*Economics Department*

*Economics Working Papers*

*The University of Auckland* *Year* 2001

# Forecasting Volatility:Evidence from the German Stock Market

Hagen Bluhm*  Jun Yu†

*

†University of Auckland, j.yu@auckland.ac.nz

# Forecasting Volatility: Evidence from the German Stock Market*

Hagen H.W. Bluhm[a], Jun Yu[b]

February 2001

**Abstract**

In this paper we compare two basic approaches to forecast volatility in the German stock market. The first approach uses various univariate time series techniques while the second approach makes use of volatility implied in option prices. The time series models include the historical mean model, the exponentially weighted moving average (EWMA) model, four ARCH-type models and a stochastic volatility (SV) model. Based on the utilization of volatility forecasts in option pricing and Value-at-Risk (VaR), various forecast horizons and forecast error measurements are used to assess the ability of volatility forecasts. We show that the model rankings are sensitive to the error measurements as well as the forecast horizons. The result indicates that it is difficult to state which method is the clear winner. However, when option pricing is the primary interest, the SV model and implied volatility should be used. On the other hand, when VaR is the objective, the ARCH-type models are useful. Furthermore, a trading strategy suggests that the time series models are not better than the implied volatility in predicting volatility.

*JEL classification*: G12, G15

*Keywords*: Forecasting Volatility; ARCH Model; SV Model; Implied Volatility; VaR; Germany

---

*We would like to thank Deutsche Boerse AG, the German stock exchange, for providing the data, and the University of Auckland Research Committee for financial support.
[a] Bankgesellschaft Berlin AG; Email: Hagen.Bluhm@ib.bankgesellschaft.de
[b] Department of Economics, The University of Auckland, Private Bag 92019, Auckland, New Zealand; Email: j.yu@auckland.ac.nz.

## 1. Introduction

Using daily data from the German stock market, this paper compares two basic approaches to forecast volatility. The first approach uses various univariate time series techniques while the second approach makes use of volatility implied in option prices. The time series models include four ARCH-type models and a stochastic volatility (SV) model. We focus on the forecast horizons of 1, 10, and 180 trading days and 45 calendar days. The evaluation criteria used are mean squared prediction error (MSPE), bounded violations, and the LINEX loss function. A trading strategy is also used to examine the usefulness of the time series models in predicting volatility.

Forecasting financial market volatility has received extensive attention in the literature by academicians and practitioners in recent time; see Poon and Granger (2000) (hereafter PG) for an excellent review of the literature. Although "volatility forecasting is a notoriously difficult task" according to Brailsford and Faff (1996) (hereafter BF), it is generally agreed that volatility is predictable and hence the market for volatility is not as efficient as that for returns. Broadly speaking there are two ways to forecast volatility. The first method uses the historical return information only while the second makes use of volatility implied in option prices. The existing empirical evidence is conflicting in three ways. First, within the first method, the performance of the models depends on the data, forecasting horizon, sampling frequency, and evaluation criteria (cf BF). Second, within the second method, because of the volatility smile, a typical feature of implied volatility, it is not entirely clear how to extract volatility from option prices (cf PG). Third, when comparing time series forecasts with option forecasts, people have found conflicting evidence; see, for example, Jorion (1995) and Canina and Figlewski (1993) for evidence for and against the option forecasts respectively.

This paper complements the literature in three ways. First, we use data from a country which receives little attention in the literature and yet is important in the international framework. Second, we compare SV forecasts with option forecasts. The SV model provides more realistic and flexible modeling of financial time series than the ARCH-type models, since it essentially involves two noise processes. The better in-sample fit of the SV model over the ARCH-type models has been documented in the literature (see, for example, Danielsson (1994), Geweke (1994), and Kim et al. (1998)), however, the SV model receives much less attention in the volatility forecasting literature. To our present knowledge, there is only one paper published. Using New Zealand data, Yu (1999) finds that the SV model performs better than all the other univariate time series models, including the ARCH-type models. Although recent efforts have compared ARCH forecasts and option forecasts, nothing has been done to compare SV forecasts and option forecasts. Third, forecast horizons and error measurements have been arbitrarily chosen in the literature. In this paper, the forecast horizons and error measurements are selected based on the utilization of volatility forecasts in the financial industry. In particular, we use option pricing and Value-at-Risk (VaR) as the practical guidance to choose forecast horizons and error measurements.

The paper is organized as follows. Section 2 reviews the features of the German stock market. The data and descriptive statistics are in Section 3. Section 4 outlines the methods used in this paper for volatility forecasts. Section 5 describes the forecast horizons and the error measurements. Section 6 discusses the empirical results and Section 7 concludes.

## 2. The German Stock Market

The importance of the German economy is reflected in its stock market, which ranks fourth in terms of market capitalization and third in terms of turnover. In particular, the fast-growing German option and future exchange, which is the second largest in the world, demonstrates that the German stock market has drawn a lot of international attention; see Figure 1 for the comparison of the market capitalization and turnover at six major international markets in the world.

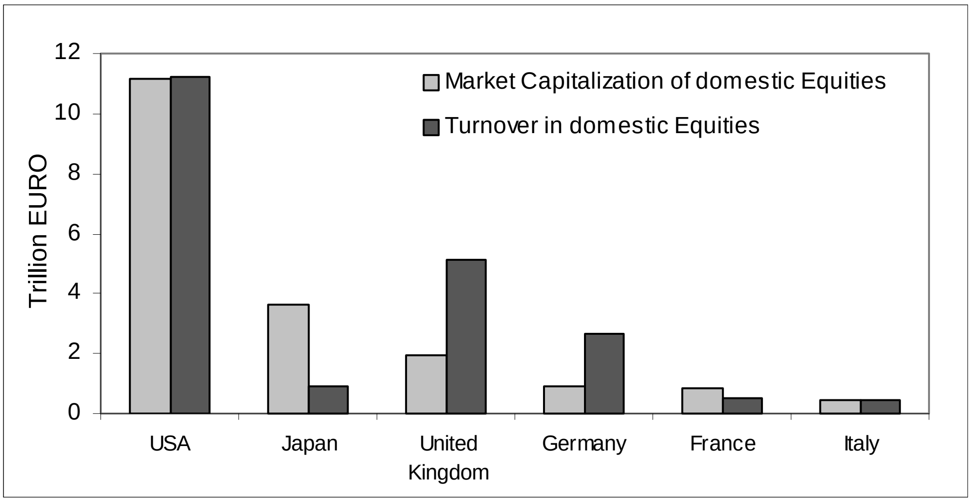


Figure 1: Importance of International Stock Markets

The German stock market consists of eight regional stock markets, where the Frankfurt exchange with 78% of the combined turnover is the most important. All these regional markets are based on open-outcry trading. Stocks at the regional stock markets are traded in two different ways. The opening, midday and closing prices of each stock are calculated using an auction system. Trading in between takes place in the conventional (continuous) way. All orders in one stock that arrive before the opening auction at 8:30am are gathered. The opening price is the one at which the highest number of stocks is traded. The same happens at the 1pm auction and at the closing auction at 5pm. Since there is a minimum order size for continuous trading, the auction system ensures that all orders, especially small orders, are executed. Most importantly, the closing auction system ensures that for most stocks the closing price will be based on trade at 5pm.

In addition to the regional open-outcry markets, Germany has an electronic trading system called XETRA. Since its introduction in 1991 XETRA steadily increased its turnover share in the biggest stocks. At the end of 1998, for example, XETRA represented 68% of the turnover in the DAX equities traded in the whole country. Trading in XETRA takes place from 8:30am until 5:15pm, with an opening auction at 8:30am and a closing auction at 5:15pm

The Deutsche Aktien index (DAX) represents the 30 largest domestic stocks traded in Germany. At the end of 1998, stocks included in the DAX index accounted for 76% of the total market capitalization in Germany and 80% of the equity turnover in Frankfurt, making the DAX a highly representative index for the German stock market.

To more easily signal volatility to investors, the German stock exchange introduced a volatility index based on implied volatilities of DAX options in December 1994. It is called VDAX. The VDAX index is based on linear interpolation of the volatilities of the two sub-indices that are nearest to a remaining lifetime of 45 calendar days. More details about the VDAX index are provided in Section 4.

## 3. DATA

### 3.1 *DAX Returns*

One data series we have is the daily DAX index from January 1, 1988 to June 30, 1999 and is based on daily closing auction prices at the Frankfurt stock exchange. We use logarithmic returns calculated from the DAX series, resulting in 2,876 daily return observations.

Inspecting the DAX return series as depicted in Figure 2 reveals that, as expected, volatility is not constant over time and moreover tends to cluster. Periods of high volatility can be distinguished from low volatility periods. Apart from the clustering of large negative returns in August 1998, there are three outstanding returns that are large in absolute value. The first is the 13.7% fall on October 16, 1989 in the wake of the burst merger bubble in the United States. The second is the -9.9% return on August 19, 1991, the day of the coup against Gorbatchev in the Soviet Union, an event that severely affected the German market. The main reason for the fall of 8.4% on October 28, 1997 was the Asian crisis.

The mean daily return of the DAX series is 0.0585%. The standard deviation of the daily returns is 0.01259, which is equivalent to an annualized volatility of 20%. The series also exhibits a negative skewness of –0.802 and an excess kurtosis of 9.8, indicating that the returns are not normally distributed. The Jarque-Bera statistic of 11,824 also shows that we have to reject normality with a p-value of one. These findings are consistent with other financial time series.

Figure 3 plots a histogram of the data and a normal density whose mean and variance match sample estimates. It shows that numerous returns are above four standard deviations, which is highly unlikely in the normal distribution. This is evidence of leptokurtosis and commonly found in financial time series.

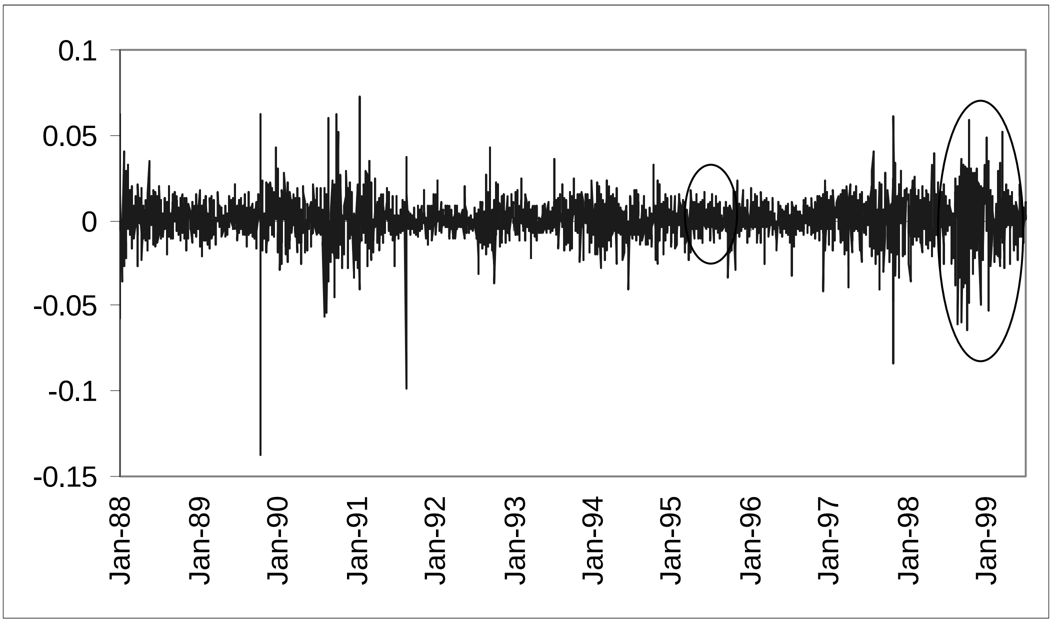


Figure 2: Return Series for FRA-DAX**.** The ovals indicate a low volatility period and a high volatility period.

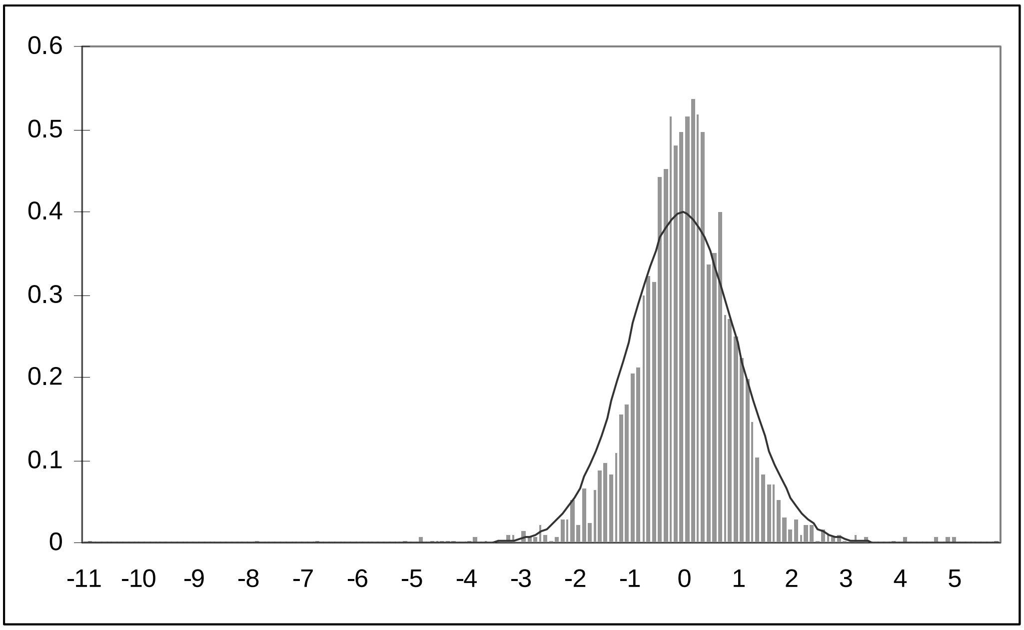


Figure 3: FRA-DAX Normalized Return Distribution

The Ljung-Box Q-statistics for the first 50 lags is 24.22 for returns and 314.5 for squared returns. Therefore, we cannot reject the hypothesis that there is no serial correlation in the level of returns but we have to reject the same hypothesis in squared returns.

Using the augmented Dickey-Fuller (ADF) unit root test we can clearly reject, as expected, the hypothesis of a unit root in the return process. The ADF t-statistic is –24.36 which rejects the unit root hypothesis with a confidence level of more than 99%. Furthermore we also have to reject the hypothesis of a unit root in the squared return process, which is an approximation of the volatility process, where the ADF-t test statistic is equal to –19.66.

Based on the various univariate time series models which will be reviewed in Section 4, the DAX return series is used to forecast volatility.

3.2 *VDAX*

Figure 4 plots VDAX for the period from January 1992 to July 1999. Two special events stand out in this figure. First, the sharp increase in October 1997 in the wake of the Asian crisis. Second, the jump in implied volatility after the Russian bond default, where the VDAX increased from 29.73% on August 21, 1998 to a high of 56.31% on October 2, 1998. The VDAX index will also be used to forecast volatility.

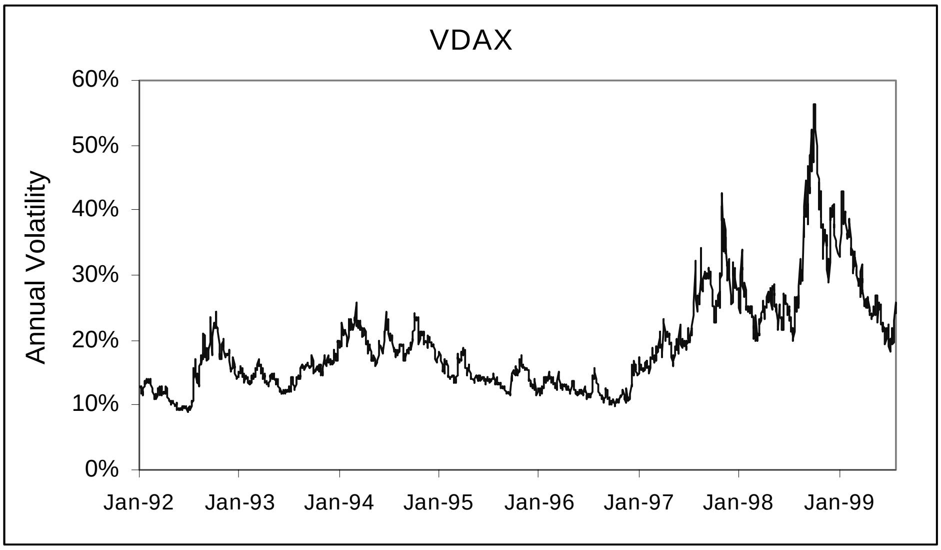

Figure 4: DAX Volatility Index (VDAX)

3.3 *Actual Volatility*

To assess the performance of various methods, we need to compare forecasted volatilities with actual volatilities. Unfortunately, the actual volatility is not directly observed and hence it has to be estimated. A common approach in the literature is to use the absolute or squared daily return to estimate the daily volatility. In the more recent literature, daily volatility has been estimated from high frequency data.

When using high frequency data, such as tick-by-tick data, particular attention has to be paid to the possible negative autocorrelation caused by the "bid-ask bounce". If prices are recorded from transactions, the price of each transaction might bounce between the bid and ask price causing negative autocorrelation. Neglect of the negative serial autocorrelation will lead to an upward-biased estimation of daily volatility.

It has been found that improvements can be made in estimating actual volatility based on high frequency data. For example, using 5-minute intra-daily data, Andersen and Bollerslev (1998) find that the GARCH model provides accurate volatility forecasts. Blair, Poon and Taylor (2000) support this by reporting an 11.5% to 41.4% increase in one-day-ahead prediction $R^2$ when daily volatility has been estimated using high-frequency data. Since we only have a data set at the daily frequency, we calculate the volatility in a certain period simply as the square root of the sum of squared daily returns in that period; that is,

$$\sigma_T = \sqrt{\sum_{t=1}^{N_T} r_t^{\,2}} \tag{3.1}$$

where $r_t$ is the daily return on day t and $N_T$ is the number of trading days in that period.[1]

## 4. Volatility Forecasting Techniques

### *4.1 Forecasting models using the historical return information only*

The forecasting models using the historical return information only include the historical mean model, the exponentially weighted moving average (EWMA) model,[2] the GARCH model (Bollerslev (1986)), the GJR-GRACH model (Glosten, et al (1993)), the EGARCH model (Nelson (1990)), the GARCH-M model (Engle et al (1987)), and the SV model (Taylor (1986)). The historical mean model, the EWMA model, the GARCH(p,q) model and the GJR-GRACH(p,q) model are defined in BF. We now define the other models. In all cases we assume $r_t$ is the return at period t and $\varepsilon_t \mid I_{t-1} \sim \quad N(0,1)$.

The EGARCH(p,q) model is defined by

$$r_t = \mu + \sigma_t \varepsilon_t$$

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \left| \varepsilon_{t-i} \right| + \sum_{i=1}^{q} \gamma_i \varepsilon_{t-i} + \sum_{j=1}^{p} \beta_j \log \sigma_{t-j}^2 .$$

The GARCH-M(p,q) model is defined by

$$r_t = \mu + \delta \sigma_t^2 + \sigma_t \varepsilon_t ,$$

[1] Obviously, volatility is defined as standard deviation instead of variance in this paper.
[2] This model is referred to in BF as the exponential smoothing model although it is more conventional to refer it to in the forecasting literature and industry as the EWMA model.

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} a_i r_{t-i}^2 + \sum_{j=1}^{p} \beta_i \sigma_{t-j}^2 .$$

The system parameters in all the ARCH-type models are estimated using the maximum likelihood method. The lag parameters p and q are selected using the BIC criterion. To obtain h-step ahead forecast of the volatility, we follow Baillie and Bollerslev (1992).

The SV model is defined by

$$r_t = \mu + \exp(h_t / 2)\varepsilon_t ,$$

$$h_t = \omega + \alpha h_{t-1} + \eta_t ,$$

where $\eta_t \sim N(0, \sigma_\eta^2)$. Compared to the ARCH-type models, the SV model provides a more flexible modeling of financial time series, since it essentially involves two noise processes, one for the observations, and one for the latent volatilities. Unfortunately, the likelihood function for the SV model has no closed form expression and therefore maximum likelihood (ML) estimation is applicable.

