Comparing Risk Neutral Density Estimation Methods using Simulated Option Data

Amine Bouden *

Abstract—**In this paper I use Monte Carlo simulated option data to investigate the empirical power of six Risk Neutral Density (RND) estimation techniques. Three alternative approaches are used for comparison and the choice of the most suitable method depends on: their performance to correctly price options, their capacity to fit the true density - which is estimated from the underlying asset series - and their ability to forecast the future realization of the underlying asset. I found that the decision depends on the purpose of the framework: for pricing purpose, the interpolation techniques would be adequate. However, if the aim is to extract market expectations, the Edgeworth expansion should be used.**

Keywords: Risk Neutral Density, Monte Carlo simulation, Probability Integral Transform

1 Introduction

Options traded on financial assets allow market participants to take views on the future values of the assets themselves. This projection towards the future confers to the option a sort of forward-looking information on the future evolution of the underlying assets. For example, in the case of Call option, the larger today's premium is, the higher is the probability to exert the option and greater is the likelihood that the asset price finishes above the strike price at maturity. These probabilities, taken together, form the so-called implied Probability Density Function also referred as Risk Neutral Density function (henceforth RND). More precisely, when the market operators evaluate theoretically an option price, they formulate their own estimates of these probabilities and their 'feelings' about future evolutions of the underlying asset are summarized in the RND. The reason why it is called 'Risk Neutral' density is that because it is estimated in a risk neutral world (in opposition with the real world) under which all investors are assumed to have no risk preferences. A relatively large number of estimation methods has been proposed in the empirical literature and almost all are based on one of the following approaches:

i. Assuming that the underlying asset follows a particular stochastic process and constructing the RND from the estimated process (Geometric Brownian Motion giving the lognormal distrbution as a RND).

ii. Fitting a particular parametric form for the density which would reflect as well as possible the market conditions (mixture of 2 lognormals or the Edgeworth expansion).

iii. Using interpolation techniques to smooth the Call pricing function. The RND is then obtained by numerical differentiation (2 interpolation methods in spaces (volatility / strike) and (volatility / delta)).

iv. Semi-parametric techniques based on implied binomial trees.

Given these alternative techniques, a question arises: 'what about their reliability?' Numerous comparison tools dealing with the stability and the accuracy of the estimated RND were considered: [1] tested the relative stability of the alternative RNDs and their robustness to small perturbations. [3] chose to deal with the pricing errors i.e. the performance to provide a good assessment of the observed option prices; while [6] studied the ability of the RND to closely recover the realized summary statistics. All these studies provided rather different conclusions concerning the most suitable method to adopt. For example, [4] proved the superiority of the non parametric techniques over the parametric ones using FTSE index option data whereas the study of [2] performed on the Norwegian option market found exactly the opposite result. According to this, a natural concern arises: the choice of a specific method seems to be related to the market data used. This problem would be resolved using unreal simulated data. Therefore, the simulation framework into which the comparison is carried out would not relate to any specific market. In this case, the best estimation method would be considered as the best ever.

To achieve this purpose, the attention is focused on the RND itself. Indeed, some authors argued that the RND corresponds to the real-world density[1] only if investors were risk neutral, and the difference between the two densities arises because of the existence of a risk premium[2]. But since simulated data are used, the risk premium is

---

*University of Paris 2 Pantheon-Assas - ERMES - CNRS, 12 Place Panthon 75 230 Paris Cedex 5, Tel: +33 6 81 76 16 49, Email: amine.bouden@u-paris2.fr

[1] Also called true density, physical density, subjective density or empirical density. In this study I will retain the term "true density".
[2] See [7] for more relevant discussion of this point.

null and the two densities should be the same. On the basis of this, two questions have to be addressed: how close is the estimated RND to the true density? And do the RND derived from option prices provide a good forecast of the true density? This paper attempts to address these concerns and six different estimation methods (which are the simple lognormal, the mixture of 2 lognormals, the Edgeworth expansion, 2 interpolation methods in spaces (volatility / strike) and (volatility / delta) and the implied binomial trees[3] are compared according to: their capacity to correctly price option prices, their ability to assess accurately the true density and their performance to correctly forecast future outcomes of the underlying asset.

The rest of the paper is organised as follows: The next section presents the simulation framework into which the estimation will be performed. Three different comparison tools are then presented and implemented in section 3. The paper concludes in a brief summary.

2 Monte Carlo approach to generate simulated data

For some financial markets, option contracts are not always available neither for all strike prices, nor for all maturities and therefore, some estimation methods of the RND are not easily applicable. For example, the non parametric techniques related to the volatility Smile interpolation are intensive in data and require the availability of a minimum number of options to facilitate the implementation of the interpolation. Therefore, it would be preferable to have a sufficient number of options that disables any data discrimination that may block the estimation process. That is why it is more useful to work with simulated option data rather than real data. Furthermore, disposing of sufficiently large number of data makes tests more robust and statistical indicators more relevant.

Before presenting the simulation procedure, a Data Generating Process (DGP) has to be chosen to simulate trajectories of the underlying asset. The question is which process to choose? Given its proprieties, the Heston stochastic volatility model, would be the best to encompass all observed market characteristics i.e. non constant volatility, asymmetries and fat tails. The underlying price dynamics are described by the following diffusion equations:

$$dS = \mu S dt + \sqrt{\sigma} S dW_1$$

$$d\sigma = \kappa(\theta - \sigma)dt + v\sqrt{\sigma}dW_2$$

Where μ is the drift of the underlying process, θ is the long-term mean of the volatility, κ governs the rate at which the volatility converges to its long-term value,

---

[3]See [4] for an overview of these techniques

v represents the variance of the volatility, dW_1 and dW_2 are Wiener processes correlated by a coefficient ρ: $dW_1 dW_2 = \rho\, dt$.

Note that the Heston's model has been subject of many studies in the empirical literature and largely employed to model the behavior of financial asset prices. This would be an additional motivation for this choice.

The parameters defining Heston model must be fixed in a logical sense before launching the simulations. I fix $r = 0.05$, $\kappa = 1$, $\theta = 0.37$, $v = 0.1$ and $\rho = 0.27$. The choice of these values is not arbitrary. Indeed, the rate of reversion and the variance of volatility, respectively 1 and 0.1, have been voluntarily selected low in order to limit the dispersion of the estimations. Intuitively, when its variance trends toward 0, the volatility becomes less random and one can expect the RND to have thin tails. Long-term volatility parameter $\theta = 0.37$ materializes the phenomenon of volatility clustering according to which high volatility periods succeed or precede low volatility periods, resulting in a median long-term value. Finally, correlation between the underlying returns and the volatility is positive and equal to 0.27. This means that as positive shocks occur, the likelihood for getting further large upward movements increases, resulting in a positive asymmetry in the RND with fat left tail and thin right tail.

Once the DGP chosen and its parameters defined, the simulation procedure can start.

A first simulation consists of generating 1 path of 11000 steps to up to 44 years which results in 11000 daily values of the underlying $S_0, S_{\Delta t}, S_{2\Delta t}..., S_{11000}$ and 11000 daily values of the volatility $\sigma_0, \sigma_{\Delta t}, \sigma_{2\Delta t}..., \sigma_{11000}$ (1 step equals to 1 day: $\Delta t = 0.004$). The starting values are set to $S_0 = 48$ and $\sigma_0 = 0.32$. This first simulation is referred as **Case A**. The first 499 observations are then eliminated from the sample in order to avoid distortions due to the initial assumptions of the simulation process. Thus, a first sample of 10501 observations of S ($S_{500}, S_{501}, ..., S_{11000}$) and σ ($\sigma_{500}, \sigma_{501}, ..., \sigma_{11000}$) is set up and will be used thereafter as a starting point for estimations and tests.

A second simulation gives a new sequence of the underlying asset $\{F_i\}_{i=1}^N$ as well as a new sequence of the volatility $\{V_i\}_{i=1}^N$. It consists of generating 1 step 5000 random paths using the same DGP and the same diffusion parameters as previously. This second simulation is referred as **Case B**. "Today's" values of both underlying asset and volatility are observed on the generated random path of **Case A** when the simulation is launched: $F_0 = S_{500}$ et $V_0 = \sigma_{500}$.

After that, Call option prices are numerically calculated using Monte Carlo integration; the Call evaluation formula becomes:

$$C = \frac{1}{N}e^{-r\Delta t}\sum_{i=1}^{N}\max(0, F_i - K) \tag{1}$$

Here, the maturity is fixed to 1 month which corresponds to 21 working days, $\Delta t = 0.004$ and $N = 5000$. A total of 9 Call options are calculated for 9 different strike prices. The 9 exercise prices are set equal to the average F_a of the 5000 simulated data F plus and minus 4 rates below and above this average, resulting in 1 at the money, 4 in the money and 4 out of the money options. The equally spaced exercise prices are: $K = (1 \pm x)F_a$ where $x = 2.5$, 5, 7.5 and 10 respectively.

