

Forecasting Time-dependent Conditional Densities: A Semi-non- parametric Neural Network Approach

CHRISTIAN SCHITTENKOPF,^{1*} GEORG DORFFNER¹ and
ENGELBERT J. DOCKNER²

¹ *Austrian Research Institute for Artificial Intelligence, Vienna, Austria*

² *University of Vienna, Austria*

ABSTRACT

In financial econometrics the modelling of asset return series is closely related to the estimation of the corresponding conditional densities. One reason why one is interested in the whole conditional density and not only in the conditional mean is that the conditional variance can be interpreted as a measure of time-dependent volatility of the return series. In fact, the modelling and the prediction of volatility is one of the central topics in asset pricing. In this paper we propose to estimate conditional densities semi-non-parametrically in a neural network framework. Our recurrent mixture density networks realize the basic ideas of prominent GARCH approaches but they are capable of modelling any continuous conditional density also allowing for time-dependent higher-order moments. Our empirical analysis of daily FTSE 100 data demonstrates the importance of distributional assumptions in volatility prediction and shows that the out-of-sample forecasting performance of neural networks slightly dominates those of GARCH models. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS conditional densities; forecasting; GARCH; neural networks; volatility

The concepts of risk and volatility are central to the modern theory of finance in general and to asset pricing in particular. For instance, the price of an option depends on the volatility of the underlying asset. An accurate prediction of future volatility is thus crucial to obtain reasonable option price forecasts. However, there is no generally accepted definition of volatility. While implied volatility is determined directly from market-based option prices together with an option price model, any historical measure relies on a time series model of asset returns. The time series

* Correspondence to: Christian Schittenkopf, Austrian Research Institute for Artificial Intelligence, Schottengasse 3, A-1010 Vienna, Austria. E-mail: chris@ai.univie.ac.at

Contract grant sponsor: Austrian Science Fund.
Contract grant number: SFB 010.

model is in fact a model of the conditional (probability) density (function) where the time-dependent conditional variance is interpreted as the volatility of the underlying asset.

The first models in this direction were the ARCH (Engle, 1982) and the generalized ARCH (GARCH) models (Bollerslev, 1986) which assume a conditional normal distribution. Despite their simplicity, GARCH models are able to capture several stylized facts of asset return series, namely heteroscedasticity, volatility clustering, and excess kurtosis. Meanwhile there exists a large body of literature dealing with extensions of the GARCH approach. One common idea is to include leverage effects into the model, i.e. negative returns are supposed to influence future volatility more strongly than positive returns (Nelson, 1990; Engle and Ng, 1993; Glosten *et al.*, 1993). Recently, neural networks have been proposed to model leverage effects in a non-linear and semi-non-parametric way (Donaldson and Kamstra, 1997). The second direction in which GARCH models have been generalized is the specification of the conditional density.

In Bollerslev (1987) it is suggested to use the standardized Student's- t distribution instead of the normal distribution in order to allow for fat tails in the conditional distribution (besides heteroscedasticity). The kurtosis of the conditional distribution, however, is fixed (not time-dependent). A survey of other conditional distributions which have been proposed in the literature, can be found in Bollerslev, Chou, and Kroner (1992). Besides these parametric density models, semi-parametric models (Engle and González-Rivera, 1991) and non-parametric models (Pagan and Schwert, 1990; Boudoukh, Richardson, and Whitelaw, 1997) have been specified to model return series. A semi-non-parametric framework has been applied in a similar context in Gallant and Tauchen (1989). Despite the large number of extensions proposed in recent years, GARCH models are probably still the most widely used approach to estimate volatility.

In this paper the issue of conditional density modelling and forecasting is addressed from a semi-non-parametric point of view. We present a neural network-based model, a so-called *recurrent mixture density network*, which approximates the conditional distribution by a mixture of Gaussians. The parameters of the mixture, which determine the shape of the conditional distribution, are estimated from the elements of the time-dependent information set. Therefore *the conditional density is itself time-dependent* including higher-order moments such as conditional skewness and conditional kurtosis. This extension of the concept of heteroscedasticity to higher-order moments is an attractive feature of our model. Indeed, it is rather unrealistic to expect that a single parametric specification of the conditional density is suitable for all kinds of return series data (Bera and Higgins, 1993). A natural solution is thus to modify the shape of the conditional density in dependence of the data as in our model. In addition, the neural network-based model allows for non-linear dependencies in the conditional mean and in the conditional variance. In the following sections we will concentrate, however, on the distributional aspects of density modelling.

We compare three different density specifications: a standard GARCH model with a normal distribution (heteroscedastic but neither skewed nor leptokurtic), a GARCH model with a Student's- t distribution (heteroscedastic, not skewed but leptokurtic), and a recurrent mixture density network (heteroscedastic, skewed and leptokurtic in a time-dependent manner). The models are evaluated not only with respect to likelihood but also with respect to their volatility forecasting performance. Strictly speaking, the predicted volatility, i.e. the variance of the conditional density predicted by the models is compared to two historical volatility measures and one measure of implied volatility. Generally we find that a proper specification of the conditional density is crucial for the out-of-sample performance of volatility prediction. In that respect the classical GARCH model is dominated by the GARCH- t model and the recurrent mixture density

network. While both the GARCH- t model and the recurrent mixture density network are capable of capturing fat tail elements in the conditional distribution only the recurrent mixture density network allows for time-varying skewness and kurtosis which are common in financial markets.

The paper is organized as follows. The next section recalls traditional GARCH models as well as neural network-based models for conditional density estimation. In the third section the concept of mixture density networks is extended in a recurrent way to allow for GARCH effects. The empirical analysis of the volatility models with special emphasis on the prediction performance is described in detail in the fourth section. The final section presents conclusions.

CONDITIONAL DENSITY MODELLING IN FINANCE

Today it is widely accepted that return series of frequently traded assets such as stocks and stock indices are characterized by several stylized facts such as fat tails in the unconditional distribution of returns (see, for example, Mandelbrot, 1963; Fama, 1965) and the observation that returns are usually uncorrelated. However, large returns are commonly clustered which means that there is a tendency that large price changes are followed by other large price changes. This phenomenon of volatility clustering has been first modelled by Engle (1982) and Bollerslev (1986) introducing ARCH and GARCH models.

A simple return series model capturing the stylized facts to some extent is given by an autoregressive model of first order (AR(1)) for the conditional mean and a GARCH specification for the conditional variance. More precisely, the conditional density of the next return r_{t+1} is given by

$$\rho(r_{t+1}|I_t) = k(\mu_{t+1}, \sigma_{t+1}^2) \quad (1)$$

$$\mu_{t+1} = a_0 + a_1 r_t \quad (2)$$

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 e_t^2 + \beta_1 \sigma_t^2 \quad (3)$$

$$e_t = r_t - \mu_t \quad (4)$$

where I_t denotes the information available at time t , and $k(\mu_{t+1}, \sigma_{t+1}^2)$ is the density of a normal distribution of mean μ_{t+1} and variance σ_{t+1}^2 . e_t is the prediction error at time t . Stationarity of this model is guaranteed if $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, and $\alpha_1 + \beta_1 < 1$ hold (besides $|a_1| < 1$). In contrast to the time-dependence of the variance, the skewness and the kurtosis of the conditional distribution are constant.¹

Bollerslev later proposed to substitute the normal conditional distribution by a Student's- t distribution with ν degrees of freedom in order to allow for kurtosis in the conditional distribution that goes beyond that driven by heteroscedasticity (Bollerslev, 1987). The conditional density $\rho(r_{t+1}|I_t)$ of a GARCH- t model is given by

¹ Since the conditional distribution is normal, they are 0 and 3, respectively.

$$t_v(\mu_{t+1}, \sigma_{t+1}^2) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} (\pi(v-2) \sigma_{t+1}^2)^{-\frac{1}{2}} \left(1 + \frac{(r_{t+1} - \mu_{t+1})^2}{(v-2)\sigma_{t+1}^2}\right)^{-\frac{v+1}{2}} \quad (5)$$

with $v > 2$). For $v \rightarrow \infty$, the density of equation (5) converges to the normal density $k(\mu_{t+1}, \sigma_{t+1}^2)$. Therefore GARCH- t models are more general than GARCH models. The degrees of freedom v are an additional parameter which determines, among other characteristics, the kurtosis of the conditional distribution.² Since the conditional density in equation (5) is symmetric, the conditional skewness is 0. In particular, the skewness and the kurtosis of the conditional distribution are again not time-dependent.

The number of publications dealing with extensions of the GARCH model is very large. For a comprehensive overview the reader is referred to Bollerslev *et al.* (1992) and Bera and Higgins (1993). Many extensions concentrate on the specification of the conditional variance. Our focus, however, is on the time-dependent shape of the conditional density.

The model which is introduced as a generalization of the GARCH models is a new neural network-based model,³ or more precisely, it is a recurrent extension of a so-called mixture density network (MDN; Bishop, 1995; Neuneier *et al.*, 1994; Ormoneit, 1998). MDNs are able to represent conditional densities of non-constant variance (thereby allowing for heteroscedastic dependencies in the data) and they are able to approximate any continuous density (thereby allowing for skewness and fat tails) to arbitrary accuracy (McLachlan and Basford, 1988). This concept has been applied to stock index data from the German (Ormoneit and Neuneier, 1996⁴) and the Austrian market (Schittenkopf, Dorffner, and Dockner, 1998). In the latter paper, MDNs are compared to traditional volatility models on the basis of daily returns. While MDNs show a better performance than ARCH models, the results of the comparison of MDNs and GARCH models suggest that the recursive specification of the latter is essential in order to model long-term dependencies in volatility. This conjecture naturally leads to neural network models with recurrent dynamics.

RECURRENT MIXTURE DENSITY NETWORKS

The main idea of the model is to predict the parameters of the conditional density in dependence of the data. These parameters are the priors, the centres, and the widths of a weighted sum of normal densities (mixture of Gaussians). More precisely, a *recurrent mixture density network* with n normal densities (RMDN(n)) approximates the conditional density by

$$\rho(r_{t+1}|I_t) = \sum_{i=1}^n \pi_{i,t+1} k(\mu_{i,t+1}, \sigma_{i,t+1}^2) \quad (6)$$

where the parameters $\pi_{i,t+1}$, $\mu_{i,t+1}$, and $\sigma_{i,t+1}^2$ of the n Gaussian components are estimated by

² For $v > 4$, the conditional kurtosis is given by $3(v-2)/(v-4)$.

³ See Refenes (1995) and Adya and Collopy (1998) for neural network applications in economics and finance.

⁴ There MDNs are used to predict intraday volatility.

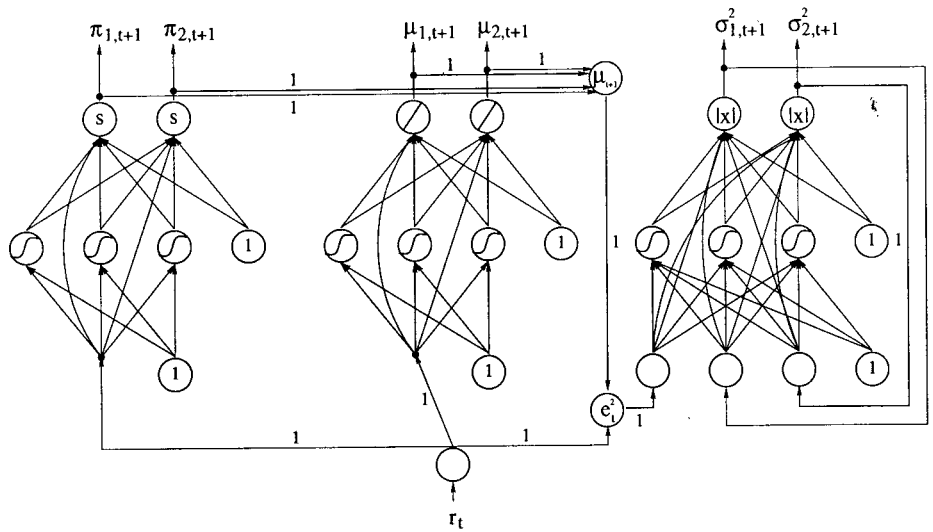


Figure 1. A recurrent mixture density network with two Gaussian components

different multi-layer perceptrons (MLPs⁵). For the priors, the softmax function $s(\tilde{\pi}_{i,t})$ is applied to the MLP outputs to ensure that the priors $\pi_{i,t}$ are positive and that they sum up to 1 which makes the right-hand side of equation (6) a density function:

$$\pi_{i,t+1} = s(\tilde{\pi}_{i,t+1}) = \frac{\exp(\tilde{\pi}_{i,t+1})}{\sum_{j=1}^n \exp(\tilde{\pi}_{j,t+1})} \tag{7}$$

$$\tilde{\pi}_{i,t+1} = \text{MLP1}_i(r_t) \tag{8}$$

The outputs of the second MLP are the centres of the Gaussian components which are not restricted: $\mu_{i,t+1} = \text{MLP2}_i(r_t)$. Finally, as a generalization of the GARCH models, the MLP estimating the conditional variances is recurrent and it has $n + 1$ inputs. One input is the squared error e_t^2 of the previous time step and the other n inputs are the variances $\sigma_{k,t}^2$, $k = 1, \dots, n$, of the Gaussian components of the previous time step. The mapping of this MLP⁶ is given by

$$\text{MLP3}_i(e_t^2, \sigma_{1,t}^2, \dots, \sigma_{n,t}^2) = \sum_{j=1}^H v_{ij} \tanh\left(w_{j0}e_t^2 + \sum_{k=1}^n w_{jk}\sigma_{k,t}^2 + c_j\right) + s_{i0}e_t^2 + \sum_{k=1}^n s_{ik}\sigma_{k,t}^2 + b_i \tag{9}$$

The variances $\sigma_{i,t+1}^2$ are obtained by taking absolute values of the MLP outputs which avoids negative values. An RMDN with two Gaussian components ($n = 2$) and three hidden units (for each MLP) is depicted in Figure 1.⁷

⁵The basic MLP maps the one-dimensional input r_t onto an n -dimensional output. The i th component is given by $\text{MLP}_i(r_t) = \sum_{j=1}^H v_{ij} \tanh(w_j r_t + c_j) + s_i r_t + b_i$, $1 \leq i \leq n$, where H denotes the number of hidden units, w_j and v_{ij} the weights of the first and second layers, c_j and b_i the biases of the first and second layer, and s_i the shortcut connection from the input to the i th output component.

⁶Since there are several inputs, the weight vector w_j and the shortcut vector s_i are replaced by a weight matrix w_{jk} and a matrix of shortcuts s_{jk} .

⁷Weights which are fixed are set to 1. The calculation of the conditional mean μ_{t+1} , which is fed back as μ_t in the next time step, and the calculation of e_t^2 is indicated by extra units.

For the described model, the conditional mean μ_{t+1} and the conditional variance σ_{t+1}^2 are given by

$$\mu_{t+1} = \sum_{i=1}^n \pi_{i,t+1} \mu_{i,t+1} \quad (10)$$

$$\sigma_{t+1}^2 = \sum_{i=1}^n \pi_{i,t+1} (\sigma_{i,t+1}^2 + (\mu_{i,t+1} - \mu_{t+1})^2) \quad (11)$$

For a model with only one Gaussian component ($n = 1$), these equations are reduced to trivial identities because $\pi_{1,t+1} = 1$. In this case an RMDN is reduced to a non-linear GARCH model where the autoregressive specifications for the conditional mean and the conditional variance of a standard GARCH model are replaced by MLPs. In general ($n \geq 2$), however, the shape of the conditional density depends on the particular mixture of Gaussians the parameters of which are themselves functions of the elements of the current information set I_t . In particular this means that higher-order properties of the conditional distribution such as skewness and kurtosis are also time-dependent. The exact expressions for the conditional skewness s_t and the conditional kurtosis k_t are obtained as

$$s_{t+1} = \frac{1}{\sigma_{t+1}^3} \sum_{i=1}^n \pi_{i,t+1} (3\sigma_{i,t+1}^2 (\mu_{i,t+1} - \mu_{t+1}) + (\mu_{i,t+1} - \mu_{t+1})^3) \quad (12)$$

$$k_{t+1} = \frac{1}{\sigma_{t+1}^4} \sum_{i=1}^n \pi_{i,t+1} (3\sigma_{i,t+1}^4 + 6\sigma_{i,t+1}^2 (\mu_{i,t+1} - \mu_{t+1})^2 + (\mu_{i,t+1} - \mu_{t+1})^4) \quad (13)$$

This time-dependence of higher-order moments is an appealing feature of RMDNs, and it is in contrast to the properties of GARCH and GARCH- t models. We recall that for the latter models, although conditional leptokurtic, k_{t+1} is constant. As noted in Bera and Higgins (1993), no single specification of the conditional density appears to be suitable for all conditional heteroscedastic data. In this sense, semi-non-parametric RMDNs represent a promising candidate for modelling different types of time-dependent conditional distributions.⁸

EMPIRICAL ANALYSIS

In this section we first describe the data sets which are used to estimate three different volatility models, namely the GARCH model, the GARCH- t model and the RMDN(2) model as defined above. Then several error measures which quantify different aspects of the performance of the models, are described. Finally, the out-of-sample modelling and prediction performance of the volatility models are analysed.

⁸ Another interesting neural network-based approach has been presented in Ormoneit (1998). The main concept is similar to the one formulated in Gallant and Tauchen (1989), namely to approximate the conditional density by a series of Hermite polynomials (Gram-Charlier expansion). The parameters of this approximation are predicted by so-called recurrent conditional density estimation networks.

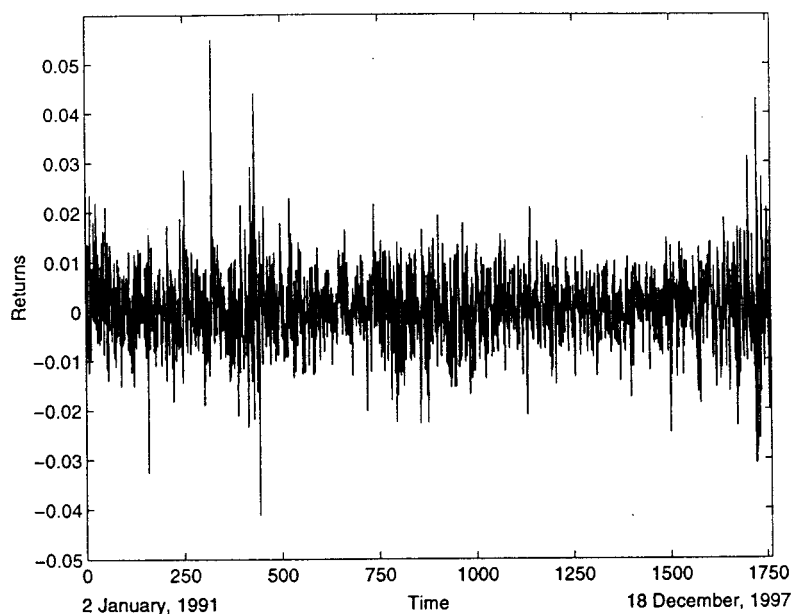


Figure 2. The time series of daily returns of the FTSE 100 from 2 January 1991 to 18 December 1997

FTSE 100 data set

The data set used in the empirical analysis is a series of daily closing values of the FTSE 100 and of daily closing prices of call and put options on this index with different maturities and exercise prices. The series covers the 1762 trading days from 2 January 1991 to 18 December 1997. Daily closing values s_t of the FTSE 100 are transformed into continuously compounded returns r_t by

$$r_t = 100 \log \frac{s_{t+1}}{s_t} \quad (14)$$

The time series of daily returns is depicted in Figure 2.

In order to take care of stationarity issues and to put the empirical analysis on a broader basis, the time series is divided into five overlapping segments of three years. The first segment covers the years 1991–1993, the second the period 1992–1994 and so on. The first two years of each segment are used to estimate the parameters of the volatility models whereas the third years obtained from the segments form independent test sets. In particular, the test sets are not overlapping. By this procedure, which is similar to cross-validation, the reliability of the results is increased in comparison to experimental studies on a single data set. Some basic statistics of daily FTSE 100 returns are summarized for each year in Table I. In particular, the mean, the standard deviation, the skewness, and the kurtosis of the unconditional distribution of returns are reported.

Error measures

The commonly used procedure for estimating parameters of heteroscedastic models is the maximum likelihood approach. In this paper the average negative loglikelihood of the sample

Table I. Basic statistics of daily FTSE 100 returns from 1991 to 1997

Year	Mean ($\times 10^{-3}$)	Std. ($\times 10^{-3}$)	Skewness	Kurtosis
1991	0.621	8.305	0.162	3.705
1992	0.529	9.866	0.872	8.537
1993	0.705	6.305	0.230	3.832
1994	-0.415	8.499	-0.192	2.466
1995	0.740	6.217	-0.075	3.386
1996	0.434	5.862	-0.493	3.725
1997	0.982	9.944	-0.081	4.809

(apart from some initial condition) which is denoted \mathcal{L} , is minimized. \mathcal{L} will also be called the *loss function* of a data set since \mathcal{L} can be calculated for data sets which were not used to estimate the model parameters, as well. In other words, the loss function can also be evaluated out-of-sample.

In addition to the loss function, alternative error measures are applied to analyse the performance of the models. These measures require the definition of a 'true' volatility of the underlying return series which is problematic since volatility is not directly observable. As a first rough approximation, the squared daily returns are taken as a measure of 'true' volatility. Due to the shortcomings of this approach,⁹ we turn to the notion of realized volatility which has recently been proposed in literature (Andersen *et al.*, 1999). In this framework daily volatility is approximated by summing high-frequency intraday squared returns which leads to an essentially model-free estimate of 'true' volatility. The data set available for this study consists of intraday quotes of bid-ask prices of options on the FTSE 100 which include the value of the FTSE 100 at that time. From this high-frequency data set the value of the FTSE 100 is extracted every five minutes between 8:30 am and 4:30 pm which gives a series $\tilde{s}_0, \tilde{s}_1, \dots, \tilde{s}_{96}$. Following Andersen *et al.* (1999), the variance over the 5-minute intervals is approximated by calculating 5-minute squared returns $\tilde{r}_t = (\log \tilde{s}_t / \tilde{s}_{t-1})^2$, $t = 1, \dots, 96$. Additionally, the overnight change of the index must be taken into account. At this point the question arises how the squared overnight return should be scaled or, in other words, whether time outside the trading hours is the same as time during the trading hours. This item will be discussed more thoroughly in a later section. For the moment, the overnight period of 16 hours is taken as 192 independent 5-minute intervals, and the realized volatility of the FTSE 100 is estimated by weighting the squared overnight return \tilde{r}_0^2 and the intraday 5-minute squared returns \tilde{r}_t^2 such that an unbiased estimator of daily volatility is obtained:

$$\sigma_t^2 = \frac{288}{97} \left(\frac{1}{192} \tilde{r}_0^2 + \sum_{t=1}^{96} \tilde{r}_t^2 \right) \quad (15)$$

Denoting the 'true' volatility (of the next time step) σ_{t+1}^2 and the corresponding conditional variance predicted by a model¹⁰ $\hat{\sigma}_{t+1}^2$, the following two error measures are calculated for both measures of 'true' volatility:

⁹Squared returns are considered to be an unreliable measure since they are contaminated by substantial measurement error (Andersen and Bollerslev, 1998).

