GLOBAL ACADEMIC RESEARCH INSTITUTE

COLOMBO, SRI LANKA

International Journal of Management (IJM)

ISSN 2424-6433

Volume: 01 | Issue: 01

On 31st March 2016

http://www.research.lk

Author: Tea Poklepović, Zdravka Aljinović, Ante Rozga

Faculty of Economics, University of Split, Croatia

GARI Publisher | Financial Economics | Volume: 01 | Issue: 01

Article ID: IN/GARI/ICBME/2016/208 | Pages: 172-195 (23)

ISSN 2424-6433 | Edit: GARI Editorial Team

Received: 03.02.2016 | Publish: 31.03.2016

Moments Extraction from Implied Probability Distribution: Nonstructural Approach

Tea Poklepović (tpoklepo@efst.hr)
Faculty of Economics Split, Department of Quantitative Methods
Cvite Fiskovića 5, 21000 Split, Croatia

Zdravka Aljinović
Faculty of Economics Split, Department of Quantitative Methods
Cvite Fiskovića 5, 21000 Split, Croatia

Ante Rozga
Faculty of Economics Split, Department of Quantitative Methods
Cvite Fiskovića 5, 21000 Split, Croatia

Abstract

Quantifying economic uncertainty, i.e. an environment in which little or nothing is known about the future state of the economy, poses a challenge to both market participants and public authorities since it is not observable, but rather latent variable. In spite of this, it is possible to measure it indirectly through standard deviation of stock returns and implied volatility obtained from option prices. However, option prices may reveal considerable information beyond implied volatility, i.e. a set of option prices with the same maturity but with different exercise prices can be used to extract the entire implied probability distribution of the underlying asset at the expiration date. Implied probability distribution is risk-neutral according to the absence of arbitrage, i.e. the prices of options are not affected by degree of risk aversion among investors. The models for estimating risk-neutral density (RND) can be divided into two main categories: structural and nonstructural. The purpose of this paper is to extract market expectations from option prices and to investigate which of the proposed nonstructural models for estimating RNDs give the best fit. Namely, mixture of two log-normals, Edgeworth expansions and Shimko's model, i.e. representatives of parametric, semiparametric and nonparametric approaches respectively, will be compared. The model that fits sample data better than others is used to describe different characteristics (moments) of the *ex ante* probability distribution. The sample covers one-year data for DAX index options from July, 15th 2014 to July, 15th 2015. The results are compared through models and different maturity horizons.

Keywords: market expectation, risk-neutral density, mixture of log-normals, Edgeworth expansions, Shimko's model, maturity horizon, DAX index options.

JEL classification: C14, C58, G1

Introduction

Economic uncertainty refers to an environment in which little or nothing is known about the future state of the economy (Bloom et al, 2013). It poses a challenge to both market participants and public authorities in quantifying uncertainty, since it is not observable, but rather latent variable. However, it is possible to measure it indirectly. Some of the market based measures include standard deviation of stock returns and implied volatility obtained by option prices. Therefore, extracting information from financial prices is essential to gauge the uncertainty. Market participants, both small individual investors and big institutional investors, are interested in forecasting future volatility, since it is an important variable in portfolio management. Public authorities, especially central banks, are interested in understanding market expectations concerning future developments of various financial and monetary variables such as stock prices, exchange rate, interest rate and inflation. Central banks use information from market expectations when formulating and implementing monetary policy. Therefore, derivatives markets, especially futures and option prices, provide a rich source of information for gauging the market sentiment due to their forward-looking nature.

Most commonly, market expectations of future interest rates, for instance, have been extracted from the term structure of interest rates or from futures contracts on money-market instruments and bonds (Vähämaa, 2005). However, these measures reflect only central expectations, and hence provide no indication about the dispersion of market expectations, i.e. they do not provide market's assessment of the level of uncertainty, nor the asymmetries in the risk assessment. Moreover, an estimate of the standard deviation, as a risk measure, is obtained from data on stock returns however observations over several months are required if the skewness and kurtosis are to be measured accurately (Gemmill and Saflekos, 2006). Consequently, the focus has recently started to shift to information contained in option prices. Extracted information is mostly concentrated on implied volatility, i.e. volatility computed from a certain option pricing model. For example, one can invert Black-Sholes option pricing model to extract implied volatility using observable option prices in the market. Implied volatility obtained from option prices is considered to be a useful forward-looking measure of market uncertainty, and is therefore widely used among market participants and public authorities to assess the uncertainty surrounding central expectations. However, option prices may reveal considerable information beyond implied volatility (Vähämaa, 2005), i.e. a set of option prices with the same maturity but with different exercise prices can be used to extract the entire probability distribution of the underlying asset price at the maturity of the option. Therefore, in this paper the option prices are used to assess the so called "implied probability distribution". The result is the *ex ante* probability distribution that market participants would have expected if they were risk neutral. This means that the estimated implied probability distribution does not take into account the degree of risk aversion of investors. The risk-neutral probability distribution and the associated risk-neutral density function (RND) can describe different characteristics (moments) of the *ex ante* probability distribution, which includes mean, standard deviation, skewness and kurtosis. The implied probability distribution can be interpreted as the market assessment of the future probability distribution for the underlying asset on which the options are issued (Aguilar and Hördahl, 1999), where the variations over time in the implied moments provide a good indication of changes in the market's assessment of future developments in the underlying asset.

All of the techniques for estimating RND functions from option prices are related to the result of Breeden and Litzenberger (1978) and Banz and Miller (1978). They showed that a set of option prices with the same maturity but with different exercise prices can be used to extract the entire probability distribution of the underlying asset at the maturity of the option. According to Jondeau et al (2007) the models for estimating RNDs can be divided into two main categories: structural and nonstructural. Structural approaches propose a full description for the stock price dynamics, but they are rarely used due to the large number of unknown parameters that should be estimated. Nonstructural approaches yield a description of the RND without completely describing the dynamic of price. The nonstructural models can be parametric, semiparametric and nonparametric. Parametric models propose a direct expression for the RND, without referring to any price dynamic. They include mixture of distributions and a wide collection of distributions allowing for higher moments, including Black-Sholes Merton (BSM) model, mixture of two log-normals (MLN) and generalized beta distribution (GB2). Semiparametric and nonparametric models propose some approximation of the true RND. Most commonly used semiparametric models include Edgeworth expansions and Hermite polynomials, while nonparametric models include spline methods, tree-based methods, maximum entropy principle and kernel regression that all rely on Shimko model. However, they do not try to give explicit form for the RND, and "the data is left to speak for itself". Therefore, in this paper nonstructural models will be explained in detail and used in empirical research. Namely, representatives of parametric, semiparametric and nonparametric models, i.e. mixture of two log-normals (MLN), Edgeworth expansions (EE) and Shimko model (SM) respectively, are compared to find which one describes the "best" RND for prediction of *ex ante* moments of DAX index.

Since there is no consensus in the literature about the "best" model for RND estimation, the goal of this study is to compare the performances of three fundamental models for option pricing. It should be noted that this paper is written as a part of a greater project and it is relying on findings of Arnerić et al (2015) where different parametric approaches are compared and tested within different maturity horizons. This paper uses the main findings of the previous paper, i.e. uses the best parametric approach in modelling RND and compares it to the most commonly used semiparametric and nonparametric models. Due to different features of proposed models, they are extremely sensitive to different maturities of call and put options on underlying asset. Therefore, the purpose of this paper is to investigate if forecasting ability between proposed nonstructural models differs significantly considering different maturity horizon. Hence, the present paper contributes to the existing literature in a several ways. Firstly, it reveals at which maturity horizon the prediction of RND is most accurate given by the "best" nonstructural model. Second, it introduces the Diebold-Mariano test for comparison of the RND estimators. Third, it covers the most recent period which have not been in the center of research interest. The research is conducted in order to extract the implied distribution from option prices using the best nonstructural model and consequently to explain the implications of movements in implied moments for market participants and public authorities. Therefore, investors are able to apply the chosen RND estimator and obtain the "most accurate" implied probability distribution.

The remainder of the paper is organized as follows. Section 2 presents the literature review. Section 3 describes the methodology. Section 4 presents the data used in the empirical research, Section 5 presents obtained empirical results and discussion. Finally, some conclusions and directions for future research are provided in Section 6.

Literature review

Most commonly used approach for extraction of market expectations rely on Breeden and Litzenberger (1978) results and require the existence of a continuum of European options with the same time to maturity on underlying asset, spanning exercise prices from zero to infinity, and with these assumptions and in the absence of market frictions, it is possible to imply the underlying risk-neutral distribution (RND). Breeden and Litzenberger (1978) first showed how the second partial derivative of the call-pricing function with respect to the exercise price is directly proportional to the RND function. However, this method is based upon the assumption that there exist traded options for many strikes, which is not the case in practice since options contracts are only traded at discretely spaced time points. Because of that, some approximation for the second derivative is necessary and more than one implied distribution can be included depending on the approximation chosen (Gemmill and Saflekos, 2006). Shimko (1993) proposes an alternative approach by interpolating in the implied-volatility domain instead of the call-price domain. He begins by fitting a quadratic relationship between implied volatility and exercise price. The Black-Scholes formula is then used to invert the smoothed volatilities into option prices. The main limitation of the above techniques is the need for a relatively wide range of exercise prices. This can be overcome by imposing some form of prior structure on the problem. One such prior is to assume a particular stochastic process for the price dynamics of the underlying asset. In other words, the basic probability distribution is log-normal, but it can jump up or down. This framework follows Ritchey (1990), who notes that a wide variety of shapes may be approximated with a mixture of log-normal distributions.

Since then, in order to overcome the disadvantages of previous models, various techniques, methods and approaches that put more structure into the option prices have been developed. Hence, various structural and nonstructural models have been developed and studied in the literature. Due to the large number of unknown parameters included, structural models are rarely used for RND estimation. Nonstructural models can be divided into three categories: parametric, semiparametric and nonparametric models. Through the literature inspection, there are papers regarding semiparametric and nonparametric approaches for estimating RND (Jackwerth and Rubinstein, 1996; Andersen and Wagner, 2002; Tavin, 2011; Smith, 2012; Breeden and Litzenberger, 2014; Datta et al, 2014; Malz, 2014), some of them comparing parametric and nonparametric approaches (Bahra, 1997; Aparicio and Hodges, 1998; Figlewski, 2009; Mizrach, 2010) and only few are dealing with parametric approaches (Malz, 1997; Bookstaber and McDonald, 1987; Aguilar and Hördahl, 1999; Söderlind, 2000; Gemmill and Saflekos, 2000; Vähämaa, 2005; Cheng, 2010; Markose and Alentom, 2011; Grith and Krätschmer, 2011; Khrapov, 2014; Arnerić et al, 2015). It should be noted that literature lacks papers comparing representatives of nonstructural models, especially in the sense of comparing those models and finding the "best" fit model for different maturity horizons.

Jackwerth and Rubinstein (1996) derive underlying RND of European options on S&P 500 index by using a nonparametric approach. This is the method in which the risk-neutral probabilities are recovered from contemporaneous market prices of associated derivatives, and which makes no prespecification of the functional relations. The empirical research is based on database that contains all reported trades and quotes covering S&P European index options traded on CBOE, S&P 500 index futures traded on the CME, and intraday S&P 500 index levels from April 2, 1986, through December 31, 1993. A purpose of this article is to examine alternative specifications of the minimization criterion (other than quadratic) using historically observed option prices and to report and analyse the historical record of RNDs inferred from the nonparametric approach. The results identify a distinct change in shape between the precrash (log-normal) and the postcrash distributions (leprokurtosis and left-skewness), and also reveal that maximizing the smoothness of the resulting probability distribution is good objective for nonparametric methods.

Bahra (1997) gives an extensive overview of previous literature on estimation of RNDs. Using nonstructural approaches, and considering both advantages and disadvantages of each approach, the author estimates RNDs for LIFFE index options for 2 dates in March 1996 with time to expiration in June 1996 (103 and 105 days to expiration), for three-month sterling interest rate for 2 dates in May 1995 with expiration date in September 1995 (133 and 135 days to expiration) and for Philadelphia Stock Exchange (PHLX) currency options. Although there is an enormous contribution of this paper in synthesising various models for RND estimation, they are not compared to each other to provide evidence of superiority of a chosen method.

Aparicio and Hodges (1998) examine two alternative approaches to recovering the RND function from option prices. First approach is parametric which assumes that the future distribution of the underlying asset is GB2 distribution, and the second, nonparametric approach approximates the volatility smile using B-splines approximating functions and chain rule of differentiation. They use a sample of daily closing prices of the CME options on S&P 500 index future to estimate the implied volatility distributions and compare two different methods. The RNDs functions are calculated for September 1987 and September 1992 CME Future options data and two methods are compared by observing residuals between the observed implied volatilities and the estimated volatilities, which would indicate if there is a significant improvement in the goodness-of-fit of the nonparametric approach with respect to the parametric. The results show that nonparametric approach provides the best fit but satisfactory results can also be obtained with the GB2 approximation. Moreover, stability results reveal that the high level of negative autocorrelation of the innovation series obtained through the nonparametric approach may be an indication of overfitting the market data, picking up noise and pricing errors that later are reflected in the distributions. Parametric models with flexible distributions like GB2 model produce plausible distributions with the time series properties close to desirable random walk properties.

Aguilar and Hördahl (1999) present examples of estimated implied distributions for interest rates (options on three-month Eurolira interest rate on 3 dates in 1996 and 1997 with time to maturity of 77 days) and equities (Swedish OMX index on 4 dates in 1998 with time to maturity of 60 days) using a MLN method, and for currencies (Swedish krona –SEK/USD and SEK/DEM on 2 dates in 1998 and 1999 with one month to expiration) using Malz (1997) method. They conclude that implied probability distributions provide an indication of the market's assessment concerning the uncertainty of future events, which vary substantially over time. The authors give an overview of parametric methods used to estimate RND for interest rates and equities (MLN) and for currencies (Malz) and show their good potential. However, it does not compare different nonstructural approaches.

Söderlind (2000) study how market expectations about future UK monetary policy changed in period from June to November 1992 around events of ERM crisis. It extracts RNDs from daily option prices on the mark-pound exchange rate, 10-year sterling yield to maturity, 3-month sterling interest rate and 3-month mark interest rate, using MLN model and only arguing that it has to be better than BSM model, without referring to other nonstructural approaches.

Gemmill and Saflekos (2000) use option prices to construct RND to prove whether they are useful in forecasting market movements and/or revealing the investor sentiment. They estimate the implied distribution using a MLN approach and test its one-day ahead forecasting performance, using data on LIFFE's FTSE-100 index options from 1987 to 1997. The options used in this part of the analysis are chosen on one day per month (from the middle of the week) such that they have approximately 45 days to maturity. They found that the MLN performs much better than BSM approach at fitting the observed option prices, by comparing their RMSE (root mean squared errors), but it is only marginally better at predicting out-of-sample prices. Moreover, the analysis of extraordinary periods confirms that the shape of implied distribution does not anticipate these events, it rather reflects their passing. They conclude that implied distribution reflects market sentiment, however, that sentiment has little forecasting ability. Moreover, they do not compare the performances of other nonstructural approaches themselves or regarding different maturity horizons.

Andersen and Wagener (2002) extract RND from option prices by interpolating the Black-Scholes implied volatility smile, by fitting a smoothing spline of higher order polynomials with a relatively low number of knot points. They apply this to the LIFFE three-month Euribor futures option prices and calculate RNDs day by day for three months period from 20 March to 21 June 2001 and further analyse the impact of 11 September attacks on the expectation of future Euribor interest rate.

Vähämaa (2005) uses data on German government bond futures options to examine the behaviour of market expectations around monetary policy actions of the ECB and focuses on the asymmetries in bond market expectations, as measured by the skewness of option-implied probability distributions of future bond yields. By analysing the asymmetries in bond market expectations, this paper provides new evidence on how monetary policy affects financial markets. In this paper, option-implied probability distributions are extracted based on a parametric approach, i.e. MLN. This approach may be considered theoretically more competent as it ensures proper behaviour of the tail probabilities. The empirical analysis in this paper is performed using daily settlement prices of options on short-term German government bond futures traded at Eurex. The bond underlying the futures contract is Bundesschatzanweisungen (Schatz), a notional German government bond with 6% coupon and maturity of 1.75–2.25 years. The settlement prices for the Schatz futures and futures options are obtained from Eurex. The sample period used in the analysis extends from July 1999 to July 2003.

Jondeau et al (2007) give much newer overview of various methods for extracting RNDs using both structural and nonstructural methods, and parametric, semiparametric and nonparametric approaches. They also compare several techniques to estimate RNDs using data on CAC40 options. The first one is the numerical approximation of RND based on Breeden and Litzemberger (1978) approach, then with GB2 model, then using a MLN approach and finally using EE around the log-normal distribution. All of them are compared to BSM. The first method gives unstable results, unlike three other that yield with good approximations indicating negative skewness and fat tails. The MLN approach provides the best fit measured by

MSE (mean squared errors) and ARE (absolute relative errors). Although it uses different approaches and compares them, it does not provide any information about maturity of options and other information important for estimation purposes.

Cheng (2010) uses double log-normal approach and multi log-normal approach to extract RNDs for various asset classes including commodities, S&P 500, Dow-Jones index, euro/dollar exchange rate and the US 10-year treasury note and for different maturities. Data is collected on March 24-25, 2010. Monte Carlo study suggests that the multi log-normal approach outperforms the double log-normal using SSE to compare them. The results indicate that quadruple model is better than triple, which is in most cases better than double log-normal distribution. However, there might be problems of finding optimal solution for models with large number of parameters and one may consider choosing more parsimonious model.

Mizrach (2010) examines a variety of methods for extracting implied probability distributions from option prices and the underlying asset. The paper first explores nonparametric procedures for reconstructing densities directly from options market data. It then considers local volatility functions, both through implied volatility trees and volatility interpolation. It then turns to alternative specifications of the stochastic process for the underlying. This paper estimates a MLN, applies it to exchange rate data (US Dollar/British Pound (US$/BP) option which trades on the Philadelphia exchange from 1992 to 1993), and illustrate how to conduct forecast comparisons. Finally, this paper turns to the estimation of jump risk by extracting bipower variation.

Tavin (2011) presents how to obtain the RND of an underlying asset price as a function of its option implied volatility smile. It derives a known closed form nonparametric expression for the density and decompose it into a sum of log-normal and adjustment terms. By analysing this decomposition this paper also derives two no-arbitrage conditions on the volatility smile. The methodology is applied first to the pricing of a portfolio of digital options in a fully smile-consistent way. It is then applied to the fitting of a parametric distribution for log-return modelling, the Normal Inverse Gaussian. The underlying asset is DJ Eurostoxx 50, it is an equity index composed of 50 large Eurozone companies, and the date is 20.12.2007.

Markose and Alentom (2011) argue that the use of the Generalized Extreme Value (GEV) distribution to model the RND provides a flexible framework that captures the negative skewness and excess kurtosis of returns, and also delivers the market implied tail index. The GEV based option pricing model successfully removes the in-sample pricing bias of the BSM model, and also shows greater out of sample pricing accuracy, while requiring the estimation of only two parameters. They explain how the implied tail index is efficacious at modelling the fat tailed behaviour and negative skewness of the implied RND functions, particularly around crisis events. The data used in this study are the daily settlement prices of the FTSE 100 index call and put options published by the London International Financial Futures and Options Exchange (LIFFE). The period of study is from 2-Jan-1997 to 1-Jun-2009.

Grith and Krätschmer (2011) provide an overview of parametric models for RND estimation and uses paired European call options written on the underlying DAX index which mature in one month (21 days) and strike prices observed on 21.1.2004. Parametric models include BSM, MLN and generalized gama distribution, however they do not provide the conclusion about the best fit model and do not compare different nonstructural approaches.

Datta et al (2014) investigate the informational content of option-implied probability density functions (PDFs) for the future price of oil. They use daily data on

West Texas Intermediate (WTI) crude oil futures and options on WTI futures traded on New York Mercantile Exchange between January 2000 and December 2013, around episodes of high geopolitical tensions, oil disruptions and macroeconomic data releases, to explore the extent to which oil price movements are expected or not, and whether agents believe these movements are persistent or temporary. Using a semiparametric variant of the methodology of Breeden and Litzenberg (1978) they investigate the fit and smoothness of distributions derived from alternative PDF estimation methods and develop a set of robust summary statistics. They implement four options-implied volatility curve-fitting methods: cubic spline, fourth-order spline, smoothing spline and one knot spline, and the results are in favour of the smoothing spline method. The other problem is to estimate the tail of the distribution, and the results are in favour of extrapolating the implied volatility curve using first order polynomial.

Malz (2014) describes a method for computing RNDs based on the option-implied volatility smile. The aim is to reduce complexity and provide guidance through the estimation process. The technique is robust and avoids violations of option no-arbitrage restrictions that can lead to negative probabilities and other implausible results. Paper provides examples for equities (3-month options on the S&P 500 index, as of 21.12.2012), foreign exchange (1-month options on EUR-USD, as of 31.12.2012), and long-term interest rates (2-year options on 10-year swaps for 1.5.2013 and 5.9.2013).

Given a cross-section of daily option prices, Khrapov (2014) finds the RNDs, or equivalently the set of parameters, that minimize the option pricing errors using a MLN and a GB2 distributions. This estimation is repeated every Wednesday from 10.1.1996 to 25.7.2012 for VIX options with maturities up to one year to collect the time series of density parameters. It turns out that these parameters are highly predictable. The author exploits this predictability by fitting a simplest autoregressive model to each series of parameters and then forecasts these parameters out-of-sample and computes option prices corresponding to the forecast of the RND. The results are compared using RMSE and IVRMSE for different maturities. It can be concluded that MLN always outperforms the BSM model, whereas for GB2 nothing is concluded.

Arnerić et al (2015) investigate which of the parametric models for extracting RND, i.e. BSM, MLN and GB2 model, gives the best fit. The empirical findings indicate that the best fit is obtained for short maturity horizon, but when comparing models in short-run, the MLN gives statistically significant smaller MSE.

The abovementioned papers provide an extensive literature review, detailed description of the methods chosen and clear start for further investigation. Some of the papers only give a theoretical proof of advantages of one over the other method; most of them do not even compare the models or do it improperly, even sometimes giving an advantage to less parsimonious models. Moreover, different maturities are taken into account in some of the papers, but nothing is concluded regarding that with regard to different non-structural models. Since previous researches only conclude that none of the nonstructural model is clearly superior to the others, different approach is needed. Hence, this paper relies on findings of Arnerić et al (2015) where different parametric approaches are compared and tested within different maturity horizons, however in this paper different nonstructural approaches are explained, analysed and compared for extracting implied RNDs. Moreover, different models are compared regarding different maturities in order to conclude which method has the best forecasting ability in different time horizons.

Methodology

In this paper nonstructural models are used to infer the RND from option prices: mixture of two log-normals (MLN), Edgeworth expansions (EE) and Shimko's model (SM), i.e. representatives of parametric, semiparametric and nonparametric models respectively. Structural models are not observed as they are rarely used for estimating RND's due to the large number of unknown parameters.

The option pricing model based on the log-normal assumption with constant variance across exercise prices and across maturities has been developed by Black and Scholes (1973) and Merton (1973). Log-normal assumption originates from a geometric Brownian motion (GBM), which means that the price of the underlying asset is log-normally distributed variable, and its returns are normally distributed. Based on these assumptions the price of a European call option can be written as:

$$c = SN\left(d_1\right) - Xe^{-r(T-t)}N\left(d_2\right). \tag{1}$$

The price of a European put option can be written as:

$$p = SN\left(d_1\right) - Xe^{-r(T-t)}N\left(d_2\right). \tag{2}$$

or based on put-call parity, if the price of call option is known, as:

$$p = Xe^{-r(T-t)} - S + c, \tag{3}$$

where

$$d_1 = \frac{\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{(T-t)}}, \qquad d_2 = d_1 - \sigma\sqrt{(T-t)} \tag{4}$$

and where c is the price of European call option, p is the price of European put option, S is the price of the underlying asset, X is the strike price or the exercise price, r is the risk-free interest rate, $(T-t)$ is the time to expiration date, σ is the standard deviation of returns, i.e. volatility of the underlying asset, and $N(\cdot)$ denotes the standard cumulative normal distribution. Therefore, the price of the call option is a function of five variables, whose values are observable, except for the volatility.

To be able to price an option it is required to estimate the future value of volatility. It can be extracted from the Black-Scholes formula if the price of the option on market is observed, and it is known as implied volatility. In the same way as the implied volatility can be found from the option prices, the whole distribution of the underlying asset can also be estimated from option prices, and it is known as implied probability distribution. From implied probability distribution the moments that describe different characteristics of the distribution can be computed. These moments include mean (expectation), standard deviation (variance), skewness and kurtosis.

If a call option is $c = e^{-r(T-t)} \int_0^{\infty} \max\left(S_T - X, 0\right) q\left(S_T\right) dS_T$, based on Breeden and Litzenberger (1987), the risk neutral density is given by differentiating this equation twice, with respect to exercise price X

$$q\left(S_T\right) = e^{r(T-t)} \frac{\partial^2 c}{\partial X^2}. \tag{5}$$

Using simultaneously observed call and put option prices with the same maturity but with different exercise prices, the parameter vector $\theta \in \Theta$ can be estimated by minimising the sum of squared deviations between the observed (c_j^*, p_j^*) and theoretical prices (c_j, p_j) as:

$$\min_{\theta \in \Theta} \sum_{j=1}^{m} \left(c_j - c_j^*\right)^2 + \sum_{j=1}^{n} \left(p_j - p_j^*\right)^2. \tag{6}$$

The log-normal assumption does not hold in practice, i.e. implied volatilities of options on different exercise prices are not the same, nor are volatility smiles for different maturities. Therefore, it is required to use more general option pricing model than BSM. One alternative is to use the mixture of log-normal (MLN) densities, since it can be seen as an extension of the BSM model that uses only one log-normal density. The prices of European call and put options at time t can be written as:

$$c = e^{-r(T-t)} \int_X^{\infty} q\left(S_T\right)\left(S_T - X\right) dS_T, \tag{7}$$

$$p = e^{-r(T-t)} \int_0^{X} q\left(S_T\right)\left(X - S_T\right) dS_T, \tag{8}$$

where $e^{-r(T-t)}$ is the discount factor based on risk-free interest rate r and $q\left(S_T\right)$ denotes the risk-neutral density (RND) function for the price of the underlying asset on the expiration date T. From the previous equations it can be seen that using observed option prices, one should be able to extract the markets' estimate of the probability distribution $q\left(S_T\right)$. It should be noted that, in theory, any density functional form $q\left(S_T\right)$ can be used in equations (7) and (8), however this paper assumes that a mixture of two log-normal distributions is suitable to describe the underlying distribution for $\left(S_T\right)$:

$$q\left(S_T\right) = \theta L\left(\alpha_1, \beta_1, S_T\right) + \left(1 - \theta\right) L\left(\alpha_2, \beta_2, S_T\right), \tag{9}$$

where

(

$$L\left(\alpha_i, \beta_i, S_T\right) = \frac{1}{S_T \beta_i \sqrt{2\pi}} e^{\left(-\frac{(\ln S_T - \alpha_i)^2}{2\beta_i^2}\right)}, i = 1, 2, \quad (10)$$

and where θ is the weighting parameter that determines the relative influence of two log-normal distributions on the terminal distribution. Parameters α_i and β_i indicate location and dispersion for each log-normal distribution, which determine the mean and variance of the distributions according to:

$$\mu_i = e^{\alpha_i + \frac{\beta_i^2}{2}} \quad (11)$$

and

$$\sigma_i^2 = e^{2\alpha_i + \beta_i^2}\left(e^{\beta_i^2} - 1\right). \quad (12)$$

Inserting equations (9) and (10) into equations (7) and (8), the values of the call and put options can be expressed as:

$$c = e^{-r(T-t)} \int_X^\infty \left[\frac{\theta}{S_T \beta_1 \sqrt{2\pi}} e^{\left(-\frac{(\ln S_T - \alpha_1)^2}{2\beta_1^2}\right)} + \frac{1-\theta}{S_T \beta_2 \sqrt{2\pi}} e^{\left(-\frac{(\ln S_T - \alpha_2)^2}{2\beta_2^2}\right)} \right] \left(S_T - X\right) dS_T \quad (13)$$

$$p = e^{-r(T-t)} \int_X^\infty \left[\frac{\theta}{S_T \beta_1 \sqrt{2\pi}} e^{\left(-\frac{(\ln S_T - \alpha_1)^2}{2\beta_1^2}\right)} + \frac{1-\theta}{S_T \beta_2 \sqrt{2\pi}} e^{\left(-\frac{(\ln S_T - \alpha_2)^2}{2\beta_2^2}\right)} \right] \left(X - S_T\right) dS_T \quad (14)$$

Using at least five simultaneously observed call and put option prices with the same maturity but with different exercise prices, the five parameters (θ, α_1, β_1, α_2, β_2) can be estimated by minimising the sum of squared deviations between the observed prices (c_j^*, p_j^*) and theoretical prices (c_j, p_j) as:

$$\min_{\theta, \alpha_1, \beta_1, \alpha_2, \beta_2} \sum_{j=1}^{m} \left(c_j - c_j^*\right)^2 + \sum_{j=1}^{n} \left(p_j - p_j^*\right)^2 \quad (15)$$

Moreover, according to Bahra (1997), additional information can be exploited by including the forward price as an additional variable in the minimisation problem, where the forward price is the mean of the risk-neutral distribution, given by:

$$F\left(t, T\right) = \theta e^{\alpha_1 + \frac{\beta_1^2}{2}} + \left(1 - \theta\right) e^{\alpha_2 + \frac{\beta_2^2}{2}} \quad (16)$$

This yields with:

$$\min_{\theta,\alpha_1,\beta_1,\alpha_2,\beta_2} \sum_{j=1}^{m}\left(c_j - c_j^*\right)^2 + \sum_{j=1}^{n}\left(p_j - p_j^*\right)^2 + \left(\theta e^{\alpha_1+\frac{\beta_1^2}{2}} + (1-\theta)e^{\alpha_2+\frac{\beta_2^2}{2}} - e^{-r(T-t)}S\right) \quad (17)$$

subject to

$$\begin{aligned}&\beta_1, \beta_2 > 0\\ &0 \le \theta \le 1\end{aligned}. \quad (18)$$

Since evaluating equations (13) and (14) is computationally demanding due to the definite integral incorporated in these equations, the minimisation problem in (17) can be solved using closed-form solutions to equations (13) and (14) given as:

$$c = e^{-r(T-t)}\left\{\theta\left[e^{\alpha_1+\frac{\beta_1^2}{2}}N(d_1) - XN(d_2)\right] + (1-\theta)\left[e^{\alpha_2+\frac{\beta_2^2}{2}}N(d_3) - XN(d_4)\right]\right\} \quad (19)$$

$$p = e^{-r(T-t)}\left\{\begin{aligned}&\theta\left[-e^{\alpha_1+\frac{\beta_1^2}{2}}N(-d_1) - XN(-d_2)\right] +\\ &(1-\theta)\left[-e^{\alpha_2+\frac{\beta_2^2}{2}}N(-d_3) - XN(-d_4)\right]\end{aligned}\right\} \quad (20)$$

where

$$\begin{aligned}&d_1 = \frac{-\ln X + \alpha_1 + \beta_1^2}{\beta_1}, d_2 = d_1 - \beta_1\\ &d_3 = \frac{-\ln X + \alpha_2 + \beta_2^2}{\beta_2}, d_4 = d_3 - \beta_2\end{aligned}. \quad (21)$$

The closed-form solution uses the cumulative normal distribution rather than the log-normal density function which eases the numerical computation. Based on Liu et al (2004) the standard deviation, skewness and kurtosis of a mixture can be derived from:

$$E\left[S_T^n\right] = \theta\alpha_1^n e^{\frac{1}{2}\left(n^2-n\right)\beta_1^2 T} + (1-\theta)\alpha_2^n e^{\frac{1}{2}\left(n^2-n\right)\beta_2^2 T} \quad (22)$$

The other alternative is to use Edgeworth expansion (EE) method developed by Jarrow and Rudd (1982), whose idea is to capture deviations from log-normality by an Edgeworth expansion of the RND $q\left(S_T \mid \theta\right)$ around the log-normal density. The expansion approach has the advantage that the approximation, by involving parameters that can vary, allows generating more functions. It is a representative of semiparametric approach in modelling RND (Jondeau et al, 2007).

Let Q be the cumulative distribution function (*cdf*) of a random variable S_T and q its density. Characteristic function of S is defined as $\phi_Q(u) \equiv \int e^{isu} q(s)\,ds$. If moments of S_T exist up to order *n*, then cumulants of the distribution *Q*, denoted $\kappa_{Q,j}$ exist, implicitly defined by the expansion:

$$\log\left(\phi_Q(u)\right)=\sum_{j=1}^{n-1}\kappa_{Q,j}\frac{(iu)^j}{j!}+o\left(u^{n-1}\right) \tag{23}$$

Thus, if the characteristic function $\phi_Q(\cdot)$ is known, by taking an expansion of its logarithm around *u=0*, it is possible to obtain the cumulants. There is the following relationship between the cumulants and moments up to the fourth order, i.e. first four cumulants are equivalent to mean, variance, skewness and kurtosis respectively.

$$\begin{gathered}\kappa_{Q,1}=E\left[S_T\right],\ \kappa_{Q,2}=V\left[S_T\right],\kappa_{Q,3}=E\left[\left(S_T-E\left[S_T\right]\right)^3\right],\\ \kappa_{Q,4}=E\left[\left(S_T-E\left[S_T\right]\right)^4\right]-3V\left[S_T\right]^2.\end{gathered} \tag{24}$$

EE of the fourth order for the true probability distribution *Q* around the log-normal *cdf L* can be written, after imposing that the first moment of the approximating and true density are equal, ($\kappa_{Q,1}=\kappa_{L,1}$), and by denoting densities with small letters:

$$\begin{aligned}q(s)=l(s)&+\frac{\left(\kappa_{Q,2}-\kappa_{L,2}\right)}{2!}\frac{d^2l(s)}{ds^2}-\frac{\left(\kappa_{Q,3}-\kappa_{L,3}\right)}{3!}\frac{d^3l(s)}{ds^3}\\ &+\frac{\left(\kappa_{Q,4}-\kappa_{L,4}\right)+3\left(\kappa_{Q,2}-\kappa_{L,2}\right)^2}{4!}\frac{d^4l(s)}{ds^4}+\varepsilon(s)\end{aligned} \tag{25}$$

where $\varepsilon(s)$ captures terms neglected in the expansion. The various terms in the expansion correspond to adjustments of the variance, skewness and kurtosis, and the interpretation of the expansion is similar to a Taylor expansion. Moreover, under this approximated density, the price of a European call option with strike *K* can be approximated as in equation (26), where the first term is the BSM formula:

$$\begin{aligned}C(Q)&=e^{-r\tau}\int_K^\infty\left(S_T-K\right)q\left(S_T\right)dS_T\approx e^{-r\tau}\int_K^\infty\left(S_T-K\right)l\left(S_T\right)dS_T\\ &+e^{-r\tau}\frac{\left(\kappa_{Q,2}-\kappa_{L,2}\right)}{2!}\int_K^\infty\left(S_T-K\right)\frac{d^2l\left(S_T\right)}{dS_T{}^2}dS_T\\ &-e^{-r\tau}\frac{\left(\kappa_{Q,3}-\kappa_{L,3}\right)}{3!}\int_K^\infty\left(S_T-K\right)\frac{d^3l\left(S_T\right)}{dS_T{}^3}dS_T\\ &+e^{-r\tau}\frac{\left(\kappa_{Q,4}-\kappa_{L,4}\right)+3\left(\kappa_{Q,2}-\kappa_{L,2}\right)^2}{4!}\int_K^\infty\left(S_T-K\right)\frac{d^4l\left(S_T\right)}{dS_T{}^4}dS_T\end{aligned} \tag{26}$$

Additionally, the log-normal distribution has the following property:

$$\int_K^\infty (S_T-K)\frac{d^j l(S_T)}{dS_T{}^j}dS_T=\left.\frac{d^{j-2}l(S_T)}{dS_T{}^{j-2}}\right|_{S=K},\ for\ j\geq 2 \tag{27}$$

We deduce for the call option price:

$$\begin{aligned}C(Q)\approx C(L)&+e^{-r\tau}\frac{(\kappa_{Q,2}-\kappa_{L,2})}{2!}l(K)-e^{-r\tau}\frac{(\kappa_{Q,3}-\kappa_{L,3})}{3!}\frac{dl(K)}{dS_T}\\&+e^{-r\tau}\frac{(\kappa_{Q,4}-\kappa_{L,4})+3(\kappa_{Q,2}-\kappa_{L,2})^2}{4!}\frac{d^2l(K)}{dS_T^2}\end{aligned} \tag{28}$$

For the log-normal density, the first cumulants are given by:

$$\begin{aligned}&\kappa_{L,1}=S_t e^{r\tau},\kappa_{L,2}=[\kappa_{L,1}\vartheta]^2,\kappa_{L,3}=[\kappa_{L,1}\vartheta]^3(3\vartheta+\vartheta^3),\\&\kappa_{L,4}=[\kappa_{L,1}\vartheta]^4(16\vartheta^2+15\vartheta^4+6\vartheta^6+\vartheta^8),\end{aligned} \tag{29}$$

where $\vartheta=\left(e^{\sigma^2\tau}-1\right)^{1/2}$ and where the first relation follows from risk-neutral valuation. Additionally, second moment can be identified by imposing $\kappa_{Q,2}=\kappa_{L,2}$. Rather than estimating $\kappa_{Q,3}$ and $\kappa_{Q,4}$ it is possible to estimate standardized skewness and kurtosis ($\gamma_{Q,1}$ and $\gamma_{Q,2}$ respectively), which are defined as:

$$\gamma_{Q,1}=\frac{\kappa_{Q,3}}{(\kappa_{Q,2})^{3/2}}=3\vartheta+\vartheta^3,\gamma_{Q,2}=\frac{\kappa_{Q,4}}{(\kappa_{Q,2})^2}=16\vartheta^2+15\vartheta^4+6\vartheta^6+\vartheta^8. \tag{30}$$

These expressions also hold for the log-normal density, and therefore, skewness and kurtosis of the log-normal density can be derived from the above cumulants. With the assumption of equality of the second cumulants for the approximating and the true distribution, it follows:

$$C(Q)\approx C(L)-e^{-r\tau}(\gamma_{Q,1}-\gamma_{L,1})\frac{\kappa_{L,2}^{3/2}}{3!}\frac{dl(K)}{dS_T}+e^{-r\tau}(\gamma_{Q,2}-\gamma_{L,2})\frac{\kappa_{L,2}^2}{4!}\frac{d^2l(K)}{dS_T^2} \tag{31}$$

Using this expression, it is easy to estimate with nonlinear least squares the implied volatility (σ^2), skewness ($\gamma_{Q,1}$) and kurtosis ($\gamma_{Q,2}$). The RND can be obtained after twice differentiating equation (31) with respect to *K* and then evaluation over S_T :

$$q(S_T)\approx l(S_T)-(\gamma_{Q,1}-\gamma_{L,1})\frac{\kappa_{L,2}^{3/2}}{6}\frac{d^3l(S_T)}{dS_T^3}+(\gamma_{Q,2}-\gamma_{L,2})\frac{\kappa_{L,2}^2}{24}\frac{d^4l(S_T)}{dS_T^4}, \tag{32}$$

where the partial derivative can be computed iteratively using:

$$\frac{dl(S_T)}{dS_T}=-\left(1+\frac{\log(S_T)-m}{\sigma^2\tau}\right)\frac{l(S_T)}{S_T},$$
$$\frac{d^2l(S_T)}{dS_T^2}=-\left(2+\frac{\log(S_T)-m}{\sigma^2\tau}\right)\frac{1}{S_T}\frac{dl(S_T)}{dS_T}-\frac{1}{S_T^2\sigma^2}l(S_T),$$
$$\frac{d^3l(S_T)}{dS_T^3}=-\left(3+\frac{\log(S_T)-m}{\sigma^2\tau}\right)\frac{1}{S_T}\frac{d^2l(S_T)}{d^2S_T}-\frac{2}{S_T^2\sigma^2}\frac{dl(S_T)}{dS_T}+\frac{1}{S_T^3\sigma^2}l(S_T), \quad (33)$$
$$\frac{d^4l(S_T)}{dS_T^4}=-\left(4+\frac{\log(S_T)-m}{\sigma^2\tau}\right)\frac{1}{S_T}\frac{d^3l(S_T)}{d^3S_T}-\frac{3}{S_T^2\sigma^2}\frac{d^2l(S_T)}{d^2S_T}$$
$$+\frac{3}{S_T^3\sigma^2}\frac{dl(S_T)}{dS_T}-\frac{2}{S_T^4\sigma^2}l(S_T)$$

and where $m=\log(S_T)+\left(r-\sigma^2/2\right)\tau$. Those computations indicate that the RND in the Edgeworth case will be a polynomial whose coefficients directly command the skewness and kurtosis of the RND.

Nonparametric models do not try to give an explicit form of the RND. Representative of nonparametric model is Shimko model (SM). Shimko (1993) propose to implement the results of Breeden and Litzemberger (1978) directly after a preliminary smoothing of the volatility smile. Since a direct estimation leads to numerically unstable results, the idea of SM is to summarize the information contained in the volatility smile via a more or less sophisticated polynomial, for instance $\sigma(K)$ a function of strike price *K* and then to use this expression to evaluate the density. In other words, the function $\sigma(K)$ is fitted to the various volatilities. Outside the range of quoted strikes, the volatility is supposed to be constant. Focusing on a call option, formally it is written that $C(S_t,K,\tau,r,\sigma)=C(S_t,K,\tau,r,\sigma(K))$.

The first idea (Shimko, 1993) is to use a quadratic polynomial for $\sigma(K)$. Formally,

$$\sigma_i=a_0+a_1K_i+a_2K_i^2,\ \text{for } i=1,...,n, \quad (34)$$

where *n* represents the number of observed prices. The parameters of this polynomial can be easily estimated using a nonlinear least square regression. The RND is given by:

$$d_1(K)=\frac{\log\left(\frac{S}{Ke^{r\tau}}\right)}{\sigma(K)\sqrt{\tau}}+\frac{\sigma(K)\sqrt{\tau}}{2},d_2(K)=d_1(K)-\sigma(K)\sqrt{\tau}, \quad (35)$$
$$\sigma(K)=\left(a_0+a_1K+a_2K^2\right)1_{\left(K>\min(K_i)\ and\ K<\max(K_i)\right)}+\sigma_1 1_{\left(K\le\min(K_i)\right)}+\sigma_M 1_{\left(K\ge\max(K_i)\right)}$$

where the function $1_{(A)}$ is the indicator function taking the value 1 if A is true. The first- and second-order derivatives of d_1 and d_2 are given by:

$$\begin{aligned}
&d_1'(K) = -\frac{\sigma'(K)\sqrt{\tau}}{\sigma^2(K)\tau}\log\left(\frac{S}{Ke^{r\tau}}\right) - \frac{1}{K\sigma(K)\sqrt{\tau}} + \frac{\sigma'(K)\sqrt{\tau}}{2},\\
&d_2'(K) = d_1'(K) - \sigma'(K)\sqrt{\tau},\\
&d_1''(K) = -\frac{\sigma''(K)\sigma(K)\tau - 2\sigma'(K)^2\tau}{\left(\sigma(K)\sqrt{\tau}\right)^3}\log\left(\frac{S}{Ke^{r\tau}}\right) + \frac{\sigma'(K)\sqrt{\tau}}{K\sigma^2(K)\tau} +\\
&+\frac{\sigma(K)\sqrt{\tau} - K\sigma'(K)\sqrt{\tau}}{K^2\sigma^2(K)\tau} + \frac{\sigma''(K)\sqrt{\tau}}{2},\\
&d_2''(K) = d_1''(K) - \sigma''(K)\sqrt{\tau},\\
&\sigma'(K) = (a_1 + 2a_2K)1_{(K>\min(K_i)\ and\ K<\max(K_i))},\\
&\sigma''(K) = 2a_2 1_{(K>\min(K_i)\ and\ K<\max(K_i))}.
\end{aligned} \tag{36}$$

The characteristics of nonparametric approaches is their independence of the assumptions, which is viewed as their great strength and a weakness. It is a strength since it induces the structure of the problem from the data, rather than presuming complex models from which prices are deduced. However, there is no guarantee that the prices obtained from nonparametric models are going to be in accordance with rational pricing. Particularly in Shimko model it is not rare to obtain negative probabilities. Moreover, nonparametric models are found to be data-intensive, requiring large datasets and cannot be practically implementable due to limited number of quoted bid and ask option prices.

Data analysis

Data sets includes averages of the last bid and the last ask options prices, both for calls and puts on the same strikes, in period from July 2014 to July 2015. However, the selection of dates to include is not set by chance. After the observation of DAX index movements (Figure 1), and before the research is conducted, it was important to select dates at which the estimation would be performed. Dates around peeks and bottoms are selected. Moreover, each selected date corresponds to the third Friday of each month, i.e. expiration day of option prices. In the observed period from July 2014 to July 2015, DAX index was on average 10412.94 points, reaching its minimum value of 8571.95 in October 2014, and its maximum value of 12374.73 in April 2015. Time series of DAX index prices is presented by Figure 1. Moreover, the RND were calculated for one and two months in advance, i.e. time to expiration is both 28 (and 35) and 63 (or 56) days depending on number of days in a month. Therefore, RNDs are calculated for 11 months, for 2 different maturity horizons and with 3 different nonstructural models, yielding with 66 different RNDs. Three models will be compared to conclude which method has the best forecasting accuracy in different time horizons.

It should be noted that when estimating the implied probability distribution, it is important to select the type of option prices. In empirical researches, there are often used prices collected at the end of the trading day or the average of the bid and ask price. Moreover, it is common to use the out-of-the money options, because of their greater liquidity. The same is done in this research. The research is conducted in "R" software using "RND" package.

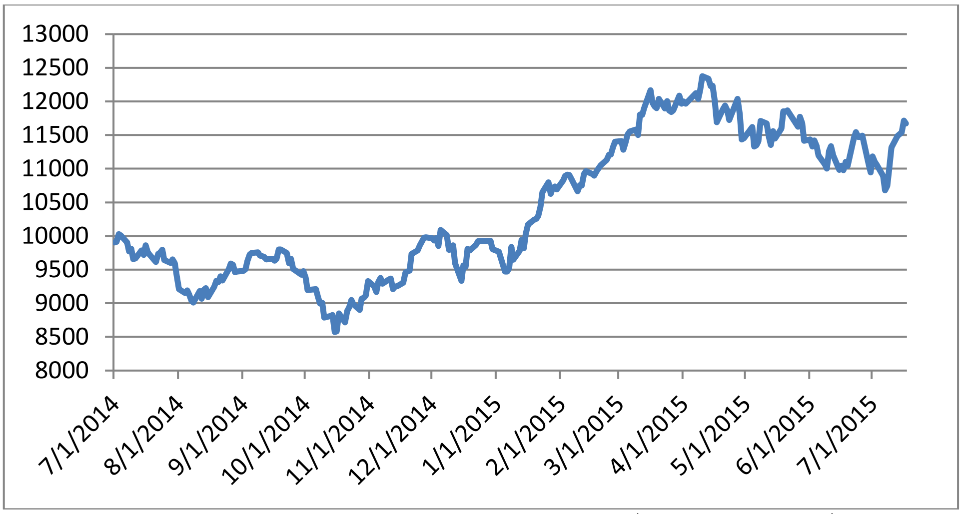

Figure 1 - Closing prices of DAX index from July, 1st 2014 to July, 1st 2015
Source: http://finance.yahoo.com/q/hp?s=%5EGDAXI+Historical+Prices [25.07.2015]

Empirical results

In this paper nonstructural models are used to infer the RND from option prices: mixture of two log-normals (MLN), Edgeworth expansions (EE) and Shimko's model (SM). Since these models are representatives of parametric, semiparametric and nonparametric models respectively, estimated parameters that have an economic interpretation and significance can be obtained only for MLN model. They are not presented and are available from the authors upon request.

Tables 1 - 3 present *ex ante* moments extracted from each RND, i.e. mean μ, standard deviation σ, skewness α_3 and kurtosis α_4 of DAX index prices at chosen expiration dates, i.e. 17th of October 2014, 16th of January 2015 and 19th of June 2015 respectively, and for 1 month and 2 months maturity horizons. In each Table 1 - 3 six RNDs have been estimated based on two different maturity horizons and using three nonstructural models. The same is done for 8 more expiration dates, but the results are not presented here due to the lack of space (results are available upon request). However, in Tables 1 - 3, and Figures 2 - 4 one can notice that there are differences between extracted three risk-neutral densities. Therefore, the purpose of this paper is to investigate which model and at which maturity horizon fits the future distribution of DAX index most accurately. As a comparison criterion, mean square error (MSE) is used. MSE was calculated as the mean square difference between observed call and put option prices and expected (theoretical) call and put option prices obtained from three nonstrustural models for the same strikes. Furthermore, Diebold-Mariano (DM) test is used to test which model has lower MSE. Model that has significantly lower MSE is considered as most appropriate.

Firstly, the testing procedure is computed to compare the performance of RND's extracted between 1 month and 2 months maturity horizons within each model. The results indicate that the short-run forecasts yield better results with smaller MSE, i.e. within each estimated model (MLN, EE and SM) the null hypothesis of DM test can be rejected in favour of one-sided alternative. The null hypothesis is that the two models have the same forecast accuracy (Diebold and Mariano, 1995).

Table 1 - RND's for 1 month and 2 months maturity horizons with ex ante moments extracted at expiration date 17.10.2014.

Model	MLN		EE		SM	
(T-t)	1m	2m	1m	2m	1m	2m
μ	9851.40	9048.34	9518.26	9041.60	9918.03	9256.14
σ	291.19	428.89	483.11	344.16	237.99	386.99
α_3	-0.06	0.62	0.12	-0.15	-0.06	-0.04
α_4	2.94	2.87	2.14	2.12	2.66	2.59
MSE	192.00	3472.00	7956.00	16500.00	741.80	6904.00
DM test	-10.16***		-5.35***		-3.93***	
MLN			-5.90***		-5.13***	
EE					-5.64***	

*Note: ***, **, * represent significance level of 1%, 5% and 10% at which the null hypothesis of Diebold-Mariano test (DM) is rejected.*

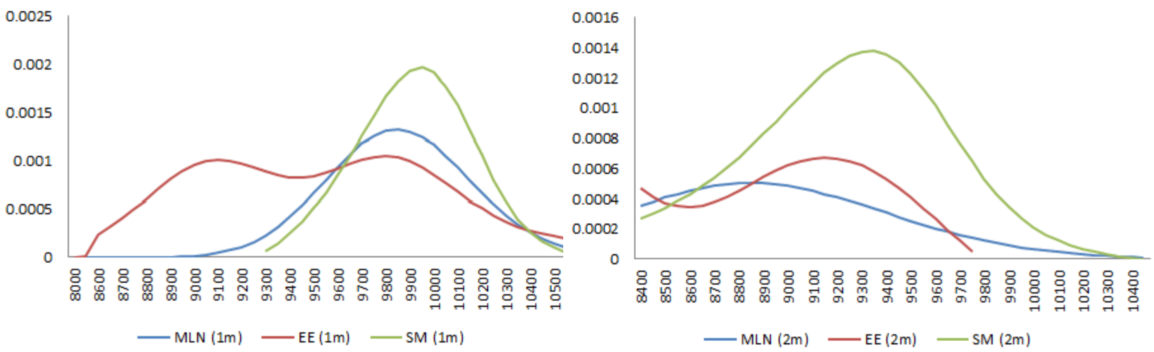

Figure 2 - Estimated RNDs for 1 month and 2 months maturity horizons at 17.10.2014.

Secondly, three proposed models are used for pairwise comparison (MLN vs. EE, MLN vs. SM and EE vs. SM) only in short-run as previously has been concluded that the short-run forecasts yield better results with smaller MSE. The results show superiority of MLN model in short run against EE and SM. The null hypothesis of DM test can be rejected, i.e. the short-run MLN forecasts have lower MSE than short-run forecasts of both EE and SM approaches. The same results were obtained and confirmed for 8 more expiration dates, but not presented here due to the lack of space. Moreover, both EE and SM model show odd results, yielding in some cases with negative probabilities and rather divergent expected call and put prices (the results are available from the authors upon request) as indicated already in the literature (Bahra, 1997) as a downside of these models.

Table 2 - RND's for 1 month and 2 months maturity horizons with ex ante moments extracted at expiration date 16.01.2015.

Model	MLN		EE		SM	
(T-t)	1m	2m	1m	2m	1m	2m
μ	9784.38	9598.07	9579.27	9402.93	9906.40	9686.48
σ	480.90	416.97	531.29	529.60	379.49	389.00

α_3	-0.86	-1.24	-0.28	-0.29	-0.19	-0.38
α_4	3.90	4.38	2.55	2.27	2.61	2.93
MSE	112.10	757.30	563.90	2151.00	111.40	994.90
DM test	-1.58*		-2.09*		-1.34*	
MLN			-10.50***		-0.05	
EE					-9.68***	

*Note: ***, **, * represent significance level of 1%, 5% and 10% at which the null hypothesis of Diebold-Mariano test (DM) is rejected.*

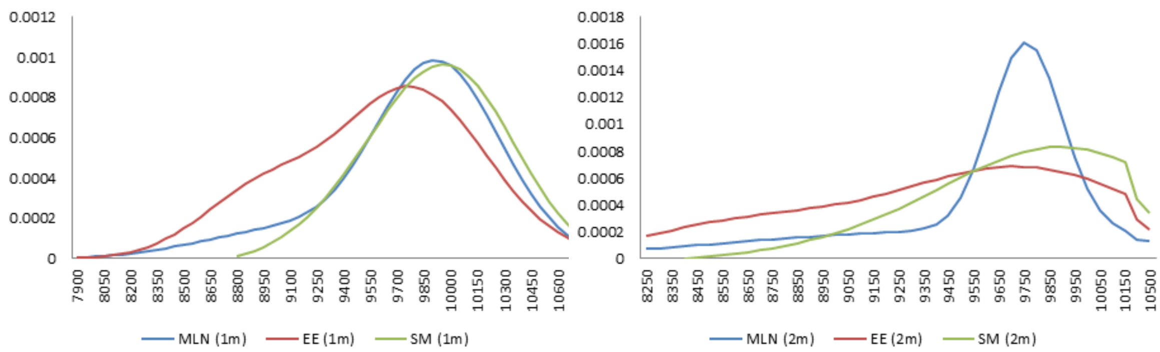

Figure 3 - Estimated RNDs for 1 month and 2 months maturity horizons at 16.01.2015.

Table 3 - RND's parameters estimates for 1 month and 2 months maturity horizons with ex ante moments extracted at expiration date 19.06.2015.

Model	MLN		EE		SM	
(T-t)	1m	2m	1m	2m	1m	2m
μ	11446.08	11667.88	11024.88	10821.68	11568.13	11863.52
σ	752.90	874.94	905.86	1226.60	493.04	675.53
α_3	-0.60	-0.92	-0.26	-0.08	0.07	-0.33
α_4	3.25	4.10	2.39	1.87	2.63	2.57
MSE	24.18	148.80	5620.00	41010.00	518.40	1543.00
DM test	-9.33***		-9.59***		-7.18***	
MLN			-11.53***		-12.22***	
EE					-10.60***	

*Note: ***, **, * represent significance level of 1%, 5% and 10% at which the null hypothesis of Diebold-Mariano test (DM) is rejected.*

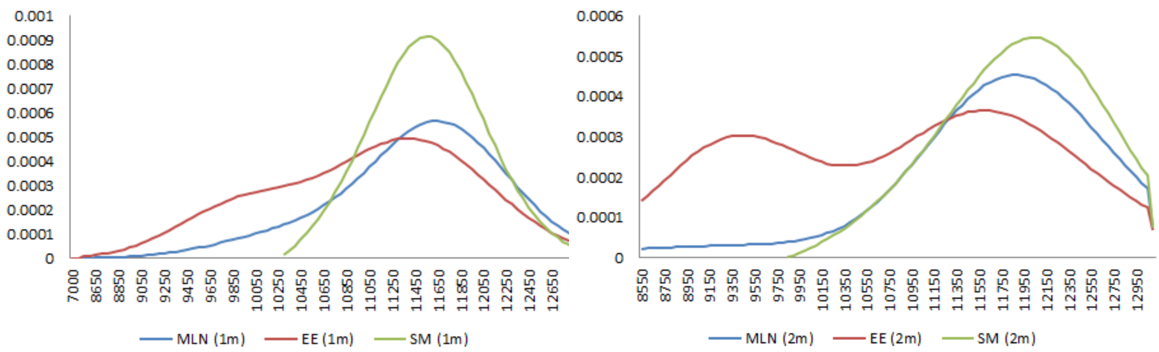

Figure 4 - Estimated RNDs for 1 month and 2 months maturity horizons at 19.06.2015.

Changes in the implied moments, extracted from RND's, between two successive time points should provide valuable information of changes in the market's assessment of future developments in the underlying asset. According to the MLN model that has most accurate predictive ability within one-month maturity horizon, three RND's are compared to describe these changes (Figure 5). For ease interpretation of changes in the market expectations of DAX index an implied (*ex ante*) moments at three chosen expiration dates are also presented in Table 4.

Table 4 - Implied (ex ante) moments extracted from estimated RND's at three chosen expiration dates (MLN model with maturity horizon of one month)

Implied moments	17.10.2014	16.01.2015	19.06.2015
μ	9851.40	9784.38	11446.08
σ	431.80	480.90	752.90
α_3	1.39	-0.86	-0.60
α_4	2.34	3.90	3.25

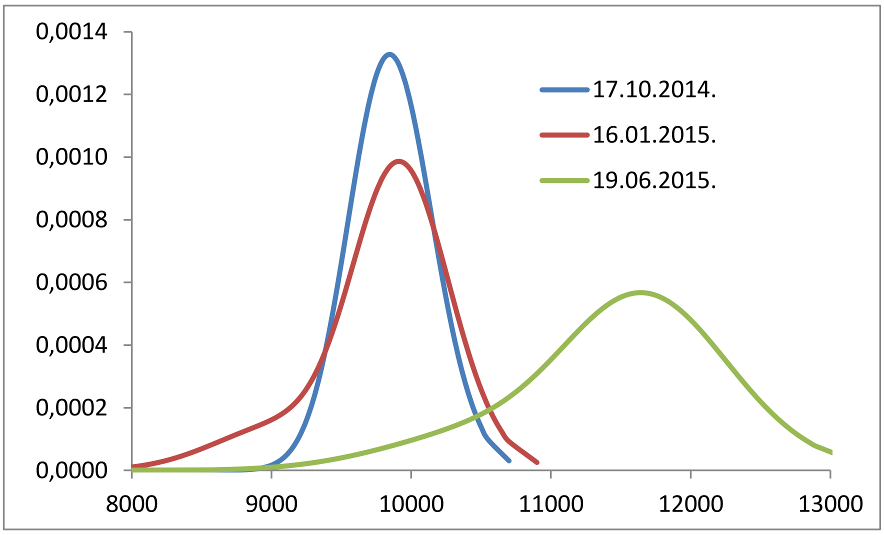

Figure 5 - RNDs at 3 expiration dates based on MLN model (maturity horizon 1m)

On third Friday in September, 2014 the value of DAX index was still recovering from its' plunge in August. Therefore, the market sentiment was optimistic at 17th of October, 2014 yielding with higher expected value of DAX index one month ahead, lower standard deviation (volatility) as an indication of smaller uncertainty and positive skewness which indicates that market perceives the probability of positive outcomes to be higher than the probability of negative outcomes. For expiration date 16th of January, 2015 results indicate similar value of the implied mean but higher implied standard deviation (volatility) with implied asymmetry concentrated on the left tail of distribution ($\alpha_3 < 0$). As implied kurtosis is increasing a distribution has heavier tails (Figure 5). It is important to note that the kurtosis is a measure of fat tails of the distribution, i.e. leptokurtic distribution has fat-tails ($\alpha_4 > 3$) regarding the curvature of distribution. The same changes in the implied moments can be notice at expiration date on 19th of June, 2015 which means that the value of DAX index is at the higher level with higher implied volatility. Distribution is also more negatively skewed on the left tail then on the right tail even both tails are heavier compared to the implied distribution at expiration date from 16th of January, 2015. This reveals that the market participants perceive great uncertainty with the development of the DAX index during the life of the options with higher probability of negative outcomes, i.e. the market sentiment was pessimistic.

Figure 6 presents all implied moments, i.e. mean, standard deviation, skewness and kurtosis for all expiration dates using MLN model for maturity horizon of one month. Expected values of DAX index move along with the observed values of DAX index (Figure 1). Standard deviation increases as the expected value increases, indicating lower risk in the beginning of the observed period and higher associated risk in the latter periods. Asymmetry is on average negative, indicating that market perceives the probability of negative outcomes to be higher than the probability of positive outcomes in the selected period. Moreover, leptokurosis and fat-taildness is observed in majority expiration dates of the sample period. Therefore, the extracted implied distribution reveals market sentiment, however it does not anticipate movements of DAX index, it rather moves along with it.

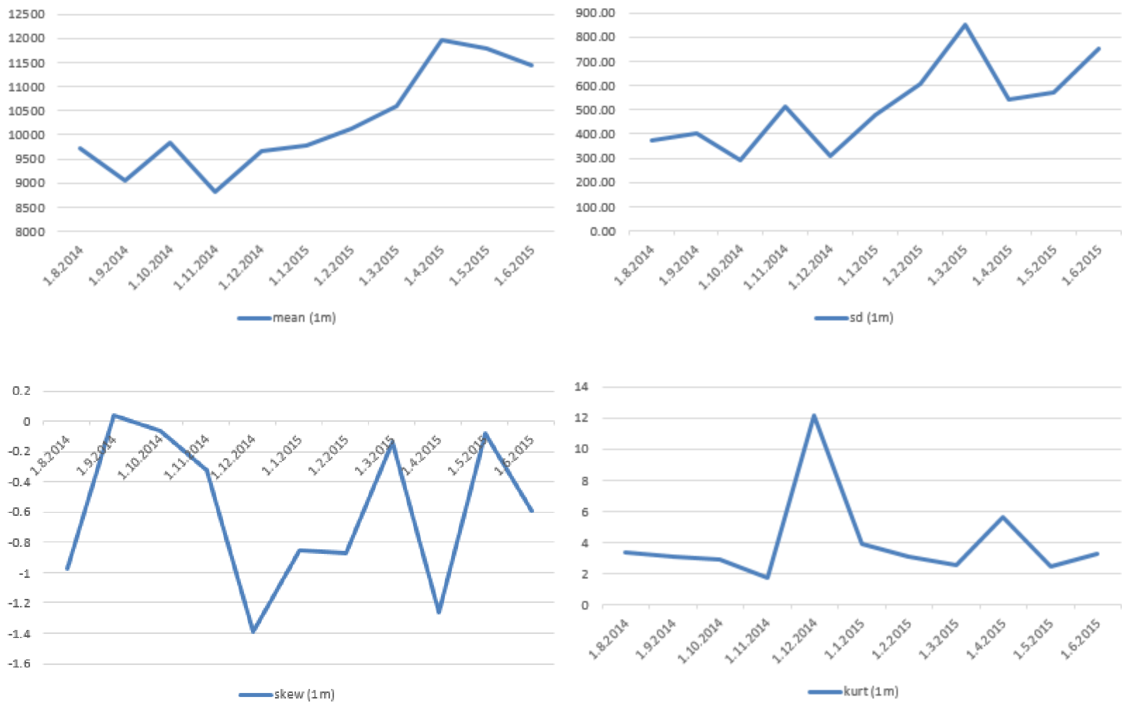

Figure 6 – Implied moments at all expiration dates based on MLN model for one-month maturity horizon

Conclusion

In this paper three nonstructural approaches for estimation of risk neutral density distribution, mixture of two log-normals, Edgeworth expansion around the log-normal distribution and Shimko model, which are representatives of parametric, semiparametric and nonparametric approaches respectively, are explained, estimated and compared using one-year data for options on DAX index. All three approaches have both their advantages and disadvantages. However, previous researches reveal that none of the approaches is clearly superior to the others. Therefore, this research is conducted in order to compare the three selected models in their forecasting ability and within the models the focus is put on comparison regarding different maturity horizons.