The procedure described in **Case B** is recurred 500 times by moving forward by 21 days on the path simulated in **Case A**. For example, to calculate the second batch of Call prices, a new sequence of F is generated similarly as previously with initial values $F_0 = S_{521}$ and $V_0 = \sigma_{521}$. I finally obtain 500 times 9 Call option prices with fixed 1-month maturity. These prices will be used for the application of RND estimation techniques presented above.

3 Comparing the selected estimation techniques

In this section, the six selected methods are implemented and tested: the Black & Scholes simple lognormal distribution which will be considered as the benchmark, mixture of two lognormals, Edgeworth expansion, both interpolation methods in spaces (volatility / strike) and (volatility / delta) and the implied binomial tree method.

3.1 Comparing the pricing errors

One approach to assess the performance of the estimation techniques is to examine how closely they fit actual option prices and the method presenting the smallest pricing errors will be considered as the best to fit option prices. For each of the six RNDs, the theoretical Call option price C^{th} is calculated numerically from the formula (1) and compared to the simulated price C^{sim} derived from the **Case B** simulation. To assess the quality of the estimation as well as to appraise the model's goodness-of-fit, two criteria are computed each day: the first one is the well-known Root mean Squared Error (RMSE):

$$RMSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(C_i^{sim} - C_i^{th})^2}$$

The RMSE is a dimensionless metric which facilitates the comparison between various models for different contracts. Nevertheless, it is sensitive to great pricing errors, which could in some cases distort the interpretations: an error of 0.5 on a Call option equal to 8 does not have the same significance on another equal to 0.5. Whereas the error accounts for 6.25% of the first Call option, it is 100% for the second Call option. Thus, it would be necessary to put into perspective the error to the corresponding observation. The second criteria, the Relative

Root mean Squared Error (RRMSE), accounts for this asymmetry:

$$RRMSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(\frac{C_i^{sim} - C_i^{th}}{C_i^{sim}})^2}$$

The figures in table (1) are obtained first by calculating the RMSE and the RRMSE for each method and then, by calculating the mean and standard deviation (between brackets) over the 500 observations ($m = 500$).

It is quite obvious that the two statistics are rather concordant and lead to almost the same results. Indeed, according to the RMSE, interpolation methods (volatility / strike) and (volatility / delta) provide the most accurate estimates with values about 0.59×10^{-3} and 0.67×10^{-3} respectively. Then, with a RMSE of 1.22×10^{-3}, the Edgeworth approximation proves to be the best parametric technique with, in addition, a relatively low dispersion; follow after the simple lognormal and the mixture of two lognormals. Finally, the implied binomial tree method provides the worst pricing performance with a very large error value equal to 99.78×10^{-3}.

The analysis of the RRMSE provides a classification quasi similar with one exception: the pricing performance of the mixture of lognormals exceeds that of the lognormal distribution. This result seems more logical insofar as the second method constitutes a particular case of the first one. The bad pricing performance of the mixture of lognormals is rather surprising. The great dispersion (equal to 0.4975×10^{-3}) characterizing RMSE averages could be the reason: only few errors but with large magnitude generate a high RMSE coefficient. This problem is not likely to occur in the case of the RRMSE since each error is reported to the corresponding observation, which confers great robustness to this metric. The implied binomial tree method has a catastrophic evaluation performance with the highest RMSE and RRMSE coefficients. This result is due more to implementation irregularities than to the bad specification of the model: indeed, during the optimization procedure, all the constraints of the equation (9) were not filled and the algorithm did not always converge towards a final solution. In particular, the forth constraint, which stipulates that the theoretical Call price should be equal to the simulated price, was often violated.

Up to now, to assess the performance of the estimation methods, I used quantitative indicators that measure the deviation between simulated and theoretical option prices. But what would be the conclusions in a statistical universe? Indeed, the question whether these deviations are statistically significant can be addressed. The relevance of this question is reinforced since a model can provide with a good quality of adjustment which,

however, could be non-significant in a statistical sense. Thus, it would be interesting to carry out a statistical test to assess more accurately the deviations between theoretical and simulated option prices. The problem is formulated as follows:

$$C^{sim} = \alpha_0 + \alpha_1 C^{th} + \varepsilon$$

I separately apply this test to the six methods by replacing each time the theoretical Call option price by that provided by the tested method. I also carry out the Wald test of the null assumption $H_0 : \alpha_0 = 0$ and $\alpha_1 = 1$ according to which the theoretical option price is an unbiased estimate of the true option price. The number of parameter restrictions is equal to 2 and therefore, the statistic of this test follows the Chi-square distribution with 2 degrees of freedom: $\aleph^2(2)$. The critical values on 95% and 99% bands are respectively 5.99 and 9.21. A statistic higher (lower) than 5.99 implies the rejection (acceptance) of the null at a 5% confidence level. The results are presented in table (2):

Not surprisingly, the coefficients of determination are very close to 100% and the regression parameters α_0 and α_1 are very close to respectively 0 and 1. Nevertheless, the Wald test indicates that the null assumption is rejected for all the regressions and consequently, any of the six estimation methods provides satisfactory results from a statistical point of view.

