.35	0.64	0.97	0.31	0.51	0.81	0.27	0.54	0.89	0.33	0.65	0.99
Min	0.06	0.13	0.21	0.05	0.11	0.18	0.05	0.11	0.21	0.03	0.05	0.10
<i>Panel C: natural logarithm of (100 × volatility)</i>												
Mean	3.58	3.93	4.14	3.53	3.88	4.09	3.46	3.89	4.09	3.33	3.69	3.96
Std.Dev.	0.15	0.17	0.21	0.15	0.17	0.21	0.11	0.16	0.23	0.23	0.26	0.30
Skew	0.07	0.34	0.66	0.00	0.29	0.61	0.22	0.75	1.49	-0.02	0.26	0.60
Exc. Kurtosis	-0.65	-0.24	0.43	-0.73	-0.28	0.30	-0.07	1.16	3.23	-0.42	-0.04	0.52
Max	4.09	4.38	4.59	4.02	4.27	4.50	3.95	4.29	4.55	4.04	4.39	4.60
Min	3.20	3.59	3.82	3.15	3.51	3.76	3.09	3.51	3.83	2.77	3.08	3.43

Table 3 Summary statistics of the correlation matrices

The table presents the summary statistics of the correlation matrices calculated for the sample period from January 1996 to December 1999. Correlation coefficients for each firm are calculated for the non-overlapping observations of volatility estimates. The mean, L_q , Med and U_q are respectively the cross-sectional mean, lower quartile, median and upper quartile of the correlation coefficients of 149 firms. σ_{MF} , σ_{ATM} , σ_{His} and σ_{RE} are, respectively, the model-free volatility expectation, ATM implied volatility, historical volatility and realized volatility calculated from Parkinson (1980)'s formula. Numbers are stated as percentages.

<i>Panel A: correlation matrices of volatility estimates</i>												
	σ_{MF}				σ_{ATM}				σ_{His}			
	Mean	L_q	Med	U_q	Mean	L_q	Med	U_q	Mean	L_q	Med	U_q
σ_{ATM}	93.7	92.3	95.2	97.1								
σ_{His}	54.7	37.3	57.6	74.5	56.2	41.2	57.7	74.7				
σ_{RE}	49.2	37.3	51.6	62.7	50.2	37.7	51.6	62.1	34.4	19.1	34.3	50.0

<i>Panel B: correlation matrices of variance estimates</i>												
	σ_{MF}^2				σ_{ATM}^2				σ_{His}^2			
	Mean	L_q	Med	U_q	Mean	L_q	Med	U_q	Mean	L_q	Med	U_q
σ_{ATM}^2	93.3	92.3	95.2	97.1								
σ_{His}^2	53.1	37.0	55.4	75.8	54.6	39.9	56.6	74.6				
σ_{RE}^2	47.1	34.1	48.7	58.7	47.8	34.9	47.9	60.3	31.0	15.9	31.2	47.9

<i>Panel C: correlation matrices of natural logarithm of (100 × volatility)</i>												
	$\ln(100\sigma_{MF})$				$\ln(100\sigma_{ATM})$				$\ln(100\sigma_{His})$			
	Mean	L_q	Med	U_q	Mean	L_q	Med	U_q	Mean	L_q	Med	U_q
$\ln(100\sigma_{ATM})$	93.8	92.0	95.1	96.9								
$\ln(100\sigma_{His})$	55.3	38.1	58.9	73.4	56.7	42.4	60.0	74.5				
$\ln(100\sigma_{RE})$	49.9	38.1	51.3	62.4	51.2	39.4	53.3	62.9	36.3	21.6	36.8	51.9

Table 4 Summary statistics of ARCH parameter estimates across 149 firms

Daily stock returns r_t are modelled by the ARCH specification: $r_t = \mu + \psi\varepsilon_{t-1} + \varepsilon_t$, $\varepsilon_t = \sqrt{h_t} z_t$, $z_t \sim i.i.d.(0,1)$, $h_t = \frac{\omega + \alpha\varepsilon_{t-1}^2 + \alpha^- s_{t-1}\varepsilon_{t-1}^2}{1 - \beta L} + \frac{\gamma\sigma_{MF,t-1}^2}{1 - \beta_\gamma L} + \frac{\delta\sigma_{ATM,t-1}^2}{1 - \beta_\delta L}$, s_t is 1 if $\varepsilon_t \leq 0$, otherwise s_t is zero. σ_{MF} and σ_{ATM} are respectively the daily measure of model-free volatility expectation and ATM implied volatility. Parameters are estimated by maximizing the quasi-log-likelihood function. Panel A contains the results for the GJR (1,1)-MA (1) model; Panel B and Panel C for models that use information provided by model-free volatility expectation and ATM implied volatility respectively. Cross-sectional statistics including mean, standard deviation, lower quartile (L_q), median and upper quartile (U_q) across 149 firms for each parameter are presented. Numbers under the column “10%/5%/1%” are the percentages of estimates which are significantly different from zero at the 10%, 5% and 1% levels. Inferences are made through t-ratios constructed from the robust standard errors of Bollerslev and Wooldridge (1992). L_{His} , L_{MF} and L_{ATM} represent the maximized loglikelihood values of each model.

	Mean	Std. Dev.	L_q	Med	U_q	10%/5%/1%
<i>Panel A: estimates of GJR (1,1)-MA (1) model</i>						
$\mu \times 10^3$	0.91	0.96	0.40	0.85	1.42	(24.2/11.4/8.7)
θ	0.00	0.06	-0.04	0.00	0.04	(27.5/16.1/11.4)
$\omega \times 10^5$	17.61	28.93	1.57	5.91	17.89	(74.5/63.8/43.6)
α	0.05	0.07	0.00	0.03	0.06	(32.2/20.1/8.1)
α^-	0.12	0.20	0.04	0.08	0.13	(54.4/43.0/25.5)
β	0.77	0.23	0.66	0.86	0.93	(93.3/93.3/90.6)
Persistence	0.87	0.18	0.81	0.94	0.98	
L_{His}	2141	340	1860	2083	2438	
<i>Panel B: estimates of ARCH specification that uses model-free volatility expectation only</i>						
$\mu \times 10^3$	0.77	0.92	0.33	0.73	1.22	(24.2/12.8/6.0)
θ	0.01	0.05	-0.03	0.00	0.05	(24.8/17.4/9.4)
$\omega \times 10^5$	10.35	25.16	0.00	0.49	10.15	(2.0/0.7/0)
γ	0.65	0.26	0.48	0.71	0.85	(61.1/50.3/39.6)
β_γ	0.19	0.26	0.00	0.03	0.34	(8.7/7.4/6.7)
$\gamma/(1 - \beta_\gamma)$	0.80	0.19	0.72	0.83	0.90	
L_{MF}	2143	339	1872	2075	2454	
<i>Panel C: estimates of ARCH specification that uses ATM implied volatility only</i>						
$\mu \times 10^3$	0.77	0.91	0.35	0.71	1.23	(20.8/11.4/6.7)
θ	0.01	0.05	-0.03	0.01	0.05	(26.2/15.4/7.4)
$\omega \times 10^5$	9.41	22.27	0.00	0.00	7.27	(0.7/0/0)
δ	0.81	0.29	0.62	0.88	1.01	(57.0/42.3/28.2)
β_δ	0.14	0.24	0.00	0.00	0.20	(6.0/4.0/3.4)
$\delta/(1 - \beta_\delta)$	0.92	0.22	0.84	0.96	1.04	
L_{ATM}	2144	340	1870	2078	2453	

Table 5 Frequency counts for the variables that best describe the volatility of stock returns

The figures are the numbers and the percentages (out of 149) of firms that satisfy the orderings in the first column. For one-day-ahead forecasts, frequency counts are based on the maximized log-likelihood values of the ARCH specifications that use historical daily returns, model-free volatility expectation and ATM implied volatility, respectively. The higher the log-likelihood value, the better the volatility estimate is. For option life forecasts, the frequency counts are based on the explanatory powers of the univariate regressions that contain historical volatility, model-free volatility expectation and ATM implied volatility, respectively. The higher the explanatory power, the better the volatility estimate is. His, MF and ATM respectively represent the historical forecast, the model-free volatility expectation and the ATM implied volatility.

	For one-day-ahead forecasts		For options' life forecasts	
His is the best	53	35.57%	23	15.44%
His>MF>ATM	21	14.09%	9	6.04%
His>ATM>MF	32	21.48%	14	9.40%
MF is the best	41	27.52%	55	36.91%
MF>ATM>His	38	25.50%	51	34.23%
MF>His>ATM	3	2.01%	4	2.68%
ATM is the best	55	36.91%	71	47.65%
ATM>MF>His	49	32.89%	60	40.27%
ATM>His>MF	6	4.03%	11	7.38%

Table 6 Summary statistics of estimates across 149 firms for univariate and encompassing regressions

The table contains the results of both univariate and multivariate OLS regressions. Panel A presents the results of the regressions for volatility, with the general specification such that:

$$\sigma_{RE,i,T} = \beta_0 + \beta_{His} \sigma_{His,i,T} + \beta_{MF} \sigma_{MF,i,T} + \beta_{ATM} \sigma_{ATM,i,T} + \varepsilon_{i,T};$$

Panel B for variance with the specification:

$$\sigma_{RE,i,T}^2 = \beta_0 + \beta_{His} \sigma_{His,i,T}^2 + \beta_{MF} \sigma_{MF,i,T}^2 + \beta_{ATM} \sigma_{ATM,i,T}^2 + \varepsilon_{i,T};$$

and Panel C for the logarithm of volatility with the specification:

$$\ln(100\sigma_{RE,i,T}) = \beta_0 + \beta_{His} \ln(100\sigma_{His,i,T}) + \beta_{MF} \ln(100\sigma_{MF,i,T}) + \beta_{ATM} \ln(100\sigma_{ATM,i,T}) + \varepsilon_{i,T},$$

where σ_{RE} , σ_{MF} , σ_{ATM} and σ_{His} respectively refer to the realized volatility calculated from Parkinson (1980)'s formula, model-free volatility expectation, at-the-money option implied volatility and historical forecast from GARCH model. The regressions are estimated for each out of 149 firms and then the cross-sectional summary statistics, including mean, lower quartile (L_q), median (Med), upper quartile (U_q), for each coefficient estimate are presented.

Numbers in parentheses are the percentages of firms whose the estimates are significantly different from zero at the 10% and the 5% levels. Inferences are made through the standard errors computed following a robust procedure taking account of heteroscedasticity [White (1980)]. SSE is the sum of squared error of the regression.

Table 6 Summary statistics of estimates across 149 firms for univariate and encompassing regressions

		β_0	β_{His}	β_{MF}	β_{ATM}	R^2	$adj.R^2$	SSE
<i>Panel A: results of volatility regressions</i>								
His	Mean	0.203	0.505			0.166	0.148	0.669
	L_q	0.046	0.271			0.040	0.020	0.252
	Med	0.158	0.504			0.117	0.099	0.542
	U_q	0.307	0.761			0.250	0.234	0.941
		(57.7/49.0)	(71.1/61.7)					
MF	Mean	0.120		0.601		0.272	0.256	0.573
	L_q	0.037		0.465		0.139	0.121	0.212
	Med	0.096		0.617		0.266	0.250	0.461
	U_q	0.179		0.750		0.393	0.380	0.835
		(40.9/29.5)		(89.9/86.6)				
ATM	Mean	0.101			0.680	0.281	0.265	0.572
	L_q	0.008			0.503	0.142	0.124	0.202
	Med	0.089			0.693	0.266	0.250	0.447
	U_q	0.179			0.855	0.386	0.373	0.825
		(36.9/26.8)			(92.6/90.6)			
His+MF	Mean	0.108	0.113	0.526		0.303	0.273	0.550
	L_q	0.006	-0.065	0.317		0.163	0.127	0.198
	Med	0.071	0.098	0.552		0.281	0.250	0.434
	U_q	0.162	0.330	0.711		0.427	0.402	0.777
		(25.5/17.4)	(30.9/22.8)	(75.2/64.4)				
His+ATM	Mean	0.104	0.065		0.621	0.311	0.281	0.548
	L_q	-0.006	-0.096		0.430	0.176	0.140	0.192
	Med	0.070	0.081		0.599	0.299	0.268	0.436
	U_q	0.144	0.324		0.830	0.417	0.392	0.770
		(24.2/14.8)	(28.9/23.5)		(79.9/70.5)			
MF+ATM	Mean	0.102		0.239	0.425	0.304	0.274	0.552
	L_q	0.009		-0.163	-0.003	0.166	0.129	0.198
	Med	0.090		0.149	0.528	0.290	0.259	0.430
	U_q	0.180		0.618	0.926	0.415	0.389	0.789
		(37.6/26.8)		(20.1/16.8)	(28.9/22.8)			
His+MF+ATM	Mean	0.107	0.058	0.201	0.408	0.331	0.286	0.531
	L_q	0.003	-0.104	-0.203	-0.077	0.195	0.141	0.186
	Med	0.070	0.077	0.151	0.479	0.316	0.270	0.412
	U_q	0.157	0.297	0.536	0.820	0.454	0.418	0.764
		(22.8/13.4)	(26.8/20.8)	(9.5/12.1)	(25.5/17.4)			

Table 6 Summary statistics of estimates across 149 firms for univariate and encompassing regressions (Cont.)

		β_0	β_{His}	β_{MF}	β_{ATM}	R^2	$adj.R^2$	SSE
<i>Panel B: results of variance regressions</i>								
His	Mean	0.126	0.422			0.142	0.124	0.022
	L_q	0.021	0.181			0.028	0.007	0.003
	Med	0.070	0.404			0.097	0.078	0.012
	U_q	0.193	0.658			0.230	0.213	0.030
		(61.7/53.7)	(63.8/54.4)					
MF	Mean	0.064		0.518		0.253	0.237	0.018
	L_q	0.010		0.359		0.116	0.097	0.002
	Med	0.045		0.518		0.237	0.221	0.010
	U_q	0.102		0.653		0.345	0.331	0.025
		(45.6/33.6)		(86.6/79.2)				
ATM	Mean	0.055			0.629	0.259	0.243	0.018
	L_q	0.000			0.422	0.122	0.103	0.002
	Med	0.034			0.614	0.230	0.213	0.010
	U_q	0.089			0.780	0.364	0.351	0.025
		(39.6/33.6)			(89.9/83.2)			
His+MF	Mean	0.068	0.051	0.475		0.286	0.255	0.018
	L_q	0.006	-0.121	0.266		0.152	0.115	0.002
	Med	0.035	0.059	0.472		0.272	0.240	0.010
	U_q	0.100	0.288	0.624		0.412	0.386	0.025
		(29.5/20.1)	(30.9/22.1)	(71.1/56.4)				
His+ATM	Mean	0.066	0.003		0.601	0.291	0.260	0.018
	L_q	-0.001	-0.142		0.342	0.149	0.112	0.002
	Med	0.029	0.041		0.553	0.279	0.248	0.010
	U_q	0.092	0.256		0.784	0.415	0.390	0.024
		(27.5/20.8)	(28.9/20.1)		(75.2/64.4)			
MF+ATM	Mean	0.055		0.226	0.375	0.288	0.257	0.018
	L_q	0.001		-0.135	-0.121	0.148	0.110	0.002
	Med	0.036		0.233	0.338	0.263	0.231	0.010
	U_q	0.098		0.553	0.824	0.415	0.390	0.024
		(43.6/32.9)		(22.8/16.1)	(29.5/21.5)			
His+MF+ATM	Mean	0.067	-0.001	0.199	0.378	0.317	0.271	0.017
	L_q	0.003	-0.174	-0.173	-0.049	0.176	0.122	0.002
	Med	0.036	0.039	0.232	0.387	0.302	0.255	0.009
	U_q	0.157	0.297	0.536	0.820	0.441	0.404	0.023
		(22.8/13.4)	(26.8/20.8)	(9.5/12.1)	(25.5/17.4)			

Table 6 Summary statistics of estimates across 149 firms for univariate and encompassing regressions (Cont.)

		β_0	β_{His}	β_{MF}	β_{ATM}	R^2	Adj. R^2	SSE
<i>Panel C: the results of logarithm volatility regressions</i>								
His	Mean	1.344	0.620			0.179	0.162	0.059
	L_q	0.250	0.378			0.050	0.030	0.044
	Med	1.288	0.618			0.136	0.117	0.053
	U_q	2.312	0.889			0.270	0.254	0.067
		(48.3/44.3)	(75.2/69.8)					
MF	Mean	0.834		0.731		0.278	0.262	0.052
	L_q	0.170		0.594		0.145	0.127	0.038
	Med	0.764		0.749		0.263	0.247	0.047
	U_q	1.374		0.887		0.390	0.377	0.064
		(33.6/24.8)		(91.9/88.6)				
ATM	Mean	0.713			0.775	0.290	0.275	0.051
	L_q	0.016			0.601	0.155	0.137	0.037
	Med	0.673			0.777	0.284	0.268	0.047
	U_q	1.310			0.960	0.396	0.383	0.061
		(31.5/22.8)			(92.6/90.6)			
His+MF	Mean	0.607	0.182	0.613		0.308	0.278	0.049
	L_q	-0.179	-0.016	0.372		0.175	0.140	0.036
	Med	0.501	0.187	0.645		0.290	0.259	0.046
	U_q	1.145	0.418	0.812		0.424	0.399	0.062
		(18.8/12.1)	(30.9/24.8)	(75.8/67.1)				
His+ATM	Mean	0.601	0.127		0.680	0.319	0.289	0.049
	L_q	-0.257	-0.074		0.477	0.180	0.144	0.036
	Med	0.412	0.154		0.692	0.307	0.277	0.045
	U_q	1.125	0.389		0.873	0.425	0.401	0.058
		(20.8/14.8)	(30.2/21.5)		(78.5/73.2)			
MF+ATM	Mean	0.714		0.252	0.519	0.311	0.281	0.049
	L_q	-0.025		-0.254	0.092	0.170	0.133	0.036
	Med	0.631		0.214	0.576	0.302	0.271	0.046
	U_q	1.399		0.676	1.053	0.425	0.400	0.060
		(31.5/23.5)		(18.8/11.4)	(32.9/23.5)			
His+MF+ATM	Mean	0.067	0.118	0.197	0.483	0.336	0.292	0.047
	L_q	0.003	-0.086	-0.280	0.049	0.196	0.142	0.035
	Med	0.036	0.153	0.138	0.572	0.330	0.285	0.044
	U_q	0.157	0.363	0.599	0.983	0.453	0.415	0.057
		(22.8/13.4)	(26.2/18.1)	(14.1/9.4)	(29.5/18.1)			

Table 7 Summary of the firm specific variables

The table contains the definitions and estimation methods of the firm specific variables that we use to investigate the properties of firms whose historical volatility, model-free volatility expectation or ATM implied volatility performs the best in measuring realized volatility. All variables for each firm are the time-series mean of daily measures from January 1996 to December 1999. In the calculation of option-related variables, only the options used in our study are included. Moneyness is defined as the option strike price over forward price that has the same time to maturity as options. The last column shows the objective that the specific variable is used to explain. OP refers to both model-free volatility expectation, MF, and ATM implied volatility, ATM, that are based upon option prices and His is historical volatility.