Recently several methods have been proposed to estimate the SV model. Such methods include Quasi ML (QML) proposed by Harvey et al (1994), simulated ML by Danielsson (1994), GMM by Andersen and Sorensen (1996), Markov Chain Monte Carlo (MCMC) by Jacquier, Polson and Rossi (1994) and Meyer and Yu (2000), Efficient Method of Moments by Gallant and Tauchen (1994), Monte Carlo maximum likelihood by Sandmann and Koopman (1998), and the empirical characteristic function method by Knight, Satchell and Yu (1998). Some of these methods, such as QML and MCMC, also produce forecasts of volatility as by-products. MCMC provides the exact optimal predictors of volatility, however, it is computationally intensive. Despite its inefficiency, the QML method is consistent and very easy to implement numerically. Following Yu (1999), we use QML to estimate parameters in the SV model and obtain h-day ahead volatility forecasts.

There remain two questions before we can perform our model evaluation. The first is how we divide the sample. The 2,126 observations from January 4, 1988 to June 28, 1996 are used to fit the models and the out-of-sample period covers the remaining 750 observations from July 1, 1996 to June 30, 1999. This division is arbitrary, but the out-of-sample period covers periods of both low volatility and extremely high volatility in 1998, making accurate volatility forecasts more difficult. The other question is what sample we should use for model fitting as additional observations beyond June 28, 1996 become available. In this paper we use the method of expanding forecast windows. That is, beginning with the last day of the out-of-sample period we compute a volatility forecast for each horizon. For the second day of the out-of-sample period the volatility forecasts are obtained in the same way, but the information set includes the realized return of the first out-of-sample day.

### 4.2 *Deriving Implied Volatilities*

While future volatility can be forecasted using historical return information, it can also be derived from option prices. For example, the Black-Scholes formula for pricing a European call option on a non-dividend paying stock needs five input variables: the maturity of the option, the strike price, the current price of the underlying, the risk-free interest rate associated to the maturity of the option and the volatility of the underlying stock over the lifetime of the option. Once the option is traded, the only unknown parameter is the volatility. Theoretically, therefore, the Black-Scholes formula can be used to derive the implied volatility which should express the market volatility forecast of the underlying asset.

Unfortunately, in practice, the implied volatility depends on the option pricing model used and is strictly associated with the maturity and strike of the option. Therefore, we cannot derive a unique implied volatility for the underlying security. A typical feature of implied volatility is the so-called volatility smile, which results from the observation that out-of-the-money and in-the-money options have a higher implied volatility than at-the-money options. Figure 5 shows an implied volatility matrix for DAX options on June 21, 1999. The price of the underlying was 5,414 (at-the-money point). Volatility smile can be easily identified.

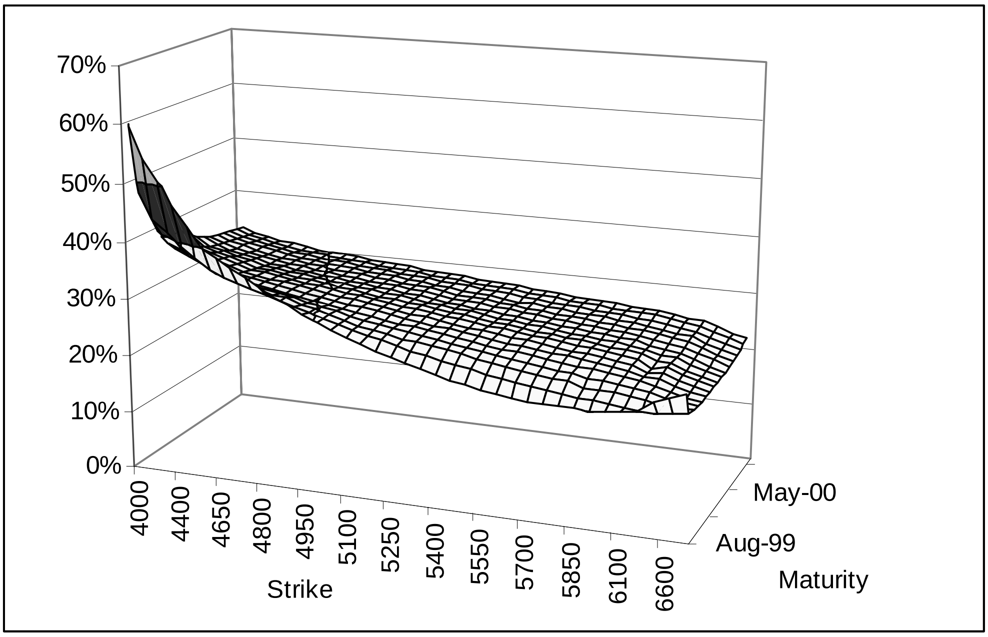


Figure 5: Implied Volatility Matrix for DAX Call Options on June 21, 1999

There are several other problems involved in deriving implied volatilities from option prices. First, most option price series are not synchronous with prices of the underlying assets. Using daily closing prices for an option and its underlying might have the effect that the last price of an option is based on an earlier price of the more liquid underlying and not on the closing price. Second, relative bid-ask spreads are large for options, especially for out-of-the-money and in-the-money options, making it difficult to derive the 'correct' option price. Third, the option forecasts work less well if the option market is less liquid.

In the literature most of the papers using the option forecasts focus on the U.S. market where the most liquid data are available.[3] To our present knowledge, there are only three papers using data from outside the U.S. market. In particular, Edey and Elliot (1992) use the Australian data while Vaselellis and Meade (1996) and Gemmill (1986) use the British data. In this paper, we use VDAX, a German data set, to forecast future volatility. The VDAX is based on linear interpolation of the volatilities of the two sub-indices that are nearest to a remaining lifetime of 45 calendar days.[4] The maturity of 45 days of the synthetic underlying option remains constant. It means that the VDAX is lifetime-independent and therefore does not expire, eliminating the effects of strong fluctuations of volatility, which typically occur close to expiry. The sub-indices correspond to the maturities of the currently traded DAX options and are calculated using four options (two calls and two puts) whose strike is closest to the current price of the underlying.[5] Using the implied volatility of these four options, based on the Black-Scholes model, the volatility value of the sub-index is calculated as follows:

$$V_i = \frac{(X_h - F_i) * (v_t^{Put} + v_t^{Call}) + (F_i - X_t) * (v_h^{Put} + v_h^{Call})}{2(X_h - X_t)},$$

where $V_i$ is the implied volatility sub index corresponding to maturity i; $F_i$ the forward or futures price corresponding to maturity i; $X$ the strike price of an individual option with h referring to a strike price of the option above the current underlying price and t to a strike price of the option below the current underlying price; $v$ the implied volatility of an individual option.

It should be stressed that VDAX takes only at-the-money options into account, neglecting the out-of-the-money and in-the-money options that are not liquid. By construction VDAX can overcome some of the aforementioned problems in deriving implied volatility from option prices.[6] First, it records every ten seconds all necessary input data and calculates new values for sub-indices only when the relevant option prices have changed. Since the underlying is traded very frequently, the option price always corresponds to the underlying price. Second, only options whose bid-ask spread is not more than 15% of the bid price are taken into account and hence can reduce

[3] See Blair, Poon and Taylor (2000) and the references therein.

[4] Using the same idea as obtaining volatility forecasts for horizon of 45 calendar days from VDAX, we can obtain volatility forecasts for other horizons.

[5] Instead of the DAX index itself as the underlying, the forward price of the DAX index is used, which can be easily obtained from the DAX future prices.

[6] How to forecast future volatility from options with different strikes has recently received a great deal of attention in the literature; see, for example, PG for a review.

measurement errors. Third, VDAX is based on the second largest option and future exchange market in the world and hence is reasonably liquid.

## 5. Forecast Horizons and Error Measurements

### 5.1 *Forecast horizons*

Although a variety of forecast horizons have been used in the literature, they are arbitrarily chosen without resorting to practical guidance. It is known that volatility forecasts have been widely used in financial institutions for various purposes, of which option pricing and VaR are of particular interest from a practical viewpoint. In this paper we use the forecast horizons by taking into account the practical requirements of volatility forecasts in option pricing and VaR.

Until recent years volatility forecasts have been almost exclusively needed as an input variable in the Black-Scholes option pricing model. This typically requires forecast horizons between a month and a year. Furthermore, the most liquid options traded at option exchanges are those with a short maturity, usually one to three months. To compare the time series forecasts with the option forecasts we choose a forecast horizon of 45 calendar days, which is exactly the same as VDAX. Since many exchanges also trade long-term options with maturities up to 2 years, we want to test the ability of the models to forecast volatility over longer horizons. In addition to the 45 calendar days, we also use a forecast horizon of 180 trading days (about nine calendar months).

With the growing usage of VaR as a risk management tool, the need for short-term variance and covariance forecasts becomes greater. Most investment banks are interested in forecasting the risk of their portfolios until the closing of the next trading day. In addition, the Basle Capital Accord in 1998 (Basle Committee On Banking Supervision (1999)) requires banks to compute VaR over a horizon of ten trading days. A related application for volatility forecasts is margining at future and option exchanges. To ensure that market participants are able to fulfill her financial obligation resulting from future and option contracts, the exchange requires them to deposit a margin. This often-called initial margin should cover the anticipated price risk of their positions. The German future and option exchange (EUREX) currently calculates initial margins based on 30 and 250 days of historical volatilities. These volatilities are used to forecast the 99% confidence interval for tomorrow's price of the underlying. Resulting from this price interval and the position of the market participants, the EUREX calculates the initial margin. Therefore, for the purpose of VaR and risk-based margining we use forecast horizons of one and ten trading days.

### 5.2 *Error Measurements*

In the literature a variety of statistics have been used to evaluate and compare forecast errors These include root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), mean mixed error (MME), the Theil-U statistic, and the LINEX loss function. RMSE, MAE and MAPE are conventionally used error statistics. MME is proposed by BF while Yu (1999) advocates the use of the Theil-U and LINEX loss function. As with the choice of forecast horizons, however, the evaluation statistics are also arbitrarily chosen.

Deciding which error measurement should be applied to the volatility forecasts, we believe, two questions have to be answered. First, is the absolute or relative deviation important? Second, does it matter whether the volatility was over-predicted or under-predicted?

Where option pricing is concerned, the first question is equivalent to whether relative or absolute trading profits matter. This is because the price of an at-the-money option is a linear function of the volatility according to the Black-Scholes model. Since in the financial world relative profits are more important it can be assumed that an option trader is interested in the relative deviation between forecasted and realized volatility. This is why we use MAPE which is defined by,

$$MAPE = \frac{1}{T}\sum_{\tau=1}^{T}\frac{\left|\hat{\sigma}_{\tau+h} - \sigma_{\tau+h}\right|}{\sigma_{\tau+h}}, \qquad (5.1)$$

where T is the number of out-of-sample observations minus the number of days of the forecast horizon; $\sigma_t$ the actual volatility at the period t; $\hat{\sigma}_t$ the forecasted volatility at the period t.

Since the Basle Committee's Market Risk Amendment to the Capital Accord in 1998 specifies that banks have to calculate the price risk of their financial activities and set aside sufficient capital to cover this market risk, VaR is today a standard tool used to comply with Basle Accord requirements. VaR uses a volatility forecast and a confidence level based on normally distributed returns to yield a potential loss. According to the Basle Capital Accord, banks have to set aside reserve capital as big as three times the potential loss that will not be exceeded with 99% probability. Depending on the accuracy of the volatility forecast, however, there can be more or less than 1% of the returns outside the boundary implied by the potential loss.

The regulation introduced by the Capital Accord emphasizes the need for accurate volatility prediction. Over-prediction of future volatility over a long period requires more costly capital, while under-prediction leads to more boundary violations than implied by the confidence level. If VaR is used only as an internal risk management tool, the boundary violations would be the concern and they should at least be as low as what the confidence level implies. If VaR is used in line with the Basle Capital Accord, the price risk of the portfolio has to be covered by equity, and hence over- and under-prediction matters. It can be assumed that an under-prediction matters more than over-prediction. This is because the model will not be accepted by the regulatory body if the boundary violations are higher than implied. The situation is certainly worse than providing a little bit more capital when the price risk was overestimated.

Where VaR is concerned, therefore, we will apply two error measurements, the number of boundary violations that occur during the forecast horizon and the LINEX loss function. Deviating from the LINEX loss function used in Yu (1999) we will use a LINEX loss function defined by

$$L(a) = \frac{1}{T}\sum_{\tau=1}^{T}\left|\hat{\sigma}_\tau - \sigma_\tau\right| + \left|\exp\left(\alpha(\hat{\sigma}_\tau - \sigma_\tau)\right) - 1\right| \quad , \tag{5.2}$$

where $\alpha$ is a given parameter which captures the degree of asymmetry. The LINEX loss function in Yu (1999) would yield for $\alpha = 0$ a forecast error equal to zero even though $\hat{\sigma}_\tau$ differs from $\sigma_\tau$. In (5.2), however, we still have a non-trivial symmetric loss function when $\alpha = 0$. Figure 5 plots the LINEX loss function with different values of $\alpha$.[7]

There are no rules as to how to choose the optimal $\alpha$, unfortunately. In general it should depend on how much one dislikes/likes under-prediction relative to over-prediction. In this paper we choose $\alpha = -30$, which implies the LINEX value from under-prediction is about 25% higher than that from over-prediction. Also we choose $\alpha = 0$.

Combinations of the forecast horizons and the error measurements produce four test settings, which are summarized in Table 1.

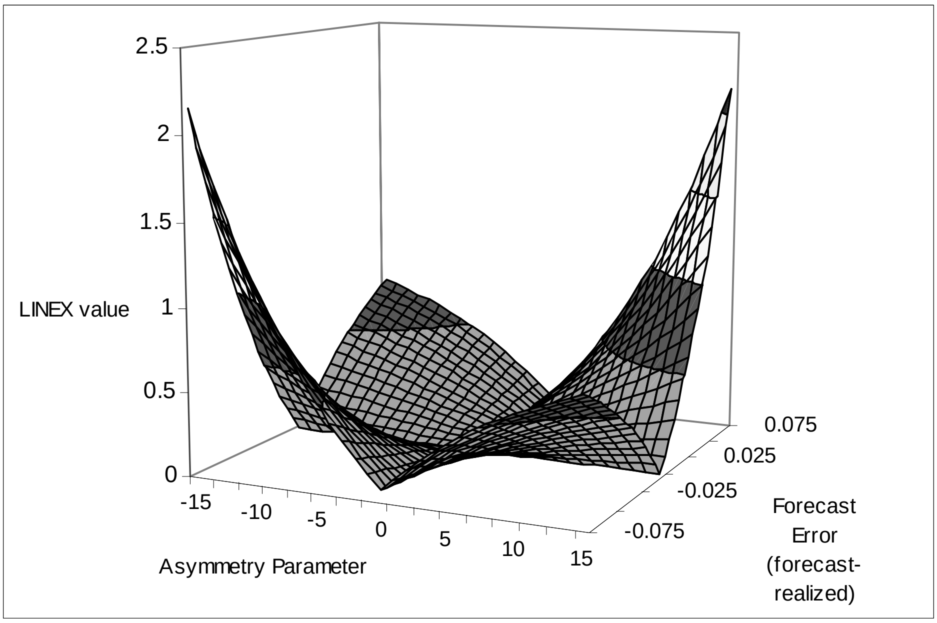


Figure 5: LINEX Loss Function

## 6. Empirical Results

### *6.1 In-sample Fit*

To frame our discussion, we briefly present some results from in-sample fit based on the first of our expanding samples. For the EWMA model, the value of damper coefficient is chosen to produce the best fit by minimizing the sum of squared in-sample forecast errors. This yields the optimal value of 0.9568. As to the choice of p and q in the ARCH-type models, the BIC criterion selects GARCH(1,3), GJR-GARCH(1,3), EGARCH(2,1) and GARCH-M(1,3) respectively. Moreover, within the ARCH family, the EGARCH model has the lowest BIC value and hence the best in-sample fit. The superior in-sample fit of the EGARCH model is also supported by the news impact curve test introduced by Engle and Ng (1993). In terms of the estimated persistence in the variance equation, the GJR-GARCH model and the EGARCH model are the lowest, followed by the GARCH and the GARCH-M models. More detailed in-sample results are available on request.

| **Test** | **Utility** | **Forecast Horizon** | **Error Measurement** |
|---|---|---|---|
| 1 | Option pricing | 45 calendar days | MAPE |
| 2 | Option pricing | 180 trading days | MAPE |
| 3 | VaR | 1 trading day | Boundary violations/ LINEX loss function |
| 4 | VaR | 10 trading days | Boundary violations/ LINEX loss function |

Table 1: Overview of the Test Settings

### *6.2 Out-of-sample Forecasts*

In Table 2 we report the actual value of forecast error measurements and ranking for all the time series models and VDAX under Test 1 and Test 2, and those for the time series models only under Test 3 and Test 4.

There are several results emerging from Table 2. Firstly, no single method is clearly superior. For example, Test 1 indicates that the SV model provides the most accurate forecasts followed by VDAX as second while under Test 2 VDAX ranks first. Test 3 favors the GJR-GARCH model and the EGARCH model, but Test 4 favors EWMA, the GARCH-M model, and the SV model. The model rankings depend on the forecast horizon as well as the forecast error measurement. This observation is consistent with the existing literature when only the various time series models compete; see, for example, BF.

Secondly, as expected, implied volatility appears to be a good predictor of future volatility. It ranks first under Test 1 and second under Test 2 and clearly outperforms the ARCH-type forecasts under both tests. This observation is consistent with the findings of Jorion (1995). When comparing

[7] Unfortunately the LINEX loss function given in (5.2), unlike Yu's LINEX loss function, has no closed form expression for the LINEX optimal forecast.

SV forecasts and implied volatility forecasts, however, neither is dominant and what we can conclude is there is a trade-off between them.

Thirdly, when VaR is the concern, it appears that various ARCH-type models are useful. For example, the GJR-GARCH ranks first for 1-day-ahead forecast under boundary violations and the LINEX loss L(-30); the EGARCH model ranks first for 1-day-ahead forecast under L(0); GARCH-M model ranks first for 10-day-ahead forecast under L(-30). Although no ARCH-type model ranks first for 10-day-ahead forecast under boundary violations or for 10-day-ahead forecast under L(0), the GARCH model and the EGARCH model rank a close second in both cases respectively.

| | Test 1 | Test 2 | Test 3 | | | Test 4 | | |
|---|---|---|---|---|---|---|---|---|
| Horizon | 45cds | 180tds | 1 td | | | 10 tds | | |
| Statistics | MAPE | MAPE | BV(99%) | L(-30) | L(0) | BV(99%) | L(-30) | L(0) |
| Historical Mean | 29.45% (7) | 27.35% (4) | 3.87% (7) | 246.7 (6) | 7.757 (7) | 5.13% (5) | 150.2 (7) | 4.687 (7) |
| EWMA | 29.07% (4) | 30.87% (5) | 2.80% (2) | 244.6 (4) | 7.706 (6) | 2.83% (1) | 140.4 (4) | 4.366 (6) |
| GARCH | 29.58% (8) | 22.68% (2) | 3.07% (3) | 244.7 (5) | 7.688 (5) | 2.97% (2) | 134.1 (2) | 4.150 (4) |
| GJR-GARCH | 29.39% (5) | 34.38% (7) | 2.67% (1) | 238.5 (1) | 7.469 (3) | 5.13% (5) | 142.7 (6) | 4.177 (5) |
| EGARCH | 28.92% (3) | 49.79% (8) | 3.47% (6) | 240.2 (2) | 7.286 (1) | 5.13% (5) | 142.3 (5) | 4.077 (2) |
| GARCH-M | 29.39% (5) | 22.79% (3) | 3.07% (3) | 243.3 (3) | 7.621 (4) | 3.24% (3) | 133.9 (1) | 4.129 (3) |
| SV | 23.09% (1) | 33.09% (6) | 3.20% (5) | 247.6 (7) | 7.339 (2) | 4.45% (4) | 135.7 (3) | 3.947 (1) |
| VDAX | 25.16% (2) | 21.24% (1) | | | | | | |

Table 2: Results of out-of-sample forecasts. MAPE is defined by (5.1); BV(99%) is the percentage boundary violations for a confidence level of 99%; L(-30) and L(0) are the LINEX values for $\alpha$=-30 and $\alpha$=0 scaled by 1000. The number in brackets is the ranking under each test.

