Risk-neutral Density Extraction from Option Prices: Improved Pricing with Mixture Density Networks

Christian Schittenkopf
Georg Dorffner

Report No. 47
August 2000
August 2000

SFB
‘Adaptive Information Systems and Modelling in Economics and Management Science’

Vienna University of Economics
and Business Administration
Augasse 2–6, 1090 Wien, Austria

in cooperation with
University of Vienna
Vienna University of Technology

http://www.wu-wien.ac.at/am

Papers published in this report series
are preliminary versions of journal articles
and not for quotations.

This paper was submitted for publication in:
IEEE Transactions on Neural Networks

This piece of research was supported by the Austrian Science Foundation (FWF)
under grant SFB#010 (‘Adaptive Information Systems and Modelling in Economics
and Management Science’).
Risk-neutral Density Extraction from Option Prices: Improved Pricing with Mixture Density Networks

Christian Schittenkopf, Georg Dorffner
Austrian Research Institute for Artificial Intelligence
Vienna, Austria

Abstract
One of the central goals in finance is to find better models for pricing and hedging financial derivatives such as call and put options. We present a semi-nonparametric approach to risk-neutral density extraction from option prices which is based on an extension of the concept of mixture density networks. The central idea is to model the shape of the risk-neutral density in a flexible, non-linear way as a function of the time horizon. Thereby, stylized facts such as negative skewness and excess kurtosis are captured. The approach is applied to a very large set of intraday options data on the FTSE 100 recorded at LIFFE. It is shown to yield significantly better results in terms of out-of-sample pricing in comparison to the basic Black-Scholes model and to an extended model adjusting the skewness and kurtosis terms. From the perspective of risk management, the extracted risk-neutral densities provide valuable information about market expectations.

1 Introduction
The phenomenal growth in turnover and volume of financial derivatives\(^1\) over the past decades indicates that derivatives have become key components of investors' portfolios around the world. The correct pricing of derivatives such as call and put options is thus of substantial interest for market participants. The main problem with the standard Black-Scholes model \(^6\) is that the predicted prices differ significantly from the prices observed at financial markets. These systematic errors are usually referred to as the term structure of volatility\(^2\) and the volatility smile\(^3\). To overcome these shortcomings, at least partially, other option pricing models like the time-discrete GARCH model \(^8\) or the time-continuous stochastic volatility models (see, for instance, \(^10\)) have been derived. As an alternative approach, neural networks have been successfully applied to estimate pricing formulae of financial derivatives, also in combination with the Black-Scholes model \(^3, 9, 13\).

---

\(^1\)See \(^{12}\) for an excellent overview.

\(^2\)The term structure of volatility is the dependence of the implied volatilities, which are the volatilities of the observed option prices for the Black-Scholes model, on the time to maturity.

\(^3\)When the implied volatilities are plotted against moneyness, the shape of the resulting curve is often reminiscent of a smile deviating substantially from a horizontal line, which the Black-Scholes model predicts. The moneyness of an option is defined by the relation of the option's strike price to the observed price (value) of the underlying asset.
The parameters of an option pricing model (such as the volatility in the Black-Scholes model) can be estimated from historical time series or from option prices observed at financial markets. Especially the latter approach has attracted increasing interest over the last decade (see, for instance, [19] and [23]). An alternative concept for extracting relevant information from market data is to estimate the risk-neutral (probability) density (function) of assets or asset returns from observed prices [16, 18]. This represents a powerful approach since the expectations of market participants about future developments are extracted and modeled, which is essential in many contexts. For instance, it is possible to analyze the deviations of the extracted risk-neutral density from the log-normal density of the Black-Scholes model and to exploit this more realistic picture to achieve higher accuracy in derivative pricing and hedging. From the perspective of risk management, the entire implied return distribution of assets provides more information for Value-at-Risk calculations than the implied volatility, i.e., the standard deviation of the implied return distribution of the Black-Scholes model. Finally, central banks can compare their view of the economy with the market’s view which is captured in the risk-neutral densities of key indicators such as stock indices, exchange rates, or interest rates [1].

In literature several approaches to risk-neutral density estimation have been described. First, one can generalize the Black-Scholes framework by formulating series expansions of the (log-)normal density and by estimating the parameters of the expansion from option prices. Jarrow and Rudd [15] derive an option pricing formula from an Edgeworth expansion of the log-normal density of stock prices whereas Corrado and Su [7] use a Gram-Charlier series expansion of the normal density of log-returns. Another approach is to model the pricing function of a European call option, e.g., by using a nonparametric estimator [2] or a neural network-based pricing model [3, 9, 13] and to differentiate twice with respect to the strike price. Finally, the risk-neutral density can also be estimated directly using nonparametric methods [14] or a mixture of Gaussians approach [21, 22].

This paper introduces a more general framework for risk-neutral density estimation from observed option prices. More precisely, the approximation of the risk-neutral density by a mixture of Gaussians is extended in the sense that the parameters of the mixture are functions of a set of input variables. This concept has originally been introduced as mixture density networks [4, 5, 20]. In this paper the concept is extended to provide a flexible tool for risk-neutral density estimation and derivative pricing and hedging. One of the strengths of the proposed method is the modeling of density characteristics in dependence of input variables such as the time horizon, the moneyness of the option and other potentially important factors. It turns out that the Black-Scholes model can be represented as a special case of an extended mixture density network.