Variables	How to measure...	To explain...
Log(TV_OP)	The natural logarithm of the firm's average option trading volume.	OP VS His
Log(TV_Stock)	The natural logarithm of the firm's average stock trading volume.	OP VS His
Log(FirmSize)	The natural logarithm of the firm's average firm size in 1000 dollars, where daily firm size, or market capitalization, is calculated as the number of shares outstanding in the market multiplied by the stock closing price.	OP VS His
Moneyness Range	The average of daily option moneyness ranges, which is the daily maximum moneyness minus the minimum moneyness.	MF VS ATM
Number of Strike Prices	The average number of available option strike prices.	MF VS ATM
Average Strike Prices Interval	The average interval between each two adjacent strike prices, in moneyness.	MF VS ATM
TV_ATM/TV_ALL	The ratio between the ATM option trading volumes over all option trading volumes.	MF VS ATM
TV_IntermediateDelta/TV_ALL	The ratio between the trading volumes of intermediate delta options over all trading volumes, where intermediate delta options are defined as the options with delta within the interquartile range for Black Scholes delta values.	MF VS ATM

Table 8 Summary statistics of the firm specific variables for firms in different groups

The table contains the sample means and standard deviations (shown in parentheses) of the selected firm specific variables for the firms satisfying the orders in the first column. MF, ATM and His refer to model-free volatility expectation, ATM implied volatility and historical volatility. The symbol “OP” includes both MF and ATM calculated from option prices. Panel B shows the statistics for the firms with OP performing better than His in both one-day-ahead and option life forecasts, and the statistics for the firms with His performing better than OP in both one-day-ahead and options’ life forecasts. Panel C shows the statistics for the firms with MF performing better than ATM in both one-day-ahead and options’ life forecasts, and the statistics for the firms with ATM performing better than MF in both one-day-ahead and option life forecasts. The t statistics in Panel B and Panel C are the Student’s t for the one-sided hypothesis tests of the differences between the means of variables in the upper group and of those in the lower group. The p – value for each Student’s t statistic is reported.

	No. of firms	Ln(TV_OP)	Ln(TV_Stock)	Ln(FirmSize)	Moneyness Range	Number of Strike Prices	Average Strike Prices Interval	TV_ATM/TV_ALL	TV_IntermediateDelta/TV_ALL
<i>Panel A: for all firms</i>									
	149	6.18 (1.42)	13.92 (1.07)	22.60 (1.65)	0.39 (0.11)	5.44 (1.54)	0.10 (0.03)	0.42 (0.06)	0.62 (0.07)
<i>Panel B: OP VS His</i>									
His>OP	16	5.27 (1.06)	13.19 (1.17)	21.96 (1.59)	0.36 (0.10)	4.63 (0.57)	0.10 (0.04)	0.43 (0.06)	0.60 (0.09)
OP>His	73	6.82 (1.29)	14.43 (0.86)	23.17 (1.58)	0.39 (0.12)	5.90 (1.78)	0.09 (0.03)	0.43 (0.06)	0.64 (0.04)
t statistics		-5.08	-4.03	-2.75	-1.21	-5.03	1.47	0.06	-1.95
p – value		0.000	0.000	0.006	0.120	0.000	0.079	0.478	0.034
<i>Panel C: MF VS ATM</i>									
MF>ATM	37	6.07 (1.40)	13.87 (1.03)	22.28 (1.49)	0.40 (0.11)	5.14 (1.18)	0.10 (0.03)	0.43 (0.06)	0.63 (0.06)
ATM>MF	60	6.20 (1.31)	13.91 (1.05)	22.81 (1.58)	0.37 (0.10)	5.35 (1.12)	0.09 (0.03)	0.42 (0.06)	0.62 (0.08)
t statistics		-0.43	-0.19	-1.66	1.37	-0.87	1.71	0.65	0.89
p – value		0.336	0.424	0.050	0.088	0.194	0.045	0.260	0.188