All in all, we can say that non-parametric methods perform better than parametric and semi parametric methods. This result is not, though, confirmed by statistical tests. Naturally, this leads us to consider alternative comparison tools. In the following, I will focus on the RNDs themselves and study their profiles, which would constitute an important source of exploitable information.

3.2 Implied RND function vs. true density function

The RND is not the true probability density. Whereas the former is inferred from quoted option prices, the latter is not available and should be estimated from observable time series of the underlying asset. The difference between the two distributions arises from the presence of risk aversion: indeed, the RND can be interpreted as the market's aggregate beliefs regarding future states of the world and consequently, incorporates investor's attitude toward risk. The RND would be equivalent to the true market density only if there is no aggregate risk in the market or assuming risk neutrality.

Working with simulated data overcomes this problem: not only the actual density is available and directly derived from various simulated scenarios of the underlying asset, but also the RND integrates no formulation about risk aversion since it is estimated from simulated option data and not from real option data. Thus, the RND should be exactly equal to the true density function. Based on this assertion, one possibility of comparing RND estimation methods raises: which one would provide the most accurate estimation of the 'real' density function?

To answer this question, I estimate first the true density q_{true} from the 5000 simulated underlying prices F in the **Case B** and second, apply separately each of the six estimation methods on option prices to derive the RND q_{RND} .The estimation of q_{true} is carried out using Kernel estimator $K(u)$ to obtain a sufficiently smooth and continuous density function:

$$q_{true}(F) = \frac{1}{Nh} \sum_{i=1}^{N} K(\frac{F - F_i}{h})$$

Where $\int K(u)d(u) = 1$ to ensure q_{true} to be a probability density function and h is the bandwidth of the Kernel estimator. I choose to work with the Gaussian Kernel which is the most commonly used.

Crucial for the Kernel estimation is the choice of the local bandwidth; a too small one will lead to an erratic distribution whereas a too large one will smooth away important details in the distribution. For this, I follow [?] who proposes an optimal value for the bandwidth h_{opt}:

$$h_{opt} = 0.9N^{-\frac{1}{5}}A$$

Where $A = \min(\sigma, iqr/1,34)$, iqr being the interquartile range.

Figures (1a-f) display estimated RNDs and true densities for each of the estimation methods[4].

The graphics above underscore the following observations:

– All RNDs underestimate the left tail of the true density. Indeed, the left true density tail lies systematically below its RND counterpart. In a real world, this report would be interpreted so as the market does not expect any crash in the underlying asset and therefore assigns low probability to the occurrence of this event.

– Interpolation methods are the best to fit correctly the true density and the two densities are almost superposed.

– On the RND provided by interpolation method in a space (volatility / strike), some erratic movements are observable around the mode. Nevertheless, they

---

[4]Over a list of 500 observations of RNDs and true densities, those represented in the figure (1) correspond to the 131^e observation.

do not appear if the interpolation is performed in the (volatility / delta) space. So, we can say that moving from a (volatility / strike) space towards a (volatility / delta) space makes the RND smoother and avoids the appearance of abrupt jumps in the probabilities. This result is not surprising since the objective to move from the first to the second space is indeed to prevent such distortions.

– The implied binomial tree method and the mixture of lognormals provide completely shifted RNDs compared to the true density. This indicates that the mean of the distribution is wrongly estimated by these two techniques.

The question of whether visible differences between the two densities are statistically significant needs to be addressed by using non parametric goodness of fit tests. I consider two goodness-of-fit tests: the two-sample Kolmogorov-Smirnov and the two-sample Cramr-von Mises. These tests compare the cumulative distribution functions rather than the probability distribution functions. More concretely, the two-sample Kolmogorov-Smirnov test (KS) is based on the maximum distance between two curves. It takes into account the whole of distribution's quantiles to test the assumption $H_0 : Q_{true} = Q_{RND}$, where Q is the cumulative distribution function: $Q = \int q(u)du$. This is based on the following test statistic:

$$D = \sup_F \left[\left| Q_{true}(F) - Q_{RND}(F) \right| \right]$$

The two-sample Cramr-von Mises (CvM) statistic measures the quadratic deviations between two curves:

$$\omega^2 = \frac{1}{4} \sum_{i=1}^{N} \left[Q_{true}(F_i) - Q_{RND}(F_i) \right]^2$$