¹⁰For the RMDN(2) models, the accumulated conditional variance defined in equation (11) is inserted.

$$\text{NMAE} = \frac{\sum_{t=1}^N |\sigma_{t+1}^2 - \hat{\sigma}_{t+1}^2|}{\sum_{t=1}^N |\sigma_{t+1}^2 - \sigma_t^2|} \quad (16)$$

$$\text{HR} = \frac{1}{N} \sum_{t=1}^N \theta_t \quad (17)$$

$$\theta_t = \begin{cases} 1: & (\hat{\sigma}_{t+1}^2 - \sigma_t^2)(\sigma_{t+1}^2 - \sigma_t^2) \geq 0 \\ 0: & \text{else} \end{cases} \quad (18)$$

Where N denotes the sample size the *normalized mean absolute error* (NMAE) relates the mean absolute error

$$\text{MAE} = 1/N \sum_{t=1}^N |\sigma_{t+1}^2 - \hat{\sigma}_{t+1}^2|$$

of a volatility model to the MAE of the naive model $\hat{\sigma}_{t+1}^2 = \sigma_t^2$ which takes the most recent volatility as the prediction of the volatility of the next time step. The naive model thus serves as a benchmark model which, of course, should be beaten. In this case the NMAE is smaller than 1. The minimum value is 0. The *hit rate* (HR) is the relative frequency of correctly predicted increases and decreases of volatility, i.e. it measures how often the model gives the correct direction of change of volatility.¹¹ The HR lies between 0 and 1. A value of 0.5 indicates that the model is not better than a random predictor generating a random sequence of ups and downs (provided that ups and downs are equally likely).

If one defines a measure of 'true' volatility, it is natural to apply also standard time series models to this defined volatility series and to calculate the performance measures NMAE and HR. In particular, autoregressive (AR) models without and with moving average (MA) terms are standard models which can be estimated from the 'true' volatility series over the time periods covered by the training sets and which can then be evaluated on the test sets. Since the squared returns are considered to be contaminated by measurement errors, it is not expected to find pronounced dependencies in these series. For the realized volatilities, however, it is a priori not clear how much structure the series contain.

An alternative notion of volatility is provided by the concept of implied volatility which completely neglects time series properties of the asset returns.¹² On the one hand, the concept of implied volatility is attractive since it is a measure of the volatility expected by the market; on the other, it requires a realistic option pricing model. It should be noted that controversial results have been reported on the statistical dependencies between implied volatility and (future) 'true' volatility (see, for example, Latané and Rendleman, 1976; Canina and Figlewski, 1993; Lamoureux and Lastrapes, 1993).

For the FTSE 100 data set, the average implied volatility of near at-the-money call and put

¹¹ This information is essential for volatility trading strategies as described, for instance, in Tompkins (1994).

¹² The implied volatility of an option is calculated on the basis of the current option price and the current price of the underlying asset. To be specific, it is the volatility for which the theoretical option price (for a given option pricing model) coincides with the current option price.

options¹³ is calculated for each trading day as a measure of the volatility currently anticipated by the market¹⁴. One way to quantify the prediction performance of a volatility model is to calculate some distance measure, say the MAE, between implied and predicted volatilities¹⁵. The drawback of this approach is the implicit assumption that the Black-Scholes model describes the option pricing by the market adequately. Indeed, since the analyzed time series are heteroskedastic, i.e., volatility is time-varying, the application of the Black-Scholes model might be problematic. As an equivalent approach with the same problem, though, one can calculate option prices from the predicted volatilities and compare them to real-world option prices, for instance, by implementing a trading strategy (Noh *et al.*, 1994, Schmitt and Kaehler, 1996). In fact, earlier experiments with the trading strategy in (Noh *et al.*, 1994) on a set of DAX options revealed that options were systematically over- or underpriced by all volatility models over different periods of time (of up to several months). Over these periods, however, the shape of the predicted price curve (the predicted volatility curve) was similar to the true price curve (the implied volatility curve) for all models. We thus think that it is a reasonable approach not to rely on the Black-Scholes model directly but to measure the *correlation* between predicted and implied volatilities. In other words, to the extent that the implied volatility increases (decreases), the model should predict an increase (decrease) of volatility relative to the current value.

As two market-oriented performance measures, the linear correlation coefficient, i.e. Pearson's r , and Spearman's rank order correlation coefficient r_s of implied volatilities $\sigma_{IV,t+1}^2$ and predicted volatilities $\hat{\sigma}_{t+1}^2$ are calculated. The rank order correlation coefficient is defined as the linear correlation coefficient where the values are replaced by their ranks (Press *et al.*, 1992). This non-parametric correlation measure is considered more robust than linear correlation. Even more importantly, r_s remains unchanged if both time series are distorted by arbitrary, increasing functions, which do not change the ranks of the series. In other words, if the 'true' volatility is not the implied volatility but a (non-)linear, increasing function of the implied volatility (which we do not have to know), the rank correlation r_s measures the correlation between the predicted volatility and the 'true' volatility.

We now turn to the methods for making multi-step volatility predictions. For the GARCH model and the GARCH- t model, analytical results can be derived since the specification of the conditional variance is linear. For the non-linear RMDN(2) model, a Monte Carlo simulation technique is applied. Starting with the GARCH models, the first goal is to calculate the expected value of the variance n time steps (days) ahead conditioned on the current information set I_t : $E(\sigma_{t+n}^2|I_t)$. For $n = 1$, the expectation operator vanishes since the deterministic relationship in equation (3) holds. For larger values of n , the expected value $E(\sigma_{t+n}^2|I_t)$ can be calculated by induction as

$$E(\sigma_{t+n}^2|I_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} + \left(\sigma_{t+1}^2 - \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \right) (\alpha_1 + \beta_1)^{n-1} \quad (19)$$

For a stationary GARCH or GARCH- t model satisfying $\alpha_1 + \beta_1 < 1$, the expected value thus converges for $n \rightarrow \infty$ to $\alpha_0/(1 - \alpha_1 - \beta_1)$ which is the unconditional variance of the model. The next

¹³ For each trading day, options maturing the next month are selected. The resulting time to maturity ranges from nine to 52 days.

¹⁴ A traditional Black-Scholes model is used as the option price model (Black and Scholes, 1973).

¹⁵ In this case the predicted volatility must be the volatility predicted for the remaining lifetime of the option (multi-step prediction).

step is to calculate the volatility (conditional variance) per time step (per day) for the remaining lifetime of the straddle. The reason for this step is that the Black–Scholes model, which is used to calculate the implied volatilities, assumes a constant volatility over the remaining lifetime. The expected values of the conditional variances are not constant¹⁶ but they can be substituted by their average value which does not change the total volatility from tomorrow until the day of maturity. Using equation (19), it is easy to derive a closed-form expression for this average volatility per day which is denoted $\hat{\sigma}_{t+1}^2$.

For the non-linear RMDN(2) models, the conditional expected values $E(\sigma_{t+n}^2|I_t)$ cannot be calculated analytically in general. Only for $n = 1$, where the conditional density $\rho(r_{t+1}|I_t)$ is known explicitly,¹⁷ the (accumulated) conditional variance can be determined analytically using equation (11). For larger n , $E(\sigma_{t+1}^2|I_t)$ can be estimated by Monte Carlo simulation. The main item is to estimate the conditional densities $\rho(r_{t+i}|I_t)$, $i = 2, \dots, n$, in a recursive fashion. More precisely, the following procedure is applied. Starting with $i = 1$, random numbers (the ‘returns’ r_{t+1}) are drawn according to the conditional density $\rho(r_{t+1}|I_t)$, and the corresponding ‘prediction errors’ $e_{t+1} = r_{t+1} - \mu_{t+1}$ are calculated.¹⁸ Then the r_{t+1} , the e_{t+1}^2 , and the actual conditional variances are fed as inputs into the trained RMDN in order to obtain (empirical) distributions of the parameters of $\rho(r_{t+i+1}|I_t)$. Additionally, by using equation (11), the distribution of σ_{t+i+1}^2 conditioned on I_t is obtained. By taking expected values, i.e. by calculating the mean values of the empirical distributions, the parameters of the conditional density $\rho(r_{t+i+1}|I_t)$ as well as $E(\sigma_{t+i+1}^2|I_t)$ can be estimated. Now the next step of the recursive procedure is performed using the new parameters and so on. Finally, as for the GARCH models, the mean value of the (accumulated) conditional variances is calculated.¹⁹

Estimation results

A GARCH model, a GARCH- t model, and an RMDN(2) model are fitted to each training set separately by minimizing the loss function \mathcal{L} . The parameter values obtained for the GARCH models and the GARCH- t models are such that the models are stationary with a mean persistence of 0.907 and 0.906, respectively. The degrees of freedom parameter ν of the GARCH- t models indicates substantial fat tails in the conditional distribution of returns for the first and second training set. Leptokurtosis in the conditional distribution has also been reported in Bollerslev (1987) for the S&P 500. For the RMDN(2) models, which are semi-non-parametric, the weights of the MLPs can be hardly interpreted. The number of hidden units was chosen ad hoc as $H = 3$. We did not try to optimize the network performance with respect to the number of hidden units.