Fourthly, the performance of the ARCH-type models with less persistence gets worse when the forecast horizon gets longer. For example, by the boundary violations and L(-30), the GJR-GARCH model, a favorite model in BF, ranks first under Test 3, and ranks dismally fifth and sixth under Test 4. Also, its MSPE increases from 29.39% to 34.38% when the forecast horizon increases from 45 calendar days to 180 trading days. The reason for this is the rapid decline in the forecasted squared returns due to the lowest persistence found in the in-sample estimation of the GJR-GARCH model. Since the volatility is determined by the sum of the squared returns, low persistence tends to generate lower volatility forecasts over longer horizons. BF also finds that the GJR-GARCH model under-predicts volatility more often than other models. However, they offer no reason for this behavior. The same observation and explanation apply to the EGARCH model. Moreover, the low persistence also attributes to the wider range of the MSPE values across the models for the horizon of 45 calendar days relative to those for the horizon of 180 trading days.

Fifthly, the forecast performance of the simple average model is rather poor, especially for short horizons. In particular, under both Test 3 and Test 4 it has the highest number of boundary violations and the LINEX loss values. The other simple model, the EWMA model, however, performs reasonably well. It has the lowest boundary violations in Test 4 and the second lowest in Test 3. The finding suggests that the EWMA model could be a suitable model for forecasting volatility in the VaR framework. Given that the EWMA estimator is extremely easy to calculate, this finding helps to justify why EWMA is often used in financial institutions. Dimson and Marsh (1990) also report evidence to support EWMA.

Sixthly and finally, judged by the relative number of boundary violations under Test 3 and Test 4, all time series models tend to under-predict volatility. However, given that the construction of boundaries relies on the assumption of normally distributed returns, a large value of boundary violations can be the result of the leptokurtic return distribution rather than the under-prediction of volatility.

### *6.3 Trading Strategy*

We need to emphasize that the tests in Section 6.2 are based on average errors. One can use a test statistic to compare the differences between two error distributions; see, for example, West and Cho (1996). Although such a test allows one to make a statistical inference about the model performance, in finance a more appealing approach is to use economic reasoning for the comparison. In this paper we use a trading strategy to test the usefulness of time series models.

This trading strategy is based on buying or selling call options on the DAX index. The option pricing formula derived by Black and Scholes (1973) is a positive function of the expected volatility of the underlying stock price. If we know the future volatility, we can construct a riskless portfolio by buying an option and selling the underlying stock or index. However, since the option's expected or implied volatility is only a forecast, this strategy is not riskless and can yield a profit or loss depending on the true volatility. If the true volatility is less than the implied volatility, buying the undervalued option and selling the underlying stock returns a profit. Similarly, selling an apparently overvalued option and buying the underlying stock will return a profit. If the volatility forecasts generated by time series models are superior to implied volatilities, one can use this trading strategy to generate profits.[8]

[8] A detailed description of this trading strategy can be found in Figlewski (1989).

Since the options used for this trading strategy have maturities between 7 and 67 calendar days, the optimal time-series model is selected based on the average of the MAPE values for Test 1 (45 days maturity) and Test 4 (about 14 days maturity). As Table 3 shows, this yields the EGARCH(2,1) model as the best model.

| | 10cds | 45cds | Mean |
|---|---|---|---|
| Historical Mean | 449% | 29.5% | 239% |
| EWMA | 466% | 29.1% | 247% |
| GARCH | 471% | 29.6% | 250% |
| GJR-GARCH | 410% | 29.4% | 220% |
| EGARCH | 374% | 28.9% | 201% |
| GARCH-M | 467% | 29.4% | 248% |
| SV | 425% | 23.1% | 224% |

Table 3: Average MAPE of Test 1 and Test 4

Using settlement (closing) prices for DAX options from July 2, 1999 to October 15, 1999, we calculate for each day an implied volatility based on an at-the-money call option with the shortest maturity available (minimum 7 days) and an implied volatility based on a medium term option with 37 and 67 days maturity. These two implied volatilities are then compared with the forecasts generated by the EGARCH model for horizons corresponding to the maturities of the options. If the EGARCH model's forecast is 20% higher than the implied volatility, the call option is bought; if the EGARCH model's forecast is 20% lower, the call option is sold. The return is the discounted costs (revenue) of buying (selling) the underlying index divided by the option premium. Since trading the index itself is extremely costly and given that the high indivisibility is actually impossible, we will use as underlying the future on the DAX index. The transaction costs that are taken into account here are the ones that arise from trading the future, which are also the main costs of a delta-hedge strategy[9]. The future's transaction costs consist of a bid-ask-spread of $\pm$ 0.5 index points of the settlement price and a fee of 1-EURO per traded future that the EUREX charges. These transaction costs are typical for institutional investors, and only a fraction of the costs that retail investors would pay performing volatility arbitrage.

The low persistence of the EGARCH model is once again obvious. For the option with the shorter maturity (denoted as option 1), on 28 out of 71 trading days the EGARCH model's forecast is 20% lower than the implied volatility and on one day 20% higher. For the option with the longer maturity (denoted as option 2), on 45 out of 51 trading days[10] the EGARCH model's volatility

[9] Since this strategy results in a spread of short position in a call option and long position in the underlying, the initial margin is small. Also, the costs for the initial option trade are negligible.
[10] The difference in trading days between option 1 and 2 arises because we have no market data in option 2 after September 11 which matures in November. .

forecast is 20% lower than the implied volatility and never higher than the implied volatility. This results in 29 trades for option 1 and 45 trades for option 2.

As a result we would expect that if our model predicts that the realized volatility is lower than implied in the option price, that trade would yield a profit. A loss would arise when the realized volatility turns out to be higher than the implied volatility.

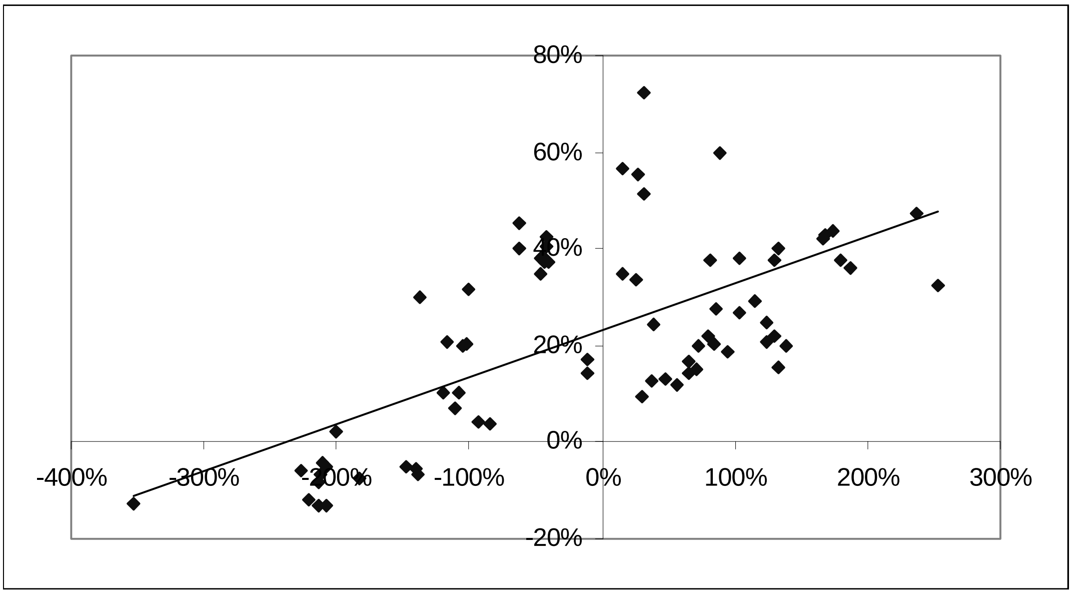


Figure 6: Scatter plot of a trade's return and the deviation between implied and realized volatility. The x-axis depicts the return in percentage terms and the y-axis the percentage difference between the implied volatility and the realized volatility. Since all the trades depicted in this figure are sales of a call option that the EGARCH model regards as overpriced, the more positive the difference between the implied and the realized volatility, the higher the expected return and vice versa.

Figure 6 shows that our data supports this expectation since most points are in the first or third quadrant, while the deviations are caused by transaction costs and market imperfections. It also shows that not all trades generate a positive return. In fact the average return is negative (-9.1%), although it is very small and partly caused by the transaction costs. More interestingly, for 29 trades of option 1 the average return is positive (0.89%), while 45 trades of option 2 generate an average return of -18.9%. This result is due to the low persistence, causing the EGARCH model to under-predict volatility more for option 2 than for option 1.

The conclusion from this trading strategy is that time series models are not better at predicting volatility than the implied volatility, although admittedly this trading strategy is hardly representative.

## 7. Conclusions

This paper has compared two basic approaches to forecast volatility in the German stock market. The first approach uses various univariate time series techniques while the second approach makes use of volatility implied in option prices. The time series models include the historical mean model, the EWMA model, four ARCH-type models and a SV model. Based on the utilization of volatility forecasts in option pricing, we choose the forecast horizons of 45 calendar days and 180 trading days and the error measurement of MSPE. Based on the utilization of volatility forecasts in VaR, we choose the forecast horizons of one and ten trading days and the error measurements of bounded violations and the LINEX loss function. The results suggest that the rankings are sensitive to error measurements as well as forecast horizons. Consistent with the evidence found in BF, the findings indicate that no single method is the clear winner. However, when option pricing is the primary interest, the SV model and implied volatility should be used. On the other hand, when VaR is the concern, the ARCH-type models are useful. Furthermore, a trading strategy suggests that time series models are not better than implied volatilities at predicting volatility, consistent with the findings of Blair, Poon, and Taylor (2000) and Jorion (1995).

## References

Andersen, T.G., and T. Bollerslev, 1998, “Answering the skeptics: yes, standard volatility models do provide accurate forecasts”, International Economic Review, 39, 4, 885-905.

Andersen, T.G., and B. Sorensen, "GMM estimation of a stochastic volatility model: A Monte Carlo study", Journal of Business and Economic Statistics, 14, 1994, pp. 329-352.

Baillie, R., and T. Bollerslev, 1992, “Prediction in dynamic models with time-dependent conditional variances”, Journal of Econometrics, 52, 91-113

Basle Committee On Banking Supervision, 1999, "Performance of models-based capital charges for market risk 1 July-31 December 1998", Basle

Black, F. and M. Scholes, 1973, “The Pricing of Options and Corporate Liabilities”, Journal of Political Economy, 81, 141-183

Blair, B., S-H Poon, and S.J. Taylor, 2000, “Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High Frequency Index Returns”, Journal of Banking and Finance, Forthcoming

Bollerslev, T., 1986, “Generalized autoregressive conditional heteroskedasticity”, Journal of Econometrics 31, 307-327

Brailsford, T.J. and R.W. Faff, 1996, "An evaluation of volatility forecasting techniques", Journal of Banking and Finance 20, 419-438

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

Danielsson, J. 1994, “Stochastic volatility in asset prices: Estimation with simulated maximum likelihood”, Journal of Econometrics 64, 375-400.

Edey, M., and G. Elliot, 1992, “Some evidence on option prices as predictors of volatility”, Oxford Bulletin of Economics and Statistics, 54, 4, 567-578.

Engle, R. F., D. M. Lilien, and R. P. Robins, 1987, “Estimating time varying risk premia in the term structure: The ARCH-M model”, Econometrica 55, 391–407.

Engle, R.F., and K.V. Ng, 1993, “Measuring and testing the impact of news on volatility”, The Journal of Finance, 48(5), 1749-1778

Figlewski, S., 1989, “Options Arbitrage in Imperfect Markets”, Journal of Finance, 44, 1289-1311

Gemmill, G., 1986, “The forecasting performance of stock options on the London Traded Option Markets”, Journal of Business Finance and Accounting, 13,4, 535-546

Geweke, J., 1992, “Bayesian comparison of econometric models”, Working Paper, Federal Reserve Bank of Minneapolis, Minnesota

Gallant, R., and G. Tauchen, 1994, “Which moments to match”, Econometric Theory, 12, 657-681

Glosten, L.R., R. Jagannathan, and D. Runkle, 1993, “On the relation between the expected value and the volatility of the normal excess return on stocks”, Journal of Finance 48, 1779–1801

Harvey, A.C., E. Ruiz, and N. Shephard, 1994, “Multivariate stochastic variance models”, Review of Economic Studies, 61, 247--264

Jacquier, E., N.G. Polson, and P.E. Rossi, 1994, "Bayesian analysis of stochastic volatility models", Journal of Business and Economic Statistics 12, 329-352

Jorion, P., 1995, “Predicting volatility in the foreign exchange market”, Journal of Finance, 50,2, 507-528

Kim S., N. Shephard, and S. Chib, 1998, “Stochastic volatility: likelihood inference and comparison with ARCH models”, Review of Economic Studies, 65, 361—393.

Knight J.L., S.S. Satchell, and J. Yu, 1998, "Efficient estimation of the stochastic volatility model by the empirical characteristic function", Working paper, The University of Western Ontario.

Meyer, R., and J. Yu, 2000, “BUGS for a Bayesian analysis of stochastic volatility models”, The Econometrics Journal, 3(2), 198-215.

Nelson, D. B., 1990, "Conditional heteroskedasticity in asset returns: A new approach", Econometrica 59, 347-370.

Poon, S., and C.W.R. Granger, 2000, “Forecast financial market volatility: A review”, University of Lancaster, Working paper.

Sandmann, G., and S.J. Koopman, 1998, “Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood”. Journal of Econometrics 87, 271-301

Taylor, S.J., 1986, "Modeling Financial Time Series", Chichester: John Wiley.

Vasilellis, G.A., and N. Meade, 1996, “Forecasting volatility for portfolio selection”, Journal of Business Finance and Accounting, 23, 1, 125-143

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

Yu, J., 1999, "Forecasting volatility in the New Zealand stock market", Applied Financial Economics, Forthcoming