The profitability of trading volatility using real-valued and symbolic models

Christian Schittenkopf[1], Peter Tino
Austrian Research Institute for Artificial Intelligence,
Schottengasse 3, 1010 Vienna, Austria
chris@ai.univie.ac.at,petert@ai.univie.ac.at

Georg Dorffner
Dept. of Medical Cybernetics and Artificial Intelligence, University of Vienna,
Freyung 6, 1010 Vienna, Austria
georg@ai.univie.ac.at

1 Overview

Essentially, there are two notions of volatility in literature: historical volatility and implied volatility. While measures of the former notion are derived from historical returns by (weighted) averaging over a time window, measures of the latter are estimated from observed option prices. Whatever particular volatility measure one is willing to apply, a central question is that of *predictability of volatility*[2]. In particular, predictability in a statistical sense and economically meaningful predictability must be distinguished.

In this paper we concentrate on the latter by analyzing the profitability of a pure volatility trading strategy which is delta-neutral and independent of an option pricing model, for the German stock index DAX. Several very different methods ranging from linear and non-linear, real-valued models to symbolic models of volatility changes[3] are applied to predict the change in volatility to the next trading day and to gain profits by buying or selling straddles accordingly. The trading performance is evaluated for one historical and one implied volatility measure. The results are carefully evaluated concerning transaction costs, stationarity issues, and statistical significance. The main contribution of this paper is that, for the first time, the trading performance of models based on different modelling paradigms (real-valued versus symbolic) is compared. Furthermore, it is shown that the combination of different models can lead to improved performance, i.e., higher profits.

In literature related work has been published in [1, 2, 3]. Harvey and Whaley [1] concentrate on implied volatilities of the S&P 100 which are predicted by multivariate models. They find statistical but not economically significant predictability. Noh, Engle and Kane [2] report that abnormal profits can be achieved by trading straddles on the S&P 500 if volatilities are predicted by a GARCH model. For a regression model of implied volatility, the trading strategy loses money (after considering transaction costs). The work of Schmitt and Kaehler [3] is related to ours in that a trading strategy using straddles on the DAX is implemented. They report that profits are significant for market makers using historical volatility models but not for those using implied volatility models .

[1] Acknowledgements: This work was supported by the Austrian Science Fund (FWF) within the research project "Adaptive Information Systems and Modelling in Economics and Management Science" (SFB 010). The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Science and Transport. The authors want thank R. Tompkins for valuable discussions.

[2] There is yet another approach where historical returns are described by heteroskedastic time series models (e.g. by a GARCH model). In this case the volatility predictions are determined by the model specification.

[3] Symbolic sequences of volatility changes are obtained by quantizing volatility differences in a preprocessing step.

2 Volatility measures and trading strategy

Our data set consists of daily closing values of the DAX and daily closing prices of call and put options on the DAX from August 22, 1991 until June 9, 1998 which covers a period of 1700 trading days. In particular, the prices of the first in-the-money and the first out-of-the-money call and put option are available[4]. As a historical volatility measure we choose the popular approach where exponentially declining weights are given to past volatilities, approximated by squared returns[5]. The implied volatility measure is the average of the implied volatilities of the in-the-money and the out-of-the-money call and put options. Both volatility series are differenced resulting in two series of daily volatility changes.

The basic trading strategy is to buy (sell) at-the-money straddles whenever volatility is predicted to increase (decrease). Since at-the-money straddles are approximately delta-neutral, there is no need to delta-hedge, and the strategy is thus a pure volatility trading strategy. In detail, every day straddles worth DM 100 are bought (sold) and on the next day, they are sold (bought). For some models, it is possible that the model is uncertain about the sign of the volatility change. In this case the amount of DM 100 is put into the bank at an annualized interest rate of 4%. For the moment we forget about transaction costs and take the average profit per day obtained by this strategy as the performance measure.

3 Volatility models

The volatility models analyzed in this paper can be divided into three classes: The *real-valued* models are fed with the volatility changes over the last days, and they are optimized to predict the next (real-valued) change in volatility by minimizing the mean squared error function. More precisely, linear autoregressive models (AR) and non-linear neural networks (NN), i.e., multi-layer perceptrons, are estimated. For the *symbolic* models, the series of volatility changes are first transformed into binary sequences by substituting a positive value (an increase in volatility) by 1 and a negative value (a decrease of volatility) by 0. Then symbolic Markov models (MM) of different order are estimated by minimizing the prediction error. The *mixed* models (MIX) are also neural networks with real-valued inputs but they are trained[6] to predict 1 (0) whenever volatility increases (decreases).

Since the data cover a period of nearly seven years, it is reasonable to apply a sliding window technique. The whole period is divided into 212 setups where each setup consists of a training set of length 500 (roughly two years), a validation set of length 125 (roughly six months), and a test set of length five (one week). Each model is estimated on the training set using the last, the last three, the last six, and the last ten volatility changes as inputs. The best of these four models with respect to profit is selected on the validation set, and finally the profit of this best model on the test set is determined. Then the time window is shifted by 5 days, the four models are reestimated etc. For the two types of neural networks, the training and the validation procedure is repeated ten times with randomly initialized networks. Then the volatility changes predicted by the ten networks are combined by majority vote[7]. All together, a test set of 1060 trading days is obtained.

[4] The at-the-money point is assumed to be the closing value of the DAX.

[5] as in RiskMetrics^{TM} (with an exponent of 0.9)

[6] by minimization of the logistic error function

[7] If at least six networks predict an increase (decrease) of volatility, the majority vote is 'increase' ('decrease') and straddles are bought (sold). If five networks predict an increase and the other five a decrease, the money is put into the bank (no trading).

Volatility Model	Historical			Implied		
	Mean	Std.	HTC	Mean	Std.	HTC
True	1.82	4.25	0.85	1.32	4.40	0.58
Random	0.10	4.41	0.05	0.10	4.41	0.05
GARCH	0.10	4.38	-	0.10	4.38	-
Simple	0.41	4.42	0.13	0.38	4.40	0.13
AR	-0.02	4.42	-	-0.28	4.28	-
NN	0.15	4.31	0.01	-0.02	4.23	-
MIX	0.11	4.14	-	0.12	4.25	-
MM	0.29	4.33	0.06	0.27	4.31	0.09

Table 1: Trading performance for different models of historical and implied volatility of the DAX. See text for detailed explanations.

4 Results and discussion

The estimation and simulation results are summarized in Table 1 where the performance of the historical (implied) volatility models is reported on the left-hand (right-hand) side. The mean and the standard deviation of daily profits are reported in the two left-most columns. In the first line of Table 1 the performance of the "true" model, i.e., knowing for all days whether volatility increases or decreases, is given. Therefore the mean profits in the first line provide an upper bound of the profits one can achieve by any volatility predicting model. As a result, using historical volatility predictions is theoretically more profitable than using implied volatility predictions (mean daily profit: 1.82% versus 1.32%). The standard deviations of the profits are 4.25% and 4.40%, respectively. Therefore a t-test[8] would reject the null hypothesis of a zero mean profit at any reasonable significance level. However, the distribution of daily profits is fat-tailed violating the assumption of a normal distribution underlying the t-test.

In our framework the influence of transaction costs can be easily evaluated. Since every day straddles worth DM 100 are bought (sold), transaction costs of, for example, 0.1% for buying or selling a straddle are taken into account by subtracting 0.1 (twice) from the average daily profit[9]. In order to get reliable statistics under transaction costs daily profits are averaged over non-overlapping windows of 62 days (roughly three months) in a first step. Thereby 17 three month performance measures are obtained which are assumed to be independent[10] and normally distributed. Then the highest transaction costs (HTC) in percent for buying (selling) a straddle are calculated so that a t-test, which is now appropriate, rejects the null hypothesis of a zero mean profit at the significance level of 0.05. The HTC for each volatility model[11] are reported in the last column. The "true" model naturally exhibits high values (0.85% for historical and 0.58% for implied volatilities).

In the second to fourth line of Table 1 the performance of three benchmark models is reported. The random model is the model where a coin is flipped every day and the outcome (head or tail) is interpreted as the change in volatility (increasing

[8] This is the standard test applied in empirical studies [1, 2, 3].

[9] Strictly speaking, this method underestimates profits since proceeds from selling straddles (plus the basic amount of DM 100) are not put into the bank overnight. Furthermore, the basic amount is also subject to transaction costs if straddles are sold. Finally, straddles of the same maturity and strike price are sometimes bought and sold on several consecutive days instead of holding the straddles over this period.

[10] Since the corresponding training and validation data sets are overlapping, some dependencies could be introduced. However, there is no significant serial correlation on the series of daily profits.

[11] If the HTC are smaller than 0.01 or the average profit is negative, the table entry is $-$.

or decreasing). The numbers reported in the second line are averages over 1000 simulations[12]. In the next line the performance of a standard return series model, namely a GARCH model, is summarized[13]. A symbolic benchmark model, denoted 'Simple' in Table 1, is specified in the following way: The four basic strategies 'Always buy', 'Always sell', 'If volatility increased (decreased), buy (sell)', and 'If volatility decreased (increased), buy (sell)' (straddles) are applied on the validation set and the most profitable strategy is used on the test set. The performance of the 'AR', 'NN', 'MIX', and 'MM' models is reported in lines 5 to 8 of Table 1.

The following main results are obtained: As for the 'true' model, the models of historical volatility produce larger profits (smaller losses) than the models of implied volatility which confirms the findings in [2, 3]. In other words, volatilities reflected in option prices do not seem to fully utilize historical information. For the real-valued and the mixed models, it is really hard to beat the random model. The mean daily profits of some models are larger than 0.10% but a paired t-test (on the three month performance measures) reveals that these additional profits are by far not statistically significant. The 'simple' model produces the highest profits of all models which are, however, not significantly larger than the profits of the random model either ($p > 0.07$). For the other symbolic model, i.e., the Markov model, the second highest profits are obtained. Consequently, the transformation of real-valued volatility changes into a sequence of ups and downs seems to reduce the noise in the volatility series and thereby facilitates the extraction of information relevant for profitable trading.

One is now tempted to ask why the 'simple' model produces higher profits than the Markov model which is more general. A closer analysis of the trading performance shows that stationarity of the data plays a crucial role. Whenever the market over the actual time window of two years is stationary such that sophisticated models can be estimated, the Markov model dominates the 'simple' model. However, if markets are unstable it is more appropriate to choose a strategy from a set of simple strategies which are selected with respect to trading performance on the validation set.

As a consequence, we started to implement *combined models* where, depending on the performance on the validation set, the 'simple' model or a Markov model of higher order (estimated on the training set) is selected. First results show that stationarity problems can be eliminated by this mechanism which leads to improved trading performance. For instance, a combined model of historical volatility generates mean daily profits of 0.66%, which is significantly better than the random model ($p < 0.001$), with HTC of 0.27%.

References

[1] Harvey, C.R., and R.E. Whaley, 1992, Market volatility prediction and the efficiency of the S&P 100 index option market, Journal of Financial Economics 31, 43-73.

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

[3] Schmitt, C., and J. Kaehler, 1996, Delta-neutral volatility trading with intraday prices: an application to options on the DAX, Discussion Paper No. 96-25, Zentrum für Europäische Wirtschaftsforschung.

[12] One simulation is one pass through the set of trading days. The entries for historical and implied volatility are of course the same.

[13] For the GARCH models, the entries for historical and implied volatility are the same since the models are trained on return series and not on volatility series.