The results reveal that no matter which nonstructural model is used, all of them give better short-term forecasts. In pairwise comparison for short-term prediction, mixture of two log-normals approach is superior to the others according to the mean squared errors and Diebold Mariano test. Moreover, MLN model has proven to be very flexible, which means that it is possible to obtain a wide variety of different implied distributions, and therefore it can capture commonly observed characteristics of financial assets, such as asymmetries and “fat-tails” in implied probability distribution. The results also reveal how the implied moments and the RND function itself respond to the arrival of new information and how market assesses risk over time. Changes in the implied moments in observed sample period reveal that the value of DAX index is at the higher level with higher implied volatility and that implied (risk-neutral) distribution is more negatively skewed on the left tail then on the right tail even both tails are heavier compared to the implied distribution at previous expiration dates. Moreover, when the underlying index moves, the RND not only moves along with the index, but it also changes shape. Although the extracted implied distribution reveals market sentiment, it does not anticipate movements of DAX index.

Since, the value of true variance, skewness and kurtosis are unknown to us, i.e. cannot be observed, these implied higher order moments can only be compared with their realized counterparts. That would be worth for further research which requires intraday observations, i.e. high frequency data. Moreover, different methods can be compared within different markets since they differ in liquidity and number of quoted bid and ask option prices. Therefore. development of option trading on emerging markets can be viewed as a new niche in modeling market expectations.

Acknowledgement

This work has been fully supported by Croatian Science Foundation under the project “Volatility measurement, modeling and forecasting” (5199).

References

Aguilar, J. and Hördahl, P. (1999), “Option prices and market expectations”, *Sveriges Riksbank Quarterly Review*, Vol. 1, pp. 43-70.

Andersen, A.B. and Wagener, T. (2002), “Extracting risk neutral probability densities fitting implied volatility smiles: some methodological points and an application to the 3M Euribor futures option prices”, ECB Working Paper No. 198.

Aparicio, S.D. and Hodges, S. (1998), “Implied Risk-Neutral Distribution: A comparison of Estimation Methods”, http://www2.warwick.ac.uk/fac/soc/wbs/subjects/finance/research/wpaperseries/1998/98-95.pdf (5.10.2015).

Arnerić, J., Aljinović, Z. and Poklepović, T. (2015), “Extraction of Market Expectations from Risk-Neutral Density”, *Zbornik radova Ekonomskog fakulteta u Rijeci, časopis za ekonomsku teoriju i praksu - Proceedings of Rijeka Faculty of Economics, Journal of Economics and Business*, Vol. 33, No. 2, pp. 235-256.

Bahra, B. (1997), “Implied risk-neutral probability density functions from option prices: theory and application”, Bank of England, Working Paper, No. 66.

Banz, R.W. and Miller, M. (1978), “Prices for State-contingent Claims: Some Estimates and Applications”, *The Journal of Business*, Vol. 51, No. 4, pp. 653-672.

Black, F. and Scholes, M. S. (1973), “The Pricing of Options and Corporate Liabilities”, *Journal of Political Economy*, Vol. 81, pp. 637-659.

Bloom, N., Kose, M.A and Terrones, M.E. (2013), “Held Back by Uncertainty”, Finance & Development, Vol. 50, No. 1

Bookstaber, R.M. and McDonald, J.B. (1987), “A General Distribution for Describing Security Price Returns”, *The Journal of Business*, Vol. 60, No. 3, pp. 401-424.

Breeden, D.T. and Litzenberger, R.H. (1978), “Prices of State-Contigent Claims Implicit in Option Prices”, *The Journal of Business*, Vol. 51, No. 4, pp. 621-651.

Breeden, D.T. and Litzenberger, R.H. (2014), “Central Bank Policy Impacts on the Distribution of Future Interest Rates (February 20, 2014) http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2642363 (30.10.2015)

Cheng, K.C. (2010), “A New Framework to Estimate the Risk-Neutral Probability Density Functions Embedded in Option Prices”, IMF Working Paper, WP/10/181.

Datta, D.D., Londono, J.M. and Ross, L.J. (2014), “Generating Options-Implied Probability Densities to Understand Oil Market Events”, Board of Governors of the Federal Reserve System, International Finance Discussion Paper, Number 1122, October.

Diebold, F. X. and Mariano, R. S. (1995), “Comparing Predictive Accuracy”, *Journal of Business and Economic Statistics*, No. 13, pp. 253-263.

Figlewski, S. (2009), “Estimating the Implied Risk Neutral Density for the U.S. Market Portfolio”, Volatility and Time Series Econometrics: Essay in Honor of Robert F. Engle, Tim Bolloerslev, Jeffrey R. Russell, Mark Watson, Oxford, UK: Oxford University Press, 2008; NYU Working Paper No. FIN-08-004.

Gemmill, G. and Saflekos, A. (2000), “How useful are implied distributions? Evidence from stock index options”, *The Journal of Derivatives*, Vol. 7, No. 3, pp. 83-91.

Grith, M. and Krätschmer, V. (2011), “*Parametric Modelling and Estimation of Risk-Neutral Densities*”, *in Jin-Chuan Duan, James E. Gentle, and Wolfgang Härdle (eds) Handbook of Computational Finance. Springer Verlag.*

Jackwerth, J.C. and Rubinstein, M. (1996), “Recovering Probability Distribution from Option Prices”, *The Journal of Finance*, Vol. 51, No. 5, pp. 1611-1631.

Jondeau, E., Poon, S and Rockinger, M. (2007), “*Financial modelling under Non-Gaussian Distributions*”, Springer-Verlag, London.

Jarrow, R. and Rudd, A. (1982) “Approximate option valuation for arbitrary stochastic processes”, *Journal of Financial Economics*, Vol. 10, pp. 347-369

Khrapov, S. (2014), “Option pricing via Risk-neutral Density Forecasting”, 2nd International Workshop on “Financial Markets and Nonlinear Dynamics” (FMND), June 4-5, 2015, Paris (France) https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SNDE2015&paper_id=110 (16.9.2015)

Liu, X., Shackleton, M.B., Taylor, S.J. and Xu, X. (2004), “Closed-form transformations from risk-neutral to real-world distributions”, Essex Finance Centre Discussion Paper No. 04-05.

Malz, A.M. (1997), “Option-implied Probability Distributions and Currency Excess Returns”, Federal Reserve Bank of New York Staff Reports, No. 32.

Malz, A.M. (2014), „A Simple and Relieble Way to Compute Option-Based Risk-Neutral Distributions“, Fedreal Reserve Bank of New York Staff Reports, No. 677.

Markose, S. and Alentom, A. (2011), “The Generalized Extreme Value (GEV) Distribution, Implied Tail Indeks and Option Pricing“, *The Journal of Derivatives*, Vol. 18, No. 3, pp. 35-60.

Merton, R. C. (1973), “Theory of Rational Option Pricing”, *The Bell Journal of Economics and Management Science*, Vol. 4, No. 1, pp.141-183.

Mizrach, B. (2010), “*Estimating Implied Probabilities from Option Prices and the Underlying*“, Handbook of Quantitative Finance and Risk Management, pp. 515-529.

Ritchey, R.J. (1990), “Call option valuation for discrete normal mixtures”, *Journal of Financial Research*, Vol, 13, No. 4, pp. 285–296.

Shimko, D. (1993), “Bounds of probability”, *Risk*, Vol 6. pp. 33-37.

Smith, T. (2012), “Option-implied probability distributions for future inflation“, Bank of England Quarterly Bulletin 2012 Q3, pp. 224-233.

Söderlind, P. (2000), “Market expectations in the UK before and after the ERM crisis”, *Economica*, No. 67, pp. 1-18.

Tavin, B. (2011), “Implied distribution as a function of the volatility smile“, *Bankers Markets and Investors*, No. 119, pp. 31-42.

Vähämaa, S. (2005), “Option-implied asymmetries in bond market expectations around monetary policy actions of the ECB“, *Journal of Economics and Business*, No. 57, pp. 23–38.