Table (3) displays the reject percentage of the null at 5% and 10% levels for each of the estimation methods:

It is clear from the results in Table (3) that the two tests do not provide exactly the same results and the CvM test seems to be more 'conciliate' by rejecting the null assumption in only two cases (the implied tree method and the interpolation in a (volatility / strike) space method). This obvious lack of precision leads us to prefer the KS test which, by taking the maximum distance between the two densities, gives more information about extreme deviations. This test indicates that Edgeworth expansion method provides incontestably the best statistical performance with an acceptance rate of 100%: for each of the 500 observations, the density estimated using this method agrees with the true density with a difference statistically non significant. Moreover, the choice of an interpolation

space proves to be very important insofar as switching between the two spaces improves appreciably the results (for example, the rejection rate decreases from 12.4% to 1.8% at 5% level by moving from the (volatility / strike) space to the (volatility / delta) space). The poor performance of the implied binomial tree method is confirmed and the null assumption is rejected into almost half of the cases regardless of the confidence band.

3.3 Testing the forecasting performance of RNDs

Forecast evaluation consists in using the true density as a point of reference to rank the RND estimation techniques relative to their ability to well approximate the true density. In this way, [10] use the Probability Integral Transform (PIT) approach, initially proposed by [8], as an appropriate mean of evaluation of density forecast. The PIT score is defined as:

$$z_{i,\tau} = \int_{-\infty}^{S_{t+\tau}} q_i(u)du$$
$$= Q_i(S_{t+\tau})$$

If forecasts and true densities coincide, then the sequence of PITs z_i of the realized outcomes $S_{t+\tau}$ observed at maturity is independently and identically distributed (i.i.d) with uniform distribution:

$$\{z_i\}_{i=1}^n \overset{iid}{\sim} U(0,1)$$

The main power of the PIT approach is that it holds regardless of the particular method followed to produce the density forecasts, and therefore, can be used for all the RND estimation techniques. For practical applications, testing whether the PIT series is i.i.d $U(0,1)$ can be achieved by using simple visual tools such as histograms or by carrying out goodness-of-fit tests.

Before starting the estimation, we need to put the problem in context by presenting the manner by which the sequence of the PITs is obtained. At every 21st observation along the sequence of realization $\{S_t\}_{t=500}^{11000}$ of **Case A**, one PIT is calculated for each of the six estimated RNDs. The forecast procedure starts at the first observation of S_t (i.e. at the 500^e observation) and the PIT is obtained by calculating the cumulative probability corresponding to the realization observed 21 days later:

$$z_{500} = Q_{500}(S_{521})$$

Then, I move 21 days forward and calculate the second PIT staring at $t = 521$ to obtain the PIT for the realization 21 days later. This procedure is repeated a total

of 500 times in order to generate 500 z-values for each of the six RND estimation methods. As it was previously mentioned, the sequence $\{z_i\}_{i=1}^{500}$ should be i.i.d $U(0,1)$ if the RND provide good forecast of the true density.

Because of their ease of interpretation, histograms of the z-series are used to verify the requirement that z is uniform over the interval $[0,1]^5$. Figure (2) provides 15 bins-histograms relative to each of the RND estimation methods. If the density forecast is correctly calibrated then, each of the histograms should be roughly flat and a random of 10% of the 15 bars should fall outside the two horizontal lines delimiting the 90% confidence interval.

It appears rather clearly that any of the estimation techniques provides a good forecast of the true density since systematic deviations from uniformity are observed in the histograms. In the best of cases, only 8 of the 15 bars are inside the confidence interval (case of the mixture of lognormals, the Edgeworth expansion and the interpolation in space (volatility / delta) method). The most significant violations are in particular observable around the tails of the distribution: indeed, it appears that there are many observations in the two extremes which suggest that RNDs have tails which are too thin. Except for the implied tree method, the RND tails tend to be above their true density's counterparts which means that the RND assigns more probability to extreme outcomes than it is observable on the effective realizations. These U-shaped histograms can also be explained by the fact that the forecast provides a too narrow variance of the true density, resulting in a shaped distribution with thin tails.

The comparison based on the uniform distribution could be inappropriate for small sample size which is the case here. Moreover, the fact that the comparison interval is limited between 0 and 1 may make the results hard to interpret. For these reasons, [9] suggests applying a transformation to normality to the PIT series:

$$ x_{i,\tau} = N^{-1}(z_{i,\tau}) = N^{-1}\big[Q_i(S_{t+\tau})\big] $$

If the sequence of z_i is iid $U(0,1)$, then the sequence of x_i must be iid according to a standard normal distribution. The normal transformation allows the application of several comparison techniques and tests. Among these techniques, the well-known QQ-plot of the normal transform variables, which displays a quantile-quantile plot of the sample quantiles of x_i versus theoretical quantiles from normal distribution. If the distribution of x_i is normal, the plot will be close to line. Figure (3) displays these plots:

We can see that the RND predicts very well the body of the true density but much less its tails. As stated

previously, the RND underestimates the right tail of the true density and the empirical right-hand quantiles of x_i are systematically above those of the standard normal distribution. However, concerning the left tail, the results differ from those obtained from the histogram analysis: the parametric methods (the simple lognormal, the mixture of lognormals and the Edgeworth expansion) overestimate the left tail of the true density (the empirical quantiles being above those of the standard normal distribution) and the non-parametric methods (Smile interpolation-based methods) underestimate the left tail of the true density (empirical quantiles being below those of the standard normal distribution). In terms of absolute performance, it appears that the Edgeworth RND provides the better forecast since it deviates rather slightly from the true density on the tails.

All RNDs, except the implied binomial tree RND, present the same profile for the body of the distribution and differ on the tails. Therefore, it becomes legitimate to focus on the extreme probabilities and the discrepancies between the true density and the RND around the tails can be studied more formally using the so-called scoring rule. The idea consists in defining a generic interval associated with the extreme quantiles $\{p_l, p_u\}$ and a binary variable B_t which takes the value of 1 if the true realization of the underlying falls into this interval and 0 otherwise:

$$ B_{t,\tau} = \begin{cases} 1 & \text{if} \quad S_{t+\tau} \in \{p_l; p_u\} \\ 0 & \text{otherwise} \end{cases} $$

The interval forecast is correctly calibrated if the expected outcome of the underlying asset inside the pre-specified interval is equal to the probability defined by the extreme quantiles: $E(B_{t,h}) = p_u - p_l$. Table (4) reports the results according to the estimation technique:

More than a confirmation of the bad forecast performance in the tails of the RNDs, already observed on the histograms of the figure (1), these results permit to quantify the magnitude of these violations. It appears that, even if all techniques fail to assess correctly the tails, some do better job than others. In fact, the Edgeworth expansion seems to be the best with the lowest score either in the left tail or in the right tail; comes then the mixture of lognormals, the two non-parametric interpolation methods, the simple lognormal and finally the implied binomial trees which should be rejected without any regret.

4   Conclusion

This paper sets out to examine the empirical performance of six RND estimation techniques using three different comparison tools (the pricing errors, the true density approximation and the forecast performance). The main result that arises is that the choice of the suitable RND estimation technique depends on the purpose of the study:

---

[5]The assessment of the i.i.d condition can be achieved by plotting correlograms of $(z - E(z))$ and powers of it.

if the RND is used within an option pricing framework, non parametric techniques would be the best; they indeed provide the lowest evaluation errors. This is rather surprising because for parametric approaches, the distribution is obtained directly by minimising the pricing errors whereas in the non parametric methods, the implied volatilities, and not the option prices, are approximated. However, if the purpose is to study market expectations, the Edgeworth expansion technique should be used with caution since it approximates very well the body of the true density and performs relatively not badly around the tails.

This result may be interesting for academics and practitioners concerned by investigating market expectations implicit in option prices. Future researches may use real data to see whether the most suitable technique depends on market characteristics not.

References

[1] B. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall Publisher, 1986.

[2] S. A. Syrdal, *A study of Implied Risk-Neutral Density Functions in the Norwegian Option Market*,Journal of Finance, 2002

[3] S. Markose, A. Alentorn, "Bank of England," *The Generalized Extreme Value (GEV) Distribution Implied Tail Index and Option Pricing*, 2005

[4] R. R. Bliss, N. Panigirtzoglou, "Bank of England," *Testing the Stability of Implied Probability Density Functions*, 2000

[5] M. Andersson and M. Lomakka, "Riksbank Research Paper Series No. 1," *Evaluating Implied RNDs by Some New Confidence Interval Estimation Techniques*, 2001

[6] N. Cooper, "Bank of England," *Testing Techniques for Estimating Implied RNDs from the Prices of European Style Options*, 1999

[7] J. C. Jackwerth, "Review of Financial Studies," *Recovering Risk Aversion from Option Prices and Realized Returns*, V13, N2, pp. 433-45

[8] M. Rosenblatt, "Annals of Mathematical Statistics," *Remarks on a multivariate transformation*, V23, pp. 470-72

[9] J. Berkowitz, "Journal of Business and Economic Statistics," *Testing density forecasts, with applications to risk management*, V19, pp. 465-74