In Figure 3 the conditional densities $\rho(r_{t+1}|I_t)$ estimated by the three volatility models are plotted for two specific days of the last test set. The first day is 26 February 1997 during a rather calm period with returns close to 0. The second day is 31 October 1997 after several large decreases of the FTSE 100.

Due to the large return preceding the second day, the conditional variances, i.e. the estimated volatilities, are much larger for the densities depicted on the right-hand side than for those on

¹⁶ Except for the case that σ_{t+1}^2 equals the unconditional variance.

¹⁷ The actual parameters are deterministic functions of the previous return, the previous squared prediction error, and the previous conditional variances.

¹⁸ μ_{t+1} is defined in equation (10).

¹⁹ In our setup where daily closing prices are used, the average volatility per day is calculated for all models as the mean value of the conditional variances from the day after tomorrow until the day of maturity, and it is compared to the implied volatility of tomorrow.

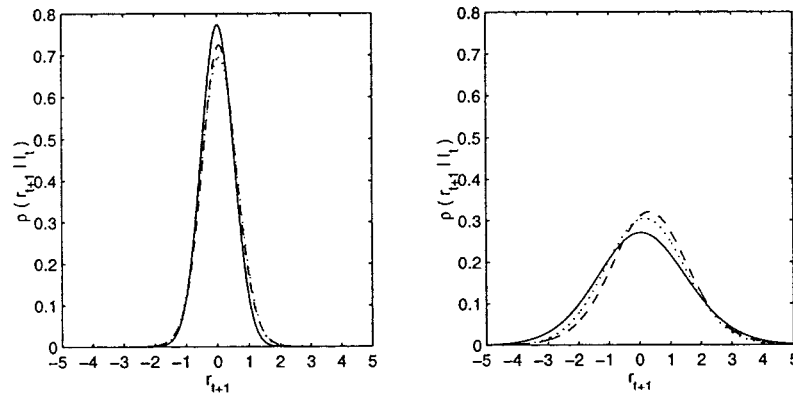


Figure 3. The conditional densities estimated by the models for 26 February 1997 and for 31 October 1997 are depicted on the left- and right-hand sides, respectively (dotted line: GARCH model, dashed line: GARCH- t model, solid line: RMDN(2) model)

the left-hand side. For both days, the two normal densities of the RMDN(2) model are weighted and combined such that the resulting distribution is leptokurtic. Furthermore, the conditional distribution of the RMDN(2) model is negatively skewed on the first day and positively skewed on the second day in contrast to the other models, which have symmetric conditional distributions.

For a better illustration of the differences between the models, the last test set, which covers the year 1997, is plotted together with the mean, the variance, the skewness, and the kurtosis of the conditional densities estimated by the models in Figure 4. The first half of this period is rather tranquil with absolute returns smaller than 2% whereas the second part exhibits larger movements in the returns. The means of the conditional densities are very similar for all models. The conditional variances, however, are only similar for the first 150 days. For the more volatile period towards the end of 1997, the RMDN(2) model exhibits much larger conditional variances than the GARCH and the GARCH- t model. In contrast to the first two conditional moments, the conditional skewness and the conditional kurtosis are time-dependent only for the RMDN(2) model. The corresponding values are 0 and 3 for the GARCH model and 0 and 3.389 for the GARCH- t model. For the RMDN(2) model, the skewness and the kurtosis vary considerably over the whole year with average values of -0.243 and 3.414 , respectively.

It has been emphasized in the literature that the real test of a volatility model is to predict volatility out-of-sample, i.e. on a data set disjoint from the training data (e.g. by Pagan and Schwert, 1990). The ability of a model to forecast volatility out-of-sample is particularly important for models with a large number of parameters (compared to GARCH models) such as the semi-non-parametric RMDN(2) model which can potentially overfit the training data. In an out-of-sample analysis the size of the volatility models can be neglected since the parameters of the models are estimated on a separate set of training data. In this section only out-of-sample results are summarized.

In our analysis, as mentioned above, each training set consists of daily returns over a period of two years. The corresponding test set covers the returns (squared returns, realized volatilities, implied volatilities) of the year following the training period. In particular, the performance on a future data set (relative to the training set) is analysed. In order to prevent the RMDN(2)

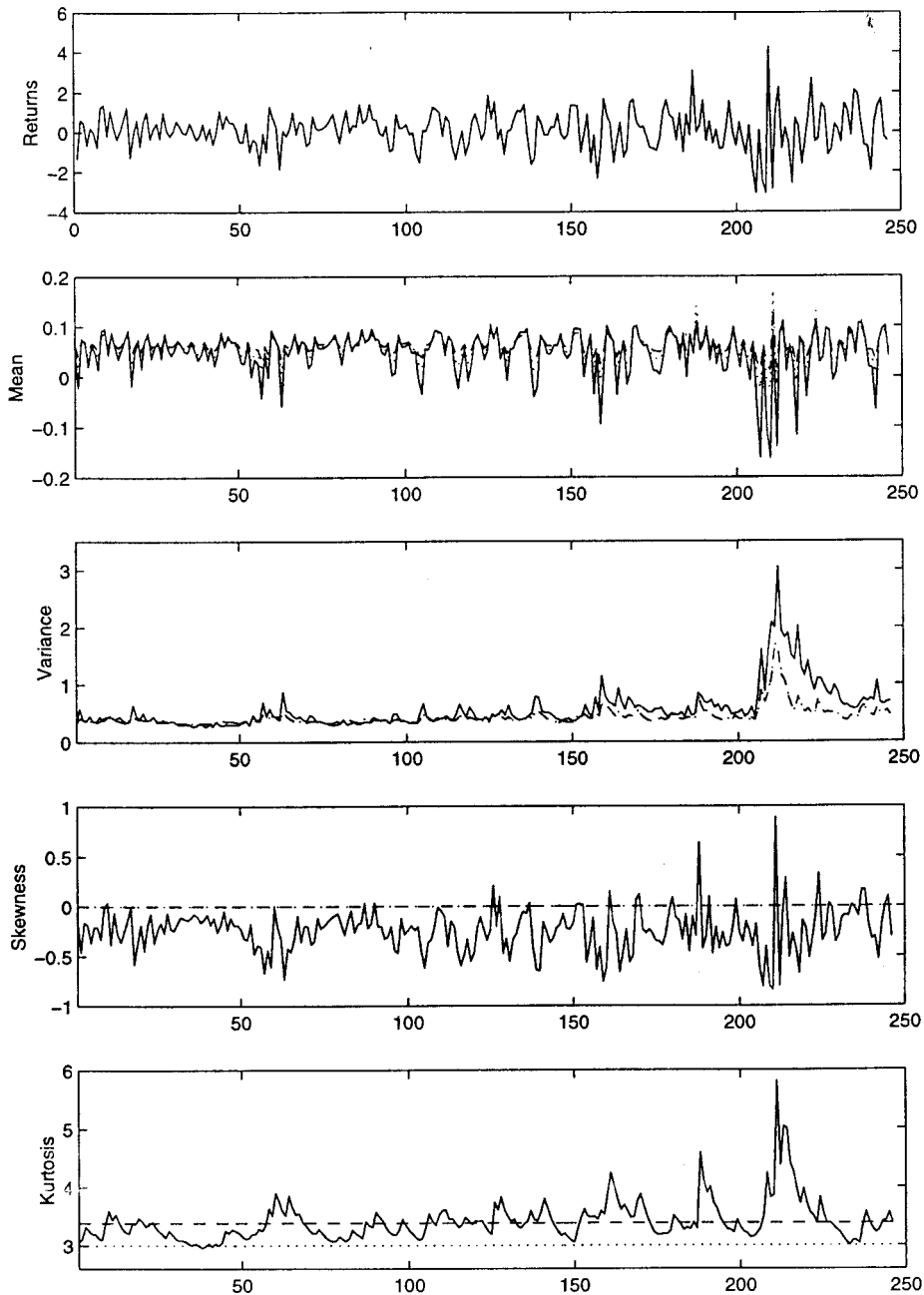


Figure 4. The series of daily FTSE 100 returns during 1997 (the last test set) and the mean, the variance, the skewness, and the kurtosis of the corresponding conditional densities estimated by the models (dotted line: GARCH model, dashed line: GARCH- t model, solid line: RMDN(2) model)

Table II. Out-of-sample statistics for the GARCH, the GARCH-*t* and the RMDN(2) models on the five test sets of FTSE 100 returns

Set	Model	Loss	Mean	Standardized residuals		
				Std.	Skew.	Kurt.
1	GARCH	0.999	0.048	0.812	0.140	3.978
	GARCH- <i>t</i>	0.984	0.063	0.794	0.167	3.903
	RMDN(2)	0.992	0.041	0.795	0.147	3.957
2	GARCH	1.275	-0.112	1.016	-0.247	2.706
	GARCH- <i>t</i>	1.297	-0.103	1.036	-0.244	2.608
	RMDN(2)	1.261	-0.072	0.986	-0.235	2.624
3	GARCH	0.953	0.061	0.930	-0.155	3.478
	GARCH- <i>t</i>	0.950	0.058	0.936	-0.170	3.510
	RMDN(2)	0.944	0.087	0.996	-0.093	3.815
4	GARCH	0.889	0.015	1.050	-0.509	3.834
	GARCH- <i>t</i>	0.887	0.011	1.053	-0.511	3.834
	RMDN(2)	0.898	0.039	1.001	-0.514	3.815
5	GARCH	1.523	0.055	1.439	-0.083	3.471
	GARCH- <i>t</i>	1.485	0.047	1.444	-0.086	3.472
	RMDN(2)	1.401	0.078	1.284	-0.068	3.134
1-5	GARCH	1.128	0.014	1.049	-0.171	3.494
1-5	GARCH- <i>t</i>	0.238	0.065	0.212	0.212	0.441
		1.121	0.015	1.053	-0.169	3.466
		0.231	0.062	0.216	0.220	0.462
1-5	RMDN(2)	1.099	0.035	1.012	-0.152	3.469
		0.197	0.057	0.156	0.218	0.511

model from overfitting the training data, the performance on a validation set is used as the criterion for model selection. More precisely, the model parameters are optimized with respect to the loss function on the training set and after each iteration the loss function on the validation set is calculated. Finally, the RMDN(2) model with the best performance on the validation set is selected. The validation sets cover the daily returns of the year preceding the period of the corresponding training set.²⁰ In particular, the GARCH, the GARCH-*t* and the RMDN(2) models are estimated from the same data sets.

Table II summarizes the performance of the models on each of the five test sets. At the bottom of the table the mean value and the standard deviation (below) of each performance measure are given. In addition to the loss function, some statistics of the standardized residuals are also reported. The standardized residuals are defined as e_t/σ_t where e_t denotes the prediction error and σ_t the predicted standard deviation at time t . The RMDN(2) model achieves the lowest value of the loss function on three test sets whereas the GARCH-*t* model performs best on two sets. There is thus substantial evidence that models with fat tails in the conditional distribution provide a better (out-of-sample) fit to daily return series than standard models with a conditional normal distribution. On average, the RMDN(2) models show the best performance. Although the differences between the loss functions are statistically not significant²¹ it seems to be favorable to

²⁰ For instance, the first validation set consists of the FTSE 100 returns of 1990 (not described in Table I).

²¹ A paired *t*-test is applied with a significance level of 5%.

Table III. Out-of-sample errors for the GARCH, the GARCH- t , and the RMDN(2) models for different volatility measures

Set	Model	Squared returns		Realized volatility	
		NMAE	HR	NMAE	HR
1	GARCH	0.938	0.640	2.197	0.538
	GARCH- t	0.956	0.640	2.320	0.538
	RMDN(2)	0.959	0.636	2.329	0.534
	ARMA	1.116	0.609	0.909	0.632
2	GARCH	0.735	0.742	1.559	0.544
	GARCH- t	0.716	0.750	1.412	0.563
	RMDN(2)	0.737	0.742	1.625	0.548
	ARMA	0.732	0.750	0.852	0.671
3	GARCH	0.780	0.714	1.643	0.567
	GARCH- t	0.778	0.718	1.624	0.567
	RMDN(2)	0.730	0.714	1.379	0.611
	ARMA	0.869	0.691	0.942	0.663
4	GARCH	0.768	0.709	0.965	0.622
	GARCH- t	0.767	0.709	0.959	0.622
	RMDN(2)	0.790	0.709	1.071	0.587
	ARMA	1.001	0.638	0.879	0.634
5	GARCH	0.730	0.740	0.846	0.691
	GARCH- t	0.730	0.740	0.847	0.687
	RMDN(2)	0.749	0.740	0.784	0.667
	ARMA	0.742	0.724	0.834	0.626
1-5	GARCH	0.790	0.709	1.442	0.592
		0.085	0.041	0.549	0.065
1-5	GARCH- t	0.789	0.711	1.433	0.596
		0.097	0.043	0.590	0.060
1-5	RMDN(2)	0.793	0.708	1.438	0.589
		0.096	0.043	0.590	0.053
1-5	ARMA	0.892	0.682	0.883	0.645
		0.167	0.059	0.044	0.020

model the time dependence of higher-order moments, e.g. by the proposed RMDN(2) model. The out-of-sample statistics of the standardized residuals are similar for all models.

The performance of the models with respect to the NMAE and the HR is reported in Table III. As expected, only weak dependencies are detected in the squared returns series. More precisely, the AR coefficients of the fitted ARMA models are between 0.1 and 0.15, if they are significant at all, and the MA coefficients are not significant. The obtained AR model is even worse than the naive predictor on the first and last sets (the NMAE is larger than 1) which is possible since the performance is evaluated out-of-sample. The GARCH, the GARCH- t and the RMDN(2) model always achieve NMAEs smaller than 1. On average, the GARCH- t model performs best. For the hit rate HR, the return series models also perform better than the AR models, and the GARCH- t model achieves the highest hit rate on average.