We describe the Black-Scholes model, the adjusted model of Corrado and Su [7], and our new model in detail in Section 2. Some remarks on the data set of option contracts on the FTSE 100 are given in Section 3. In Section 4 the framework for estimating the models from option prices and for evaluating their performance out-of-sample is described. In Section 5 the models are applied to extract risk-neutral densities in many different setups. The performance of the models is measured by their out-of-sample pricing and hedging accuracy. Due to the large number of independent setups, it is possible to compare the performance of the models by running statistical tests. This means that differences in the performance of the models can be identified as statistically significant or not significant. Additionally, the risk-neutral densities extracted by the models are analyzed. Section 6 offers some concluding remarks and gives an outlook on future research.
2 Option Pricing Models

The three option pricing models, or more generally, the three frameworks for estimating risk-neutral densities described in this section are the basic Black-Scholes model [6], the adjusted Black-Scholes model of Corrado and Su [7], and our new model which is an extension of the concept of mixture density networks [4, 5, 20]. At the beginning, we briefly recall the risk-neutral valuation framework.

Consider a financial derivative on an underlying asset $S$ with current price $S_0$ and with price $S_t$ at maturity where $t$ denotes the time to maturity. The payoff function $\phi$ of the derivative is assumed to depend only on the price of the underlying asset at maturity, i.e., $\phi \equiv \phi(S_t)$. In the risk-neutral valuation framework the relationship between the price $\pi$ of the derivative and the risk-neutral density $f_t(y)$ of continuously compounded $t$-period returns $y = \log \frac{S_t}{S_0}$ of the underlying asset $S$ is given by

$$
\pi = e^{-r t} \int_{-\infty}^{\infty} \phi(S_0e^y) f_t(y) dy \tag{1}
$$

where $r$ denotes the continuously compounded riskless interest rate over the lifetime of the derivative contract. In other words, once we fix the risk-neutral density $f_t(y)$ of the returns $y$ we can calculate the price $\pi$ of the derivative which is characterized by its payoff function $\phi$. Whether analytical pricing formulae are obtained or whether prices must be determined by time-intensive Monte Carlo simulations depends on the payoff function of the derivative and on the risk-neutral density assumed.

2.1 The Black-Scholes Model

The Black-Scholes (BS) model [6] is the most fundamental pricing model for financial derivatives. It is derived under several simplifying assumptions such as geometric Brownian motion of the price (value) of the underlying asset, constant volatility, and no transactions costs. In this framework the risk-neutral density is Gaussian with mean $(r - \frac{1}{2} \sigma^2)t$ and variance $\sigma^2 t$, i.e.,

$$
f_t^B S(z) = \frac{1}{\sqrt{2\pi \sigma^2 t}} \exp \left( -\frac{z^2}{2} \right) \text{ with } z \equiv z(y) = \frac{y - (r - \frac{1}{2} \sigma^2)t}{\sigma \sqrt{t}} \tag{2}
$$

where $\sigma$ denotes the volatility of the underlying asset $S$. For the payoff function $\phi(S_t) = \max(S_t - K, 0)$ where $K$ denotes the strike price, evaluation of the integral in Eq. (1) yields the famous BS formula for the price $\pi_C^{BS}$ of a European call option:

$$
\pi_C^{BS} = S_0 N(d) - Ke^{-r t} N(d - \sigma \sqrt{t}) \tag{3}
$$

$$
d = \frac{\log \left( \frac{S_0}{K} \right) + (r + \frac{1}{2} \sigma^2)t}{\sigma \sqrt{t}} \tag{4}
$$

where $N(.)$ denotes the standard normal cumulative distribution function. For a European put option with payoff function $\phi(S_t) = \max(K - S_t, 0)$, the price $\pi_P^{BS}$ is obtained as

$$
\pi_P^{BS} = \pi_C^{BS} + Ke^{-r t} - S_0. \tag{5}
$$

The restrictions of the risk-neutral density in Eq. (2), i.e., the linear scaling in mean and variance as a function of the time to maturity $t$ and the absence of skewness and

---

4 Dividend effects are included later.
excess kurtosis lead to substantial differences between derivative prices according to the BS model and prices observed at financial markets. Therefore, it has been one of the main goals of research in finance to improve the pricing accuracy of the BS model.

### 2.2 The Adjusted Model of Corrado and Su

One way to achieve a better performance is to provide skewness and kurtosis adjustment terms for the BS formula by means of a series expansion of the density in Eq. (2). Corrado and Su [7] apply a Gram-Charlier series expansion up to the fourth order which yields risk-neutral densities of the form

\[
j_t^{CS}(z) = j_t^{BS}(z) \left(1 + \frac{\mu_3}{6}(z^3 - 3z) + \frac{\mu_4 - 3}{24}(z^4 - 6z^2 + 3)\right)
\]  

(6)

where \(\mu_3\) is the skewness and \(\mu_4\) is the kurtosis of the corresponding distribution. In practice, it may happen that the estimated parameters \(\mu_3\) and \(\mu_4\) are such that \(j_t^{CS}(z)\) is negative for some \(z\) and is therefore not a density. The BS model is obtained as a special case (\(\mu_3 = 0, \mu_4 = 3\)). For a European call option, the price \(\pi_t^{CS}\) is given by

\[
\pi_t^{CS} = \pi_t^{BS} + \frac{\mu_3}{6}S_0\sigma\sqrt{t} \left((2\sigma\sqrt{t} - d)n(d) + \sigma^2tN(d)\right) + \frac{\mu_4 - 3}{24}S_0\sigma\sqrt{t} \left((d^2 - 1 - 3\sigma\sqrt{t}(d - \sigma\sqrt{t}))n(d) + \sigma^3t^{3/2}N(d)\right)
\]  

(7)

where \(n(\cdot)\) denotes the standard normal density and \(d\) is defined in Eq. (4). The price \(\pi_t^{PS}\) of the put option can be derived as

\[
\pi_t^{PS} = \pi_t^{CS} + Ke^{-rt} - S_0 \left(1 + \frac{\mu_3}{6}\sigma^3t^{3/2} + \frac{\mu_4 - 3}{24}\sigma^4t^2\right).
\]  

(8)

While this model takes into account fat tails in the risk-neutral density, it still suffers from the linear scaling in mean and variance as a function of the time to maturity \(t\). The skewness and the kurtosis of the density are constant (\(\mu_3\) and \(\mu_4\), respectively) and do not depend on \(t\).

### 2.3 The Extended Mixture Density Network Model

The concept of mixture density networks (MDNs) has been introduced in [4, 5, 20]. The main idea is to approximate a conditional density by a mixture of Gaussians the parameters of which are determined by multi-layer perceptrons (MLPs) as a non-linear function of some input variables (the conditioning information set). This approach is completely general in the sense that Gaussian mixture models can approximate any density with arbitrary accuracy as the number of Gaussians goes to infinity [17]. This concept can be extended to estimate risk-neutral densities. In the simplest setting, the MDN has just one input which is the time to maturity. The risk-neutral density is thus specified by

\[
f_t^{MDN}(y) = \sum_{i=1}^{n} \frac{a_{i,t}}{\sqrt{2\pi}\sigma_{i,t}^2} \exp \left(-\frac{(y - \mu_{i,t})^2}{2\sigma_{i,t}^2}\right)
\]  

(9)

\footnote{In this paper we restrict our attention to models of this type. It is easy to take into account other factors by increasing the number of network inputs.}
where \( n \) denotes the number of Gaussians and the parameters \( \alpha_{i,t}, \mu_{i,t}, \) and \( \sigma_{i,t}^2 \) are modeled separately by MLPs\(^6\) with input \( t \). The outputs \( \tilde{\alpha}_{i,t}, \tilde{\mu}_{i,t}, \) and \( \tilde{\sigma}_{i,t} \) of the MLPs\(^7\) are transformed into the parameters of the mixture distribution by applying the following activation functions:

\[
\alpha_{i,t} = s(\tilde{\alpha}_{i,t}) = \exp(\tilde{\alpha}_{i,t}) / \sum_{j=1}^{n} \exp(\tilde{\alpha}_{j,t})
\]

which ensures that the priors \( \alpha_{i,t} \) are positive and that they sum up to one, which makes \( f_t^{MDN}(y) \) in Eq. (9) a density:

\[
\mu_{i,t} = \tilde{\mu}_{i,t};
\]

\[
\sigma_{i,t}^2 = (\tilde{\sigma}_{i,t})^2
\]

which guarantees non-negative variances. From now, an MDN with \( n \) Gaussians is called an MDN\((n)\) model. The architecture of an MDN\((2)\) model is depicted in Fig. 1.

The extension of MDNs which is necessary to model risk-neutral densities affects the processing of the network outputs \( \alpha_{i,t}, \mu_{i,t}, \) and \( \sigma_{i,t}^2 \). More precisely, we have to evaluate the integral in Eq. (1) with respect to the density specified by the MDN in Eq. (9). For a European call option, one obtains the following non-linear relationship between the price \( \pi_C^{MDN} \) and the network outputs \( \alpha_{i,t}, \mu_{i,t}, \) and \( \sigma_{i,t}^2 \):

\[
\pi_C^{MDN} = e^{-rt} \sum_{i=1}^{n} \alpha_{i,t} \left( S_0 e^{\mu_{i,t} + \frac{1}{2} \sigma_{i,t}^2 t} \mathcal{N}(d_{i,t}) - K \mathcal{N}(d_{i,t} - \sigma_{i,t}) \right)
\]

\(^6\)This is a minor extension of the original MDN architecture where the parameters are the outputs of a single MLP.

\(^7\)Each MLP maps the one-dimensional input \( t \) onto an \( n \)-dimensional output. For the MLP estimating the means, for instance, the \( i \)-th output component is given by \( \tilde{\mu}_{i,t} = \sum_{j=1}^{H} \nu_{ij} \tanh(w_{ij}t + c_j) + b_i \), \( 1 \leq i \leq n \), where \( H \) denotes the number of hidden units, \( w_{ij} \) and \( \nu_{ij} \) the weights of the first and second layer, and \( c_j \) and \( b_i \) the biases of the first and second layer, respectively.
\[ d_{i,t} = \frac{\log \left( \frac{S_t}{K} \right) + \mu_{i,t} + \sigma^2_{i,t}}{\sigma_{i,t}}. \]  

For a European put option, the price \( \pi^M_D\) is given by

\[ \pi^M_P = \pi^M_D + e^{-rt} \left( K - S_0 \sum_{i=1}^{n} \alpha_{i,t} e^{\mu_{i,t} t + \frac{1}{2} \sigma^2_{i,t} t} \right). \]  

Due to these analytical solutions for option prices, the network parameters, i.e., the weights of the MDN can be determined very fast by standard optimization routines for a given set of option prices.

The important feature introduced by this model is the fact that the shape of the risk-neutral density is (via its parameters) a non-linear function of the time to maturity \( t \). In particular, the model allows for skewed and fat-tailed risk-neutral densities which change in dependence of \( t \). Any dependence of the parameters on \( t \) can be modeled since MLPs can approximate any non-linear function with arbitrary accuracy as the number of hidden units goes to infinity [11]. The BS model is a special case of an MDN(1) model: \( \alpha_{1,t} = 1, \mu_{1,t} = (r - \frac{1}{2} \sigma^2) t, \sigma^2_{1,t} = \sigma^2 t \). Once more we emphasize that the model can be easily extended by adding other factors such as moneyness to the set of input variables. Thereby, the risk-neutral density and its moments would also depend on other factors (besides the time to maturity).

3 Data Set

The data set we use to test our model comprises transactions data from European options on the FTSE 100 index traded at the London International Financial Futures and Options Exchange (Liffe). The option prices are recorded synchronously with the FTSE 100 and time-stamped to the nearest second. The high frequency observations start on 4 January 1993 and end on 22 October 1997 covering a period of 1210 trading days. The sample comprises 102211 traded call and put option contracts.

In a first step the records are carefully checked for recording errors and prices violating boundary conditions. Then all contracts with a time to maturity of less than two weeks\(^8\), with a price below 5 points\(^9\), with moneyness\(^10\) outside \([-0.1, 0.1]\), and with an annualized volatility\(^11\) below 5% or above 50% are removed. The final set consists of 65549 option contracts (33633 call options and 31916 put options).

A brief description of the option contracts is given in Table 1. We report the 0th, the 10th, the 25th, the 50th, the 75th, the 90th and the 100th percentile\(^12\) of several sample characteristics. During the sample period the FTSE 100 rose from about 2700 points to more than 5000 points. According to that, strike prices range from 2525 to 5875. The median of the time to maturity of all option contracts is 42 days. The longest contract has a time to maturity of more than one year. In Table 1 we also report the statistics

---

\(^8\) Options with a short time to maturity are usually excluded since the option price may vary heavily toward the end of a contract.

\(^9\) These options are eliminated to reduce the effect of “integer pricing behaviour” [3].

\(^10\) In this paper moneyness is defined as \((S_t - K)/K\) for call options and \((K - S_t)/K\) for put options.

\(^11\) implied by the Black-Scholes model (see Section 2)

\(^12\) The 0th and the 100th percentile are the minimum and the maximum, respectively.
of the risk-free interest rate\(^1\)\(^3\) and the ex post dividend yield of the FTSE 100 which are inputs to the option pricing models. Finally, the percentiles of the implied volatilities of all options are reported. The median of the implied volatility is about 14%, and 80% of all contracts have an implied volatility between 9% and 20%.

4 Methods

4.1 Model Estimation and Evaluation

The data set comprises 65549 option contracts traded on 1210 days, i.e., the average number of contracts per day is about 54. However, trading activity is not equally distributed. On the least active day, which is 4 May 1993, only 7 trades were recorded whereas 351 transactions took place on the most active day, which is 16 July 1997. Therefore it is not meaningful to reestimate the option pricing models every day. On the other hand, non-stationarity of the data requires that models are reestimated within relatively short time intervals. In our experimental setup a sliding window technique is applied. The size of the time window is ten trading days meaning that option contracts are collected over a period of ten days. From this set (the training set), the option pricing models are estimated by minimizing the mean squared pricing error

\[
\text{MSPE} = \frac{1}{N} \sum_{i=1}^{N} (\pi_i - \hat{\pi}_i)^2
\]

(16)

where \(N\) is the number of contracts in the time window and \(\pi_i\) and \(\hat{\pi}_i\) denote the observed and the predicted option price, respectively. Finally, the models are tested out-of-sample on the next window of ten days (the test set). Then the time window is shifted by ten days, the models are reestimated and so on. We think that this procedure is a reasonable trade-off between the size of the data sets (the amount of market information available) on the one hand and stationarity of the data on the other hand.

Since MDN models have more degrees of freedom than the basic and the adjusted BS model, they may overfit the presented set of option prices and at the same time perform

\(^1\) Daily GBP-LIBOR quotes for overnight, one week, one month, three months, six months, and one year were linearly interpolated.
poorly out-of-sample. In order to avoid this effect, the MDN models are selected with respect to their pricing error on a validation set. More precisely, the MSPE on the time window preceding the training set\(^{14}\) is calculated after each iteration during the training of the networks and finally, the model with the lowest error on the validation set is selected. Since the data set comprises transactions data from 1210 trading days, 121 time windows and 119 training/validation/test sets are obtained. The number of options in one time window ranges from 114 to 1357. The total number of options in all test sets is 65052.

In order to compare the models in a fair way, slightly generalized versions of the basic and the adjusted BS model which incorporate dividend effects are estimated\(^{15}\). This generalization is necessary since the MDN models are able to implicitly adjust the parameters of the risk-neutral density for dividend effects through their weights\(^{16}\). The weights of the MDN models are estimated by applying a conjugate gradient routine.

4.2 Pricing Errors

The out-of-sample pricing performance of the models is measured by two error measures, namely the pricing error PE and the absolute pricing error APE:

\[
PE = \pi_t - \tilde{\pi}_t, \ APE = |\pi_t - \tilde{\pi}_t|. \tag{17}
\]

While the average PE is a measure of the bias of the pricing model, the average APE quantifies the pricing accuracy. The pricing performance of the models is analyzed in two steps: First, all prices \(\pi_t\) and the corresponding predicted prices \(\tilde{\pi}_t\) of all test sets are collected and both error measures are calculated.

In a second step the average PE and the average APE are calculated for each test set. Since the test sets are disjoint and since different test sets are evaluated with different pricing models, the average PEs and APEs may be assumed to be independent\(^{17}\). Therefore, standard tests from statistics can be applied. We perform \(t\)-tests to test for a non-zero PE for each model and paired \(t\)-tests to compare the APEs of the models. The significance level for all tests is chosen as 5%.

5 Empirical Results

In this section the pricing accuracy of the models is analyzed in detail and tested for significant differences. We also study the characteristics of the extracted densities. Finally, the hedging performance of the models is evaluated and compared.

5.1 Pricing Accuracy

The mean and the standard deviation of the PE and the APE of the models are reported in Table 2. Since the PE is the difference between the true price and the model price, a

---

\(^{14}\)In a time series framework, the choice of a validation set which is recorded before the training set is suboptimal since the training set may contain information on the validation data due to persistence effects. Nevertheless, we used this setup to make sure that all models are estimated from the same set of most recently recorded option prices.

\(^{15}\)This changes the mean of the risk-neutral density of the basic/adjusted BS model to \((r - \delta - \frac{1}{2}\sigma^2)t\) where \(\delta\) denotes the dividend yield (see Eq. 2).

\(^{16}\)See [9] for a discussion of this effect for another model.

\(^{17}\)The autocorrelation function of the series of average PEs and APEs is not significantly different from zero for all lags for all models.
<table>
<thead>
<tr>
<th>Model</th>
<th>PE Mean</th>
<th>PE Std.</th>
<th>APE Mean</th>
<th>APE Std.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td>0.26</td>
<td>11.67</td>
<td>8.49</td>
<td>8.02</td>
</tr>
<tr>
<td>BS-A</td>
<td>-0.41</td>
<td>10.48</td>
<td>7.45</td>
<td>7.37</td>
</tr>
<tr>
<td>MDN(1)</td>
<td>1.36</td>
<td>10.73</td>
<td>7.50</td>
<td>7.79</td>
</tr>
<tr>
<td>MDN(2)</td>
<td>0.84</td>
<td>9.53</td>
<td>6.33</td>
<td>7.18</td>
</tr>
</tbody>
</table>

Table 2: Summary statistics of the out-of-sample pricing error PE and the out-of-sample absolute pricing error APE for the Black-Scholes model (BS), the adjusted Black-Scholes model (BS-A), and the extended mixture density network models with one Gaussian (MDN(1)) and with two Gaussians (MDN(2)).

<table>
<thead>
<tr>
<th>Moneyness</th>
<th>Time to Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ST</td>
</tr>
<tr>
<td>OTM</td>
<td>9130</td>
</tr>
<tr>
<td>ATM</td>
<td>14180</td>
</tr>
<tr>
<td>ITM</td>
<td>1759</td>
</tr>
<tr>
<td>All</td>
<td>25069</td>
</tr>
</tbody>
</table>

Table 3: The number of options in the short-term (ST), medium-term (MT), and long-term (LT) maturities classes and in the out-of-the-money (OTM), at-the-money (ATM), and in-the-money (ITM) classes.

A positive value of PE indicates underpricing of the model whereas a negative value means that the model on average overprices the option. The results can be summarized in the following way: The basic and the adjusted BS model are unbiased as well as the MDN(2) model. The MDN(1) model tends to underestimate prices (the bias is significantly positive). On average, the MDN(2) model exhibits the highest pricing accuracy. The average APE is about 6.3 points. The MDN(2) model performs significantly better than the other pricing models. The adjusted BS model and the MDN(1) model perform second best with an average APE of about 7.5 points. Both models are still significantly better than the basic BS model which performs worst. The average APE of the basic BS model is about 8.5 points.

For a closer inspection of PEs and APEs, the set of out-of-sample option contracts is divided into classes according to moneyness and time to maturity. The class of short-term (ST) options contains all options with a time to maturity of less than five weeks. Similarly, the classes of medium-term (MT) and long-term (LT) options consist of all options with a time to maturity between five and 14 weeks and more than 14 weeks, respectively. Concerning moneyness, the following classes are distinguished: All options with moneyness inside $[-0.02, 0.02]$ are called at-the-money (ATM), all options with moneyness below $-0.02$ are called out-of-the-money (OTM), and all options with moneyness above 0.02 are called in-the-money (ITM). An overview of the number of options in the different classes is given in Table 3. Obviously, OTM options (28949) are traded much more frequently than ITM options (5113) and about as frequently as ATM options (30990). The ST and MT maturity classes roughly contain the same number of options (25069 and 27448) which is about twice the number of LT options (12535).

---

18 The difference between the models is not significant.
<table>
<thead>
<tr>
<th>Moneyness</th>
<th>Model</th>
<th>Time to Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ST</td>
</tr>
<tr>
<td>OTM</td>
<td>BS</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>-0.93</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>1.53</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>0.07</td>
</tr>
<tr>
<td>ATM</td>
<td>BS</td>
<td>-2.98</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>-3.13</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>-0.88</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>-0.39</td>
</tr>
<tr>
<td>ITM</td>
<td>BS</td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>2.59</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>4.41</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>2.26</td>
</tr>
<tr>
<td>All</td>
<td>BS</td>
<td>-1.00</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>-1.93</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>-0.04</td>
</tr>
</tbody>
</table>

Table 4: Summary statistics of the out-of-sample pricing error PE for the options of the short-term (ST), medium-term (MT), and long-term (LT) maturity classes and the out-of-the-money (OTM), at-the-money (ATM), and in-the-money (ITM) classes for the Black-Scholes model (BS), the adjusted Black-Scholes model (BS-A), and the extended mixture density network models with one Gaussian (MDN(1)) and with two Gaussians (MDN(2)).

The pricing errors of the models for the various classes of options are reported in Table 4. By splitting the average PE for each class into the average PEs of the test sets, we can again apply statistical tests\(^1^9\). We find that all models significantly underprice ITM options. The MDN(1) model also significantly underprices OTM options whereas the basic and the adjusted BS model significantly overprice ATM options. For the MDN models, overpricing occurs only for ST-ATM options but is not significant. The basic and the adjusted BS model significantly overprice ST and MT options. In contrast to that, they significantly underprice LT options. The MDN models significantly underprice MT options. The entries in the lowest and right-most field of Table 4 coincide, of course, with the entries in the second column of Table 2.

The results for the more relevant absolute pricing error APE are summarized in Table 5. Our main findings are the following ones: First, the APE increases with moneyness and time to maturity for all models. This is a consequence of the fact that option prices and thus bid-ask spreads increase with moneyness and time to maturity. Secondly, the MDN(2) model outperforms the other models not only after aggregation over all classes but also for each single class\(^2^0\). The basic BS model is the worst model for all but two classes\(^2^1\). Finally, the adjusted BS model outperforms the basic BS model for all classes.

\(^1^9\)The significance level is again 5%.

\(^2^0\)The difference in performance to the adjusted BS model for ST-ITM options is negligible. The improvement achieved by the MDN(2) model is significant except for ST-ATM and MT-ATM options for the MDN(1) model.

\(^2^1\)The MDN(1) model performs worse for MT-ITM and LT-ITM options. The BS model is significantly worse except for ST-OTM options for the MDN(1) model, for ATM and LT-ITM options for the adjusted BS model, and for ST-ITM options for the MDN models.
<table>
<thead>
<tr>
<th>Moneyness</th>
<th>Model</th>
<th>Time to Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ST</td>
</tr>
<tr>
<td>OTM</td>
<td>BS</td>
<td>5.32</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>4.43</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>5.19</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>3.90</td>
</tr>
<tr>
<td>ATM</td>
<td>BS</td>
<td>7.53</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>7.47</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>6.57</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>6.35</td>
</tr>
<tr>
<td>ITM</td>
<td>BS</td>
<td>9.62</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>7.93</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>9.42</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>7.95</td>
</tr>
<tr>
<td>All</td>
<td>BS</td>
<td>6.87</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>6.39</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>6.27</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>5.57</td>
</tr>
</tbody>
</table>

Table 5: Summary statistics of the out-of-sample absolute pricing error APE for the options of the short-term (ST), medium-term (MT), and long-term (LT) maturity classes and the out-of-the-money (OTM), at-the-money (ATM), and in-the-money (ITM) classes for the Black-Scholes model (BS), the adjusted Black-Scholes model (BS-A), and the extended mixture density network models with one Gaussian (MDN(1)) and with two Gaussians (MDN(2)).

with the largest improvement for OTM and ITM options\(^{22}\). This effect is plausible since the fat tails of the risk-neutral density are more relevant for the pricing of OTM and ITM options than ATM options. Similarly, the MDN(2) model achieves a lower APE than the MDN(1) model for all classes\(^{23}\). Again, the improvement is larger for OTM and ITM options due to the fat tails in the risk-neutral density. For ATM options, the non-linear specification of the MDN(1) model has a larger impact than the non-Gaussian risk-neutral density of the adjusted BS model\(^{24}\).

All together, the basic BS model achieves a comparable pricing accuracy (relative to the other models) only for ST-ATM and MT-ATM options. The modeling of non-linear dependencies for ATM options is more beneficial than distributional variations of the model, especially for LT options. For OTM and ITM options however, distributional generalizations of the models which allow for negative skewness and excess kurtosis are of crucial importance.

### 5.2 Comparison of Risk-neutral Densities

The results of the previous section indicate that both non-linear specifications (as in the MDN(1) and MDN(2) model) and non-Gaussian densities (as in the adjusted BS model and in the MDN(2) model) are essential for modeling risk-neutral densities. As an example, the risk-neutral densities estimated by the models for the first setup are depicted

\(^{22}\)The significant differences are mentioned in footnote 21.
\(^{23}\)The significant differences can be found in footnote 20.
\(^{24}\)The APE is significantly lower for all times to maturity.
Figure 2: (Left-hand side) The risk-neutral densities estimated by the Black-Scholes model (solid line), the adjusted Black-Scholes model (dashed line), and the extended mixture density network models with one Gaussian (dotted line) and with two Gaussians (dash-dotted line) for a time to maturity of 30 days. (Right-hand side) Enlargement of the left tails of the densities depicted on the left-hand side.

in Fig. 2 for a time to maturity of 30 days. While the densities of the BS model and the MDN(1) model are Gaussian, the densities of the adjusted BS model and the MDN(2) model are characterized by negative skewness and excess kurtosis. The enlargement of the left tail shows that these two models put much more probability mass on large negative returns which investors are substantially worried about since the crash in 1987. Only for these models, returns which are several standard deviations (σ ≈ 0.046) from the mean, might be observed with non-negligible probability.

Fig. 3 reveals how important it is to allow the shape of the risk-neutral density to vary with the time horizon. In order to facilitate interpretations, the mean, the standard deviation, the skewness, and the excess kurtosis are annualized before they are plotted as a function of the time to maturity. The means of the basic and the adjusted BS model are, by definition, constant whereas the MDN models are able to find slightly time-dependent behaviour. The implied standard deviation (volatility) per year estimated by the BS model (the adjusted BS model) is about 16% (17%). For the MDN models however, the volatility varies between 14% and 17% as a non-linear function of the time to maturity. The MDN models are thus capable of modeling the term structure of volatility. Finally, the models with a non-Gaussian risk-neutral density are able to model negative skewness and excess kurtosis which can not be captured by the BS model and the MDN(1) model. While the scaling behaviour of higher-order moments is fixed for the adjusted BS model, the MDN(2) model is more flexible in changing the shape of the risk-neutral density with the time horizon. For the first setup, the skewness becomes even positive and the excess kurtosis negative for long times to maturity.

\footnote{See also the results in [2].}
Figure 3: The annualized mean, standard deviation, skewness, and excess kurtosis of the risk-neutral densities estimated by the Black-Scholes model (solid line), the adjusted Black-Scholes model (dashed line), and the extended mixture density network models with one Gaussian (dotted line) and with two Gaussians (dash-dotted line) as a function of the time to maturity (in years).

5.3 Hedging Performance

The quality of an option pricing model is not only determined by its pricing accuracy but also by its hedging performance. The hedging performance is the ability of the model to replicate option positions by portfolios of stocks and bonds. In particular, we study the hedging performance of the models with respect to short positions in the call and put options of the test sets. Following the exposition in [13], we assume that we sell each option at time 0 and undertake the usual dynamic hedging strategy in stocks\footnote{The stocks are the underlying asset on which the options are written. In our setup the underlying asset is a stock index which can not be bought or sold directly. Nevertheless, we assume in this section that the index is a traded asset. In practice, one would use futures contracts on the index to replicate option positions.} and bonds to hedge the option over its lifetime. The difference between the option payoff and the combined value of the stock and bond positions at maturity is the hedging error of the model’s replicating portfolio for this particular option. By aggregation over all options, the hedging performance of the model is determined.

More formally, let $V_{\tau}$ denote the value of the replicating portfolio at time $\tau \in [0,t]$ which is given by

$$V_{\tau} = V^S_{\tau} + V^B_{\tau} + V^{C/P}_{\tau}$$

(18)

where $V^S_{\tau}$, $V^B_{\tau}$, and $V^{C/P}_{\tau}$ denote the value of the stocks, the bonds, and the call or put option of the portfolio, respectively. At $\tau = 0$, the option is sold at the market price
\[
\pi_{C/P}^C, \text{ i.e., } V_0^{C/P} = -\pi_{C/P}^C. \text{ Simultaneously, } \Delta_0^{C/P} \text{ shares of stock at price } S_0 \text{ are bought}^{27} \\
\text{where the option delta } \Delta_0^{C/P} \text{ denotes the derivative of the price of the call or put option} \\
(\text{according to the option pricing model}) \text{ with respect to the stock price } S_0^{28}. \text{ The initial} \\
\text{composition of the replicating portfolio is given by} \\
\] \[
\begin{align*}
V_0^S &= S_0 \Delta_0^{C/P} \\
V_0^{C/P} &= -\pi_0^{C/P} \\
V_0^B &= -(V_0^S + V_0^{C/P})
\end{align*}
\]
\[\text{with } V_0 = 0. \text{ The stock purchase is thus wholly financed by the combination of riskless} \\
\text{borrowing and proceeds from the sale of the option. Since the price of the stock, i.e., the} \\
\text{value of the underlying index is changing over the lifetime of the option, the portfolio must} \\
\text{be rebalanced to hedge the option position. This is done on a daily basis by changing the} \\
\text{stock and bond positions so as to satisfy the following relations:}
\] \[
\begin{align*}
V_t^S &= S_t \Delta_t^{C/P} \\
V_t^B &= e^{\tau \Delta_t^B} V_{t-\Delta_t} - S_t(\Delta_t^{C/P} - \Delta_t^{C/P})
\end{align*}
\]
\[\text{where } \tau \in [\Delta_t, t] \text{ with } t = k \Delta_t \text{ for some integer } k \text{ and } \Delta_t \text{ is the time period of one day.} \\
\text{The hedging error of the portfolio is defined to be the value of the portfolio at maturity,} \\
i.e., V_t. \text{ From this, the discounted absolute hedging error}
\] \[
\xi = e^{-\tau l} |V_t|
\]
\[\text{is calculated as a performance measure of the model. As for the pricing errors, the} \\
\text{performance measures } \xi \text{ of all options of all test sets are collected.}
\]
\[\text{In Table 6 the mean and the standard deviation of } \xi \text{ are reported for all models. The basic} \\
\text{and the adjusted BS model exhibit practically the same hedging performance with an average} \\
discounted absolute hedging error of about 19.0 points. The second best model is the MDN(2) model} \\
\text{with an error of roughly 19.4 points which tends to be (is) significantly worse than the} \\
\text{performance of the basic (adjusted) BS model}^{29}. \text{ The model} \\
\text{with the largest error (19.6 points) is the MDN(1) model. This error is significantly larger} \\
than the error of the basic and the adjusted BS model.}
\]

\[
\begin{array}{|l|c|c|}
\hline
\text{Model} & \text{Error measure } \xi & \\
& \text{Mean} & \text{Std.} \\
\hline
\text{BS} & 18.99 & 24.26 \\
\text{BS-A} & 19.01 & 23.45 \\
\text{MDN(1)} & 19.60 & 24.86 \\
\text{MDN(2)} & 19.39 & 23.53 \\
\hline
\end{array}
\]

Table 6: Summary statistics of the out-of-sample hedging performance of the Black-Scholes model (BS), the adjusted Black-Scholes model (BS-A), and the extended mixture density network models with one Gaussian (MDN(1)) and with two Gaussians (MDN(2)).

\[\text{For simplicity, it is assumed that fractions of shares can be traded.}\]

\[\text{This derivative can be calculated analytically for all models.}\]

\[\text{The } p\text{-value for the } t\text{-test with the basic BS model is 0.053.}\]
If we concentrate on the two best models for a moment, we may conclude that higher pricing accuracy does not necessarily lead to better hedging performance. We recall that the adjusted BS model outperformed the basic BS model with respect to pricing accuracy for all classes of options. This behaviour of the models is, however, not inconsistent since pricing accuracy and hedging errors are two different performance measures and the models are optimized only with respect to the former one.

For the MDN models, the reason for the worse hedging performance is more profound. For any time to maturity, the parameters of the mixture distribution are modeled by MLPs, which can implement any functional relationship in general\textsuperscript{30}. This means that the parameters of the mixture distribution may be economically not plausible for times to maturity the model has not been trained on. In particular, the means $\mu_{i,t}$ and the variances $\sigma_{i,t}^2$, and therefore the mean and the variance of the risk-neutral density do not necessarily converge to zero as the time to maturity goes to zero\textsuperscript{31} since all options with a time to maturity of less than two weeks are excluded from the training set (see Section 3). Another consequence is that the deltas $\Delta_{C/P}^\tau$ of the options (according to the MDN models) do not necessarily converge to the correct values as the time to maturity $t - \tau$ approaches zero. For instance, a short position of one call option which is in-the-money at maturity $t$, is hedged by holding exactly one unit of the underlying asset\textsuperscript{32}. Therefore, $\Delta_{C}^\tau \to 1$ as $\tau \to t$ is the correct behaviour in this case. We find that the trained MDN models correctly hedge the option portfolio in many cases. In general however, there is no guarantee that the deltas come close to the correct values on the day of maturity.

In Table 7 the hedging errors of the models are summarized for the different classes of option contracts. We find a similar pattern as for the APE, namely that the error increases with the moneyness and with the time to maturity\textsuperscript{33}. Due to the unreliable hedging of options by the MDN models for short times to maturity, we perform significance tests only for the basic and for the adjusted BS model. From the perspective of moneyness, the differences in performance are not significant. However, the basic BS model achieves a significantly lower error for ST and MT options whereas the adjusted BS model performs significantly better for LT options. For the single classes, the differences in performance are only significant for ST-ATM, LT-OTM, and LT-ATM options.

6 Conclusions and Outlook

We have presented a new framework for extracting risk-neutral densities from observed option prices which is based on an extension of the concept of mixture density networks. Our new model yields significantly better out-of-sample results in terms of pricing accuracy than the standard Black-Scholes model and a Black-Scholes model adjusting the skewness and kurtosis terms. This holds not only for the complete set of options but also for smaller subsets determined by moneyness (out-of-the-money, at-the-money, in-the-money) and by time to maturity (short, medium, long). Both the non-Gaussian specification of the risk-neutral density and the non-linear dependence of mean, variance, skewness, and kurtosis on the time to maturity are attractive features of the model.

\textsuperscript{30} In our framework the functionality of the MLPs is slightly restricted, e.g., to avoid negative variances.

\textsuperscript{31} For the basic and the adjusted BS model, these economically reasonable properties are guaranteed by the linear specification of the mean and the variance (see Section 2).

\textsuperscript{32} We assume that the call option gives the owner the right to buy exactly one unit of the underlying asset.

\textsuperscript{33} The only exception are the classes of ST-ATM and ST-ITM options.
<table>
<thead>
<tr>
<th>Moneyness</th>
<th>Model</th>
<th>Time to Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ST</td>
</tr>
<tr>
<td>OTM</td>
<td>BS</td>
<td>8.87</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>8.95</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>9.48</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>9.43</td>
</tr>
<tr>
<td>ATM</td>
<td>BS</td>
<td>11.80</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>12.50</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>12.50</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>12.78</td>
</tr>
<tr>
<td>ITM</td>
<td>BS</td>
<td>10.82</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>10.71</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>11.05</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>10.87</td>
</tr>
<tr>
<td>All</td>
<td>BS</td>
<td>10.66</td>
</tr>
<tr>
<td></td>
<td>BS-A</td>
<td>11.08</td>
</tr>
<tr>
<td></td>
<td>MDN(1)</td>
<td>11.30</td>
</tr>
<tr>
<td></td>
<td>MDN(2)</td>
<td>11.43</td>
</tr>
</tbody>
</table>

Table 7: Summary statistics of the out-of-sample hedging performance for the options of the short-term (ST), medium-term (MT), and long-term (LT) maturity classes and the out-of-the-money (OTM), at-the-money (ATM), and in-the-money (ITM) classes for the Black-Scholes model (BS), the adjusted Black-Scholes model (BS-A), and the extended mixture density network models with one Gaussian (MDN(1)) and with two Gaussians (MDN(2)).

In terms of hedging, the performance of the extended mixture density network models is not satisfactory. The exclusion of options with a short time to maturity from the training set which is the standard approach in literature to avoid the influence of large price fluctuations toward the end of a contract, leads to estimates of the risk-neutral density for short times to maturity which are economically not plausible. A straightforward way to improve the models in this direction is to restrict their functionality appropriately. For instance, one could remove the bias terms in the multi-layer perceptrons which model the means and the variances of the mixture distribution. Since the non-linear activation function of the hidden units is the hyperbolic tangent, this would yield means and variances equal to zero for a network input, i.e., time to maturity equal to zero.

A direction of ongoing research is the comparison of extended mixture density network models and other state-of-the-art option pricing models such as GARCH models [8]. The results obtained so far are very promising. For instance, the average absolute pricing error of the GARCH model is 6.84 points whereas it is only 6.33 points for the extended mixture density network model with two Gaussians. According to a paired t-test, the difference in performance is significant ($p = 0.001$). Additionally, option prices can be calculated with negligible time effort compared to the exhaustive Monte Carlo simulations of the GARCH model.

It is also planned to refine the presented concept by including other factors. For instance, we will analyze the effects of adding the moneyness of an option to the input of an extended mixture density network model on the pricing and hedging performance of the model. Finally, risk management is an important area of application.
Acknowledgments

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 authors want to thank A. Lehar for providing the data set. They are grateful to A. Lehar and M. Scheicher for valuable discussions.

References