[10] F. X. Diebold, T. A. Gunther, A. S. Tay, "International Economic Review," *Evaluating density forecasts with applications to financial risk management*, V39, pp. 863-83

	RMSE $\times 10^{-3}$	RRMSE $\times 10^{-3}$
Simple lognormal	1.94 (0.97)	6.02 (7.95)
Mixture of 2 lognormals	9.99 (49.75)	3.55 (7.73)
Edgeworth expansion	1.22 (0.25)	1.33 (3.57)
Implied binomial trees	99.78 (49.96)	7.56 (26.92)
Interpolation (volatility/strike)	0.59 (0.84)	0.62 (8.09)
Interpolation (volatility/delta)	0.67 (0.85)	0.63 (0.81)

Table 1: RMSE and RRMSE Means and standard deviations (between brackets) for each of the estimation methods

	R^2 adjusted	Wald statistic
Simple lognormal	99,98%	2327,91 *
Mixture of 2 lognormals	99,79%	24,67 *
Edgeworth expansion	99,99%	24221,10 *
Implied binomial trees	98,40%	2555,10 *
Interpolation (volatility/strike)	99,89%	570,52 *
Interpolation (volatility/delta)	99,83%	225,57 *

Table 2: Coefficients of determination, regression parameters and Wald test results. * indicates the significance at 5% and 1% levels.

	Test de KS		Test de CvM	
$H_0 : Q_{true} = Q_{RND}$	Rejet H_0		Rejet H_0	
	5%	**10%**	**5%**	**10%**
Simple lognormal	1.6%	6.6%	0%	0%
Mixture of 2 lognormals	10%	10.2%	0%	0%
Edgeworth expansion	0%	0%	0%	0%
Implied binomila trees	44.2%	45.6%	32.2%	35%
Interpolation (volatility/strike)	12.4%	14%	3.4%	3.4%
Interpolation (volatility/delta)	1.8%	3.4%	0%	0%

Table 3: Kolmogorov-Smirnov and Cramr-von Mises tests for the goodness of fit between the RND and the estimated true density. Note: 'reject H_0' is the percentage of rejection of the null assumption at 5% and 10% levels. The critical values of the KS test at 5% and 10% levels are respectively 0.0270 and 0.0243; those of the CvM test are respectively equal to 0.4614 and 0.3473.

	$\{p_l; p_u\} = \{0; 0.05\}$	$\{p_l; p_u\} = \{0.95; 1$
Simple lognormal	0.1	0.04
Mixture of 2 lognormals	0.076	0.06
Edgeworth expansion	0.068	0.052
Implied binomial trees	0.02	0.094
Interpolation (volatility/strike)	0.08	0.074
Interpolation (volatility/delta)	0.076	0.074

Table 4: Score tail tests performed separately on the left and the right tails of the forecast density for each of the RND estimation methods.
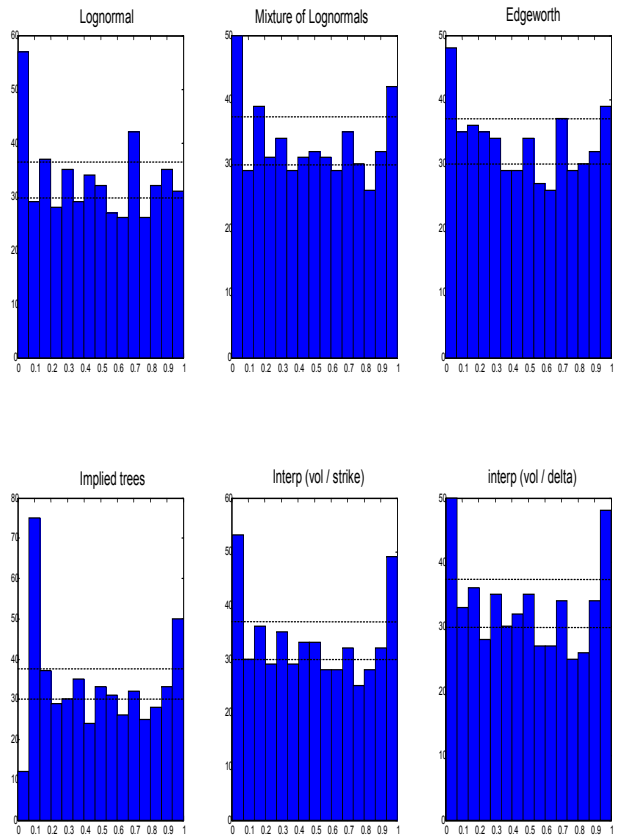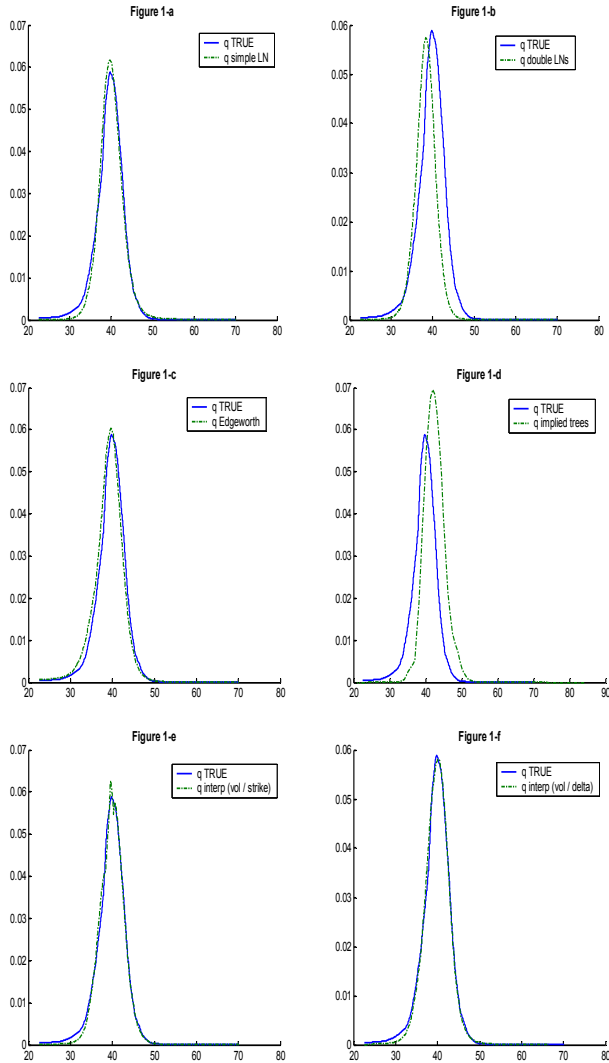
Figure 1: True probability density and RND for each estimation technique.

Figure 2: Histograms of Probability Integral Transforms with 90% confidence bands for each of the RND estimation methods.
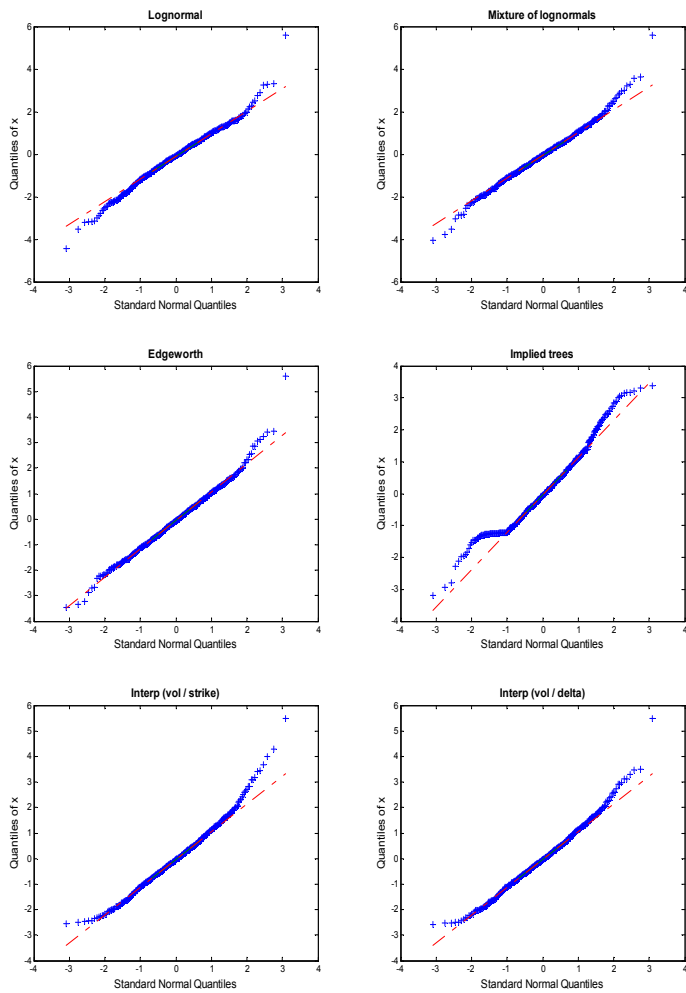
Figure 3: QQ-plot of Normal Transforms of the Probability Integral Transforms for each of the RND estimation methods.