If we take the realized volatilities as a benchmark for evaluating the forecasting performance of our different models we observe an interesting phenomenon that will be the key for understanding the corresponding out-of-sample results. Looking at the left-hand side of Figure 5 we observe that over the entire period of the first test set the realized volatilities are systematically

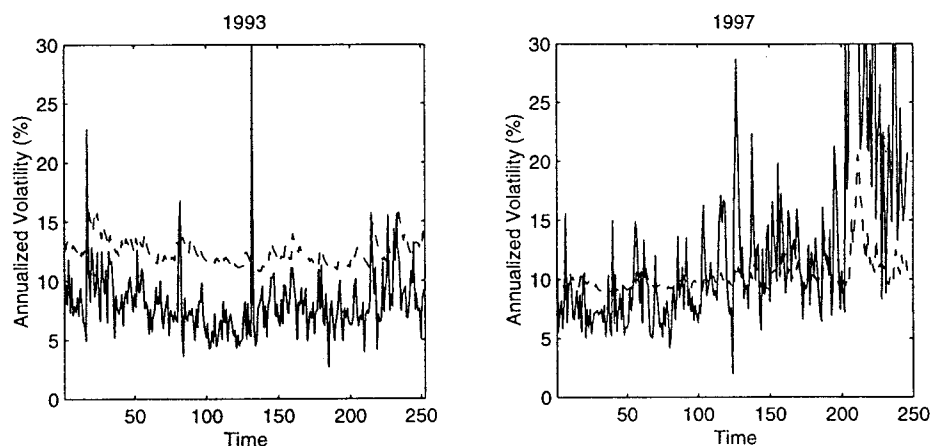


Figure 5. The (annualized) realized volatility (solid line) and the (annualized) volatility predicted by the GARCH model (dashed line) for the FTSE 100 in 1993 (on the left-hand side) and in 1997 (on the right-hand side)

below GARCH volatilities. In fact closer observation suggests that there is a parallel downward shift between GARCH and realized volatilities. Hence scaling up realized volatilities by a constant factor could eliminate this difference. Considering our way to construct realized volatilities this could be achieved by putting a higher weight on (scaling up) the overnight volatility. Changing the weight on the overnight volatility does not seem to be justified a priori. However, it might be possible that time during the trading period elapses with a different speed than during the overnight period. If that is the case this has to be reflected in different weights for intraday and overnight volatilities. In particular, if time during the overnight period passes more quickly than during the trading day we have to scale up the overnight volatility²². One possible explanation for this different 'speed of time' could be the level of (dis)integration of international capital markets. In fact one can argue that the more integrated financial markets become, the more similar time will elapse between trading and non-trading hours. Looking at the right-hand side of Figure 5 we observe that during the period of 1997 the difference between GARCH and realized volatilities is non-existent. Hence for that period it seems that the scaling of realized volatilities is correct. It is obvious that the level of integration of international financial markets in 1997 is much higher than in 1993.

Looking now at the out-of-sample results with realized volatilities as the benchmark we observe that over the years the NMAE of the return series models improves relative to the ARMA models.²³ In 1997 the results of the AR model are dominated by the RMDN(2) model with respect to the NMAE. In that year all return series models also achieve a higher hit rate than the AR model.

An alternative notion of volatility, which can be used to analyse the prediction performance of the models, is the concept of implied volatility. As mentioned earlier, the linear correlation r and the rank correlation r_s are two measures for comparing implied volatilities obtained from

²² This is achieved by a larger scaling factor of r_0^2 in Eq. (15).

²³ The AR(1) coefficient is significantly positive for all sets. For the third and fourth set the (significant) MA(1) coefficient is also included.

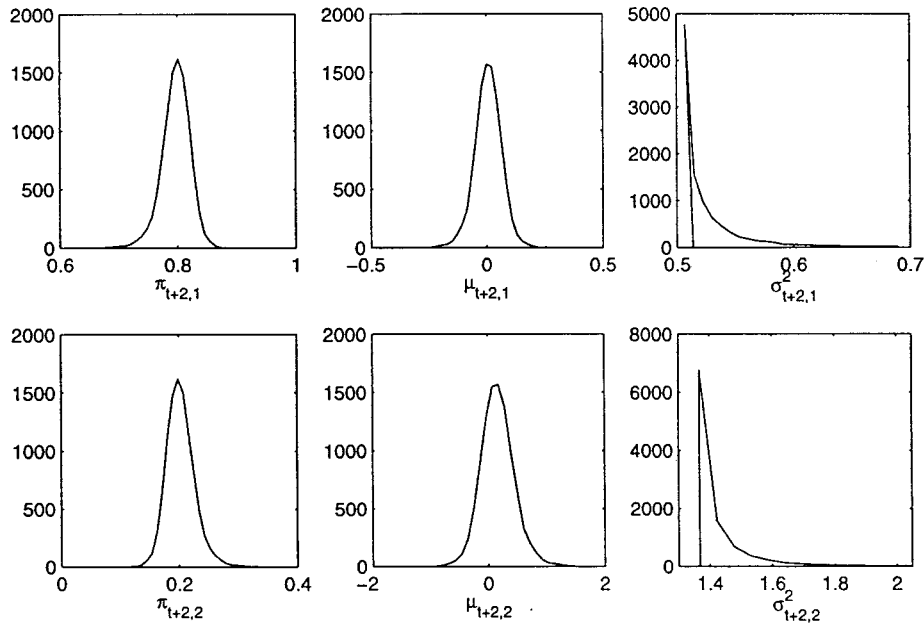


Figure 6. Histograms of the parameters of the conditional density $\rho(r_{t+2}|I_t)$ obtained from 10,000 RMDN(2) outputs. The parameters are estimated as the corresponding mean values (see text for further explanation)

the market to volatilities predicted by a model. The difference to the above error measures is not only the inclusion of real-world option data but also the prediction horizon. Now volatility is predicted *several* steps into the future, i.e. for the remaining time to maturity of the particular option.

In order to illustrate the Monte Carlo simulation technique for the RMDN(2) models, we look at the first prediction of the first test set. The parameters of the conditional density $\rho(r_{t+1}|I_t)$, which is a weighted sum of two normal densities, are determined as $\pi_{1,t+1} = 0.786$, $\pi_{2,t+1} = 0.214$, $\mu_{1,t+1} = -0.024$, $\mu_{2,t+1} = 0.310$, $\sigma_{1,t+1}^2 = 0.517$, and $\sigma_{2,t+1}^2 = 1.377$. The accumulated conditional variance is thus given by $\sigma_{t+1}^2 = 0.720$. By drawing 10000 ‘returns’ from this distribution, calculating the corresponding ‘prediction errors’ and feeding them together with $\sigma_{1,t+1}^2$ and $\sigma_{2,t+1}^2$ into the RMDN, the distributions of the parameters of the next conditional density $\rho(r_{t+2}|I_t)$ can be approximated. Histograms of these distributions are depicted in Figure 6. For $\sigma_{1,t+2}^2$ and $\sigma_{2,t+2}^2$, the histograms reflect the asymmetric distribution of e_{t+1}^2 . The mean values of the six parameters are $\pi_{1,t+2} = 0.797$, $\pi_{2,t+2} = 0.203$, $\mu_{1,t+2} = 0.007$, $\mu_{2,t+2} = 0.160$, $\sigma_{1,t+2}^2 = 0.524$, and $\sigma_{2,t+2}^2 = 1.405$. The new value of the accumulated variance, in which we are mainly interested, is $\sigma_{t+2}^2 = 0.728$. It should be noted that the accumulated variances are estimated simultaneously from the outputs of the RMDN using equation (11) mean μ (and not from the values of the parameters).

The linear correlation r and the rank correlation r_s between predicted and implied volatilities of the five test sets are reported in Table IV. Correlations which are not significantly different from 0 at the 5% level are in parentheses. The best model (the highest correlation) for each test set is in *italic*. The volatility forecasts of all models are significantly correlated with the implied

Table IV. A comparison of linear correlation (Pearson's r) and rank correlation (Spearman's r_s) between volatility implied by real-world option prices and volatility predicted by GARCH, GARCH- t and RMDN(2) models for the FTSE 100 test sets

Set	GARCH		GARCH- t		RMDN(2)	
	r	r_s	r	r_s	r	r_s
1	0.230	0.303	0.242	0.315	<i>0.318</i>	<i>0.375</i>
2	0.194	0.212	<i>0.238</i>	<i>0.296</i>	0.206	0.267
3	0.623	0.578	<i>0.627</i>	<i>0.581</i>	(0.046)	(0.006)
4	0.270	0.233	<i>0.274</i>	<i>0.238</i>	0.264	0.226
5	0.714	0.634	0.715	0.635	<i>0.879</i>	<i>0.801</i>

Correlations which are not significant (at the 5% level), are in parentheses. The best model (the highest correlation) is in italic type.

volatility on all sets except for the RMDN(2) model on the third test set²⁴. As for the loss function, it is found that the models with a leptokurtic conditional distribution produce the best results. The GARCH- t model and the RMDN(2) model achieve the highest linear and the highest rank correlation on three and two sets, respectively. The highest correlations are obtained for the RMDN(2) model on the most recent test set. This means that a conditional distribution with fat tails is essential for capturing and modelling those statistical dependencies in return series which lead to improved volatility predictions if implied volatility is defined as the reference measure.

CONCLUSIONS

We have presented a neural network-based conditional density estimator which is semi-non-parametric. In terms of time series modelling, our approach is more general than traditional GARCH models of asset return series because the shape of the conditional density depends on the actual information set and not only the variance of the conditional density. Therefore, the concept of heteroscedasticity is extended to higher-order moments of the conditional distribution such as skewness and kurtosis which may be time-dependent in our framework.

As with any newly proposed return series model, the performance of the model should be evaluated carefully. First, the performance should be measured out-of-sample so that the number of model parameters plays no role when comparing models. Furthermore, it is dangerous to evaluate the performance on a single data set or to rely on just one performance measure. This is particularly true if one is interested in volatility forecasting since volatility is not directly observable. We have tried to take into account all of these issues in our empirical analysis. The models are evaluated with respect to seven different error measures on five disjoint sets out-of-sample. As a result, the performance of each model is related to the particular data set and to the error measure applied. On average, no model can be considered superior to any other model in the sense that the differences in performance are statistically not significant. However, the

²⁴ This RMDN(2) model behaves similar to an integrated, non-stationary model. This behaviour may be obtained since the parameters (weights) of an RMDN(2) model are not restricted during the estimation procedure in contrast to the parameters of GARCH and GARCH- t models. In fact, the GARCH and GARCH- t models estimated on the third set are nearly integrated.

GARCH- t and the RMDN(2) models, which are capable of modelling fat tails in the conditional distribution, show the best performance in most cases.

Our future research activities will concentrate on the full exploitation of the information available from the whole conditional distribution of returns. In particular, it is planned to implement trading strategies based on higher-order moment properties of the conditional distribution (see, for example, Aït-Sahalia *et al.*, 1998). Furthermore, we will investigate the potential benefits of using conditional distributions with fat tails for market risk management.

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 Education, Science and Culture. The recurrent mixture density networks were implemented by extending the NETLAB software package available from <http://neural-server.aston.ac.uk/>. The authors wish to thank two anonymous referees for stimulating comments, H. Elsinger and P. Tiño for valuable discussions, and Z. Siebertova for excellent research support with the FTSE data sets.

REFERENCES

- Adya M, Collopy F. 1998. How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting* **17**: 481–495.
- Aït-Sahalia Y, Wang Y, Yared F. 1998. Do option markets correctly price the probabilities of movement of the underlying asset? Mimeo.
- Andersen TG, Bollerslev T. 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**: 885–905.
- Andersen TG, Bollerslev T, Diebold FX, Labys P. 1999. The distribution of exchange rate volatility. Working Paper 99-08, The Wharton School, University of Pennsylvania.
- Bera AK, Higgins ML. 1993. ARCH models: properties, estimation and testing. *Journal of Economic Surveys* **7**: 307–366.
- Bishop CM. 1995. *Neural Networks for Pattern Recognition*. Clarendon Press: Oxford.
- Black F, Scholes M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* **81**: 637–654.
- Bollerslev T. 1986. A generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.
- Bollerslev T. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* **69**: 542–547.
- Bollerslev T, Chou RY, Kroner KF. 1992. ARCH modelling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* **52**: 5–59.
- Boudoukh J, Richardson M, Whitelaw RF. 1997. Investigation of a class of volatility estimators. *Journal of Derivatives*. Spring: 63–71.
- Canina L, Figlewski S. 1993. The information content of implied volatility. *Review of Financial Studies* **6**: 659–681.
- Donaldson RG, Kamstra M. 1997. An artificial neural network-GARCH model for international stock return volatility. *Journal of Empirical Finance* **4**: 17–46.
- Engle RF. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica* **50**: 987–1008.
- Engle RF, González-Rivera G. 1991. Semiparametric ARCH models. *Journal of Business & Economic Statistics* **9**: 345–359.
- Engle RF, Ng VK. 1993. Measuring and testing the impact of news on volatility. *Journal of Finance* **48**: 1749–1778.

- Fama EF. 1965. The behavior of stock market prices. *Journal of Business* **38**: 34–105.
- Gallant AR, Tauchen G. 1989. Semiparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica* **57**: 1091–1120.
- Glosten LR, Jagannathan R, Runkle DE. 1993. Relationship between the expected value and the volatility of the normal excess return on stocks. *Journal of Finance* **48**: 1779–1801.
- Lamoureux CG, Lastrapes WD. 1993. Forecasting stock return variance: toward an understanding of stochastic implied volatilities. *Review of Financial Studies* **6**: 293–326.
- Latané HA, Rendleman RJ. 1976. Standard deviation of stock price ratios implied in option prices. *Journal of Finance* **31**: 369–381.
- Mandelbrot B. 1963. The variation of certain speculative prices. *Journal of Business* **36**: 394–419.
- McLachlan GJ, Basford KE. 1988. *Mixture Models: Inference and applications to clustering*. Marcel Dekker: New York.
- Nelson DB. 1990. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**: 347–370.
- Neuneier R, Finnoff W, Hergert F, Ormoneit D. 1994. Estimation of conditional densities: a comparison of neural network approaches. In *ICANN 94—Proceedings of the International Conference on Artificial Neural Networks*, Marinaro M, Morasso PG (eds). Springer: Berlin; 689–692.
- Noh J, Engle RF, Kane A. 1994. Forecasting volatility and option prices of the S&P 500 index. *Journal of Derivatives*. Fall: 17–30.
- Ormoneit D. 1998. *Probability Estimating Neural Networks*. Shaker: Aachen.
- Ormoneit D, Neuneier R. 1996. Experiments in predicting the German stock index DAX with density estimating neural networks. In *Proceedings of the 1996 Conference on Computational Intelligence in Financial Engineering (CIFEr)*. 66–71.
- Pagan AR, Schwert GW. 1990. Alternative models for conditional stock volatility. *Journal of Econometrics* **45**: 267–290.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 1992. *Numerical Recipes in C*. Cambridge University Press: Cambridge.
- Refenes AP. 1995. *Neural Networks in the Capital Markets*. Wiley: New York.
- Schittenkopf C, Dorffner G, Dockner EJ. 1998. Volatility prediction with mixture density networks. In *ICANN 98—Proceedings of the 8th International Conference on Artificial Neural Networks*, Niklasson L, Bodén M, Ziemke T (eds). Springer: Berlin; 929–934.
- Schmitt C, Kaehler J. 1996. Delta-neutral volatility trading with intraday prices: an application to options on the DAX. Discussion Paper No. 96-25, Zentrum für Europäische Wirtschaftsforschung, Mannheim, Germany.
- Tompkins R. 1994. *Options Explained*. Macmillan Press: New York.

Authors' biographies:

Christian Schittenkopf has been a research fellow at the Austrian Research Institute for Artificial Intelligence since 1998. His research interests include dynamical systems, information theory, modelling in economics and finance, neural networks, and time series analysis. He received his PhD (Dr.rer.nat.) in 1997 from Technische Universität München, Munich, Germany.

Georg Dorffner, MS PhD, is an associate professor at the Department of Medical Cybernetics at the University of Vienna and the head of the neural networks group at the Austrian Research Institute for Artificial Intelligence. His current research focus is on time series and signal processing, and the modelling of stochastic processes.

Engelbert J. Dockner is currently Professor of Finance in the Department of Business Administration at the University of Vienna. His research interests include risk management, volatility prediction, dynamical systems, and dynamic optimization. He held tenure track positions at the University of Saskatchewan, Queen's University, Canada and University of Bielefeld, Germany.

Authors' addresses:

Christian Schittenkopf and **Georg Dorffner**, Austrian Research Institute for Artificial Intelligence, Schottengasse 3, A-1010 Vienna, Austria.

Engelbert J. Dockner, Department of Business Administration, University of Vienna, Brünnerstrasse 72, A-1210 Vienna, Austria.