Can we really discard forecasting ability of option implied Risk Neutral Distributions?*

Antoni Vaello-SebastiÃ M.Magdalena Vich-Llompart

Universitat de les Illes Balears

Abstract

Estimation of risk-neutral distributions is of major importance on security valuation, risk management and asset allocation. Its forecasting ability has often been proved to fail by the literature based on the results of Berkowitz test. But, can we really discard forecasting ability of option implied risk-neutral distributions? This paper compares the extraction of risk-neutral distributions from option prices using a parametric model, mixture of two lognormal distributions and two non-parametric models: kernel regression and spline methods. Non-parametric techniques are limited within the range of the observed data, therefore we deal with the estimation of the tails by extrapolating outside the available range of data as well as by appending tails drawn from a generalized pareto distribution. In order to test whether the extracted risk-neutral distributions are good forecasters of future movements of the underlying block-bootstrap simulations are run, which conclude that Berkowitz test assumptions do not hold, and so forecasting ability of such densities cannot be rejected as the test would suggest, for any of the indexes and methodologies analyzed.

Keywords: Risk Neutral Density, Options, LogNormal Mixtures, Kernel Regression, Splines

*Authors acknowledge the financial support of the Spanish Ministry of Education, Science and Innovation through the grant ECO2013-4409.

1 Introduction

Risk-neutral distributions have become of major importance in different applications such as risk management, asset allocation, pricing securities, forecasting, among others. Focusing on the forecasting ability of such distributions, we analyze how reliable or accurate the methodology used by the literature, that is Berkowitz test, is when concluding their lack of predictability.

Regarding risk-neutral distributions, we first find in the literature the well-known Black and Scholes pricing model which assumes that prices are defined on a Geometric Brownian Motion, therefore the stock prices at a given expiration date follow a Log- Normal distribution. However, Jackwerth and Rubinstein (1996) documented that risk-neutral distributions were in fact Log-Normal distributed before the crash in 1987 becoming skewed and leptokurtic afterwards. Since the crash in 1987, researchers have noticed how theoretical prices significantly differ from the observed prices.

It seems obvious that Black and Scholes model (heretofore referred to as BSM) do not hold any longer in actual markets, since its main assumptions such as flat volatility and lognormal returns have been proved to fail. Consequently, literature has turned its focus on calculating the implied volatility from equilibrium market prices and inferring the stochastic process of the underlying instead of assuming it as previously. Under complete markets, the entire probability distribution can be extracted from security prices.

Arrow (1964) and Debreu (1959) introduced the well-known Arrow-Debreu securities which pay off only if a specific state realizes and nothing otherwise, which collectively determine a state price density being each state one possible realization of the underlying asset. As suggested by Ross (1976), Arrow-Debreu securities can be approximated using option securities which contain valuable information about investors' preferences and expectations.

In the wake of this approach, researchers have been inferring the risk-neutral distribution (RND) from option market prices which has the advantage of resulting in forwardlooking estimates, therefore these RNDs should be more responsive to changes in the market as well as being model free.

Different approaches have been used in the literature to infer the RNDs from option prices. A strand of the literature has used either polynomials or splines in order to have a spectrum of strike prices. Representatives would include Bliss and Panigirtzoglou (2002) who compared the spline method versus a mixture of two Log-Normal distributions as a method to approximate the RNDs, and they found the first technique was better. In the same line, Bu and Hadri (2007) compared the splines method versus a parametric confluent hypergeometric density and concluded that the latter performed better. Later, Alonso et al. (2005) used splines and a mixture of two Log-Normal distributions and found that both methods produced very similar results. On the other hand, Aït-Sahalia and Lo (1998) proposed a semi-parametric approach consisting in the use of the kernel regression to smooth the data. Nevertheless, they assumed that distributions were time invariant.

Even though there is no consensus on which method to use, estimation of the RNDs has been practiced for a long time now and one of the main purposes is to test its forecasting ability, that is, how well they can predict future movements of the underlying. The work of Lynch and Panigirtzoglou (2008) concluded that RNDs were not useful to predict future realizations but markets did react to events such as crisis. Hamidieh (2010) found that the left tail became thinner during the peak of the crash. A group of researchers such as Anagnou et al. (2005) for the UK market, Craig et al. (2003) for the German market, Bliss and Panigirtzoglou (2004) for the UK and US market and Alonso et al. (2005) for the Spanish market, tested the predicting power of the RNDs and they all concluded that implied RNDs do not produce accurate forecasts of actual probability density functions. Alonso et al. (2005) could not reject the null hypothesis when they considered the whole sample period, but they did reject it for the sub-periods considered. In general, the different literature advocates that differences between the two exist due to the presence of risk aversion of the representative agent, and so actual or real-world distributions would be more appropriate because they do incorporate investor's beliefs and preferences.

In this study we aim to estimate daily risk-neutral distributions on 3 major indexes which are Nasdaq100, Russell2000 and S&P500, and for time horizons of 15, 30, 45 and 60 days. As mentioned earlier, there is no consensus on the method to use in order to extract the correct risk-neutral distributions, different literature has reached different conclusions, however such literature has been using different data sets. We choose in this study a parametric technique (mixture of two Log-Normal distributions) and two non-parametric techniques (splines and kernel regressions), which will be applied to the same data set and for the same period of time; therefore we can directly compare them and build conclusions for each of the approaches. In order to check the forecasting ability of the distributions, we use the Berkowitz test and also the probability mass in the tail through the Brier's Score.

Our sample of data is of special interest because it is recent as well as one of the longest ever tested in the literature; further more, it embraces two major crisis.

Berkowitz results show that in fact risk-neutral densities are rejected as being good forecasters; however, this test is built on the assumption that observations are independent, when in fact they are not. In order to capture the autocorrelation of the data we run block-bootstrap simulations. Results of the latter show that we cannot reject the null hypothesis.

The paper is organized as follows: in Section 2 we present the different methodologies used to extract the RNDs, Section 3 contains the tests applied, Section 4 presents the data, in Section 5 we see the results and discussion, and finally in Section 6 we conclude.

2 Methodology

There exists a vast literature concerning the extraction of the RNDs. Much of these methods have to do with two techniques: parametric methods, to which major contributors would include Banz and Miller (1978) and Rubinstein (1994) among others; and non-parametric methods, being Aït-Sahalia and Lo (1998) and Bliss and Panigirtzoglou (2002) relevant references. In order to provide some robustness to our results we consider 3 different alternatives to extract the RNDs from option prices, being one parametric and two non-parametric.

Due to its simplicity, the most common method is the parametric approach, which is based on choosing a certain option pricing model built on a flexible parametric return distribution which allows for thick tails and skewed shapes. Then, RND parameters are those that best fit the observed prices. This approach is fairly easy to implement and yields to well-behaved distributions (non-negative RNDs). For the parametric methods we use the well-known Log-Normal mixture distributions (heretofore LNM)[1].

The non-parametric methods lie on the Breeden and Litzenberger (1978) result to obtain the RNDs. These methods do not assume any specific form of the probability distribution function and they are based on weaker assumptions. However, interpolation of the data is needed in order to have a continuous range of option prices across moneyness. Following the works of Aït-Sahalia and Lo (1998) and Bliss and Panigirtzoglou (2002), we use the kernel regression and the spline approaches for interpolating and smoothing the data before applying the Breeden-Litzenberger technique to finally obtain the RNDs.

2.1 Parametric RNDs

Mixture of Log-Normal distributions has been widely used in literature in different fields such as the analysis of the interest rates, see Bahra (1997) and Söderlind and Sevensson (1997), among others; Campa et al. (1998) and Jondeau and Rockinger (2000) who used it on exchange rates; as well as Bliss and Panigirtzoglou (2002), Anagnou et al. (2002) and Liu et al. (2007), who applied this technique to equity indexes. This approach consists on a weighted average of Log-Normal distributions. The main advantage of the mixture of Log-Normal distributions is that non-negativity of the distribution is ensured, as well as being easy to implement and flexible enough to fit a broad range of different shapes, allowing for bimodality.

[1] To implement this approach we follow the lines of Taylor (2005). We refer the reader to the book for further details.

2.2 Non-parametric RNDs

As per Breeden and Litzenberger (1978) the whole risk-neutral density can be extracted by taking the second partial derivative of the option pricing function with respect to the strike price. Hence, the risk-neutral density of the underlying asset at expiration $f(S_T)$, is given by

$$f\left(S_T\right) = \; e^{r(\tau)}\frac{\partial^2 C\left(S_t, X, T, t\right)}{\partial X^2}\;|_{X=S_T} \tag{1}$$

being r the risk-free rate, $C\left(S_t, X, T, t\right)$ the European call price function, S_t the current value of the underlying asset, X the strike price of the option, T the expiration date, t the current date and $\tau = T - t$ the time to expiration. The corresponding cumulative risk-neutral distribution function can be obtained as follows,

$$F(X) = e^{r_f \tau}\frac{\partial C}{\partial X} + 1 \tag{2}$$

However, non-parametric methods are challenged with two hurdles due to the nature and availability of the data. To compute numerically equation (1) by finite-differences, a thin grid of strike prices encompassing all possible future payoffs is needed. Nevertheless, available data is sparse in the strike domain, hence option prices must be interpolated. Further more, option prices may be noisy, so a smoothing technique needs to be applied. To overcome these drawbacks, we use both the kernel regression and the spline technique.

As proposed by Malz (1997) instead of interpolating on prices directly (volatility-price space), it can be done on an implied volatility-delta space. The advantage of this method is twofold: first, it groups away-from-the-money options more closely permitting the data to have a more accurate shape at the center of the distribution where information is more reliable; and second, call option delta is bounded between [0; 1], in contrast to the strike price domain which is theoretically unbounded. In order to convert option prices into implied volatilities (*iv*) and exercise prices into deltas , BSM formula is used. Once *iv*s are fitted into the corresponding smoothing technique in order to get the continuum of data, they are converted back into option prices using the same formula. [2]

2.2.1 Kernel Regression

We propose the kernel regression estimator of Nadaraya (1964) and Watson (1964) as our first non-parametric method to smooth and interpolate the data, namely

$$\hat{m}_h(x) = \frac{n^{-1}\sum_{i=1}^{n} K_h(x-x_i)Y_i}{n^{-1}\sum_{i=1}^{n} K_h(x-x_i)} \qquad ; \qquad K_{h_n}(u) = h_n^{-1} K\left(\frac{u}{h_n}\right)$$

being x and Y_i the Δ and iv of the observed options, respectively; K_{h_n} a kernel function and h the bandwidth (smoothing) parameter.

We choose as K_{h_n} the gaussian kernel, however as mentioned in Aït-Sahalia and Lo (1998) the choice of the kernel function has not as much influence on the result as the choice of the bandwidth h, being the outcome very sensitive to this value. A wide range of alternative approaches to calculate h have been studied in Silverman (1986) and Härdle (1990). However, there is no consensus in the literature about the optimal h nor the best method to use to calculate it. In this work, we choose among different values calculated using both *leave-one-out cross-validation* and *Silverman's Rule-of-Thumb*.

At this point we are faced with the limitation of being able to estimate only the part of the RND corresponding to the observed range of strikes. Extreme strike observations are scarce or even non-existent, being most of them illiquid and therefore the information embedded in such prices may be misleading and unreliable. Not because extreme events, which form the tails of the distribution, are rare means that they cannot occur; but the contrary, the information contained in the tails is of major importance in risk management to carry out value-at-risk analysis, as well as in asset allocation, among others.

We find a scarce literature exploring the issue of the tails which still remains a

[2]Note that, at this point the use of the Black-Scholes pricing formula does not presume that such formula correctly prices the options, it is merely a tool to change from prices to iv and from X to Δ, being reverted back in future steps into exercise price domain. In order to change from exercise price domain to Black-and-Scholes-delta domain we use the same volatility for all observations, which is obtained from a weighted average of the different implied volatilities.

challenge for researchers. To make estimations beyond the range of observed values, we need to extrapolate somehow the available data. One approach is to assume a parametric probability distribution to approximate the tail zone. Birru and Figlewski (2012) state that as per Fisher-Tippett Theorem, a large value drawn from an unknown distribution will converge in distribution to one of the Generalized Extreme Value distributions (GEV) family, so they propose the use of the Generalized Pareto distribution (GPD), which also belongs to the extreme value distributions family. The attractiveness of this method is that it has only two free parameters, which are η, the scale parameter and ξ, the shape parameter. We follow this approach to complete the tails of the RND, and we append GPD tails to our kernel-based RND. More details about this procedures is given in the appendix. An illustrative example is depicted in figure 1.

2.2.2 Splines

Following Bliss and Panigirtzoglou (2004), we also consider to fit *iv* using cubic smoothing splines (piece-wise polynomials). The smoothing spline is defined by the knots and polynomial coefficients that minimize the following function,

$$S_\lambda = \sum_{i=1}^{n} m_i (Y_i - g(\Delta_i, \theta))^2 + \lambda \int_{-\infty}^{+\infty} f''(x;\theta)^2 dx \tag{3}$$

where m_i is a weighting value of the squared error, Y_i is the implied volatility of the ith option observation, $g(\Delta_i, \theta)$ is the fitted *iv* which is a function of Δ_i and a set of spline parameters, θ; $g(\Delta_i, \theta)$ is any curve which can have any form and whose coefficients are estimated by least-squares. λ is the smoothing parameter, which following Bliss and Panigirtzoglou (2004) takes value 0.99, and $f''(x;\theta)^2$ is the smoothing spline.

For the m_i weight in equation 3, Bliss and Panigirtzoglou (2004) use the BSM *vegas* of the observed options. However, we slightly modify this weighting scheme using square *vegas*, which places more weight to those near-to-at-the-money observations and therefore it performs better.

Like in the kernel-based method, we are faced with the limitation of being able to estimate only the part of the RND corresponding to the observed range of strikes, missing some probability at the extremes. For the spline methodology we deal with the tails using two different approached. First, we simply extrapolate the spline outside the observed Δ domain; however this can cause implausible or negative *iv*s, as well as kinks at the ends of the RNDs. Bliss and Panigirtzoglou (2004) propose to add two extra points at both ends of the moneyness domain and assign them the *iv* value of the corresponding end point. We follow this approach and get an extended moneyness range. The second approach is to fit Pareto tails, the same way as it is done with the Kernel method.

To illustrate the different methodologies used in this paper, figure 2 exhibits the extracted RNDs calculated for S&P500 index options with 30 days to maturity for two different days: one RND is from 21 July 2005 (left hand side plots), just before the global financial crisis, while the second RND is from 23 July 2009 (right hand side plots), just after the crisis. The top plot represent the RNDs from our parametric method (Log-Normal mixture), while the bottom plots depict the non-parametric ones (Kernel with Pareto tails, Splines with extrapolation and Splines with Pareto tails). From this figure, two facts arise: First, the different methodologies used in this paper seem capable to capture the main features of option implied RNDs; second, comparing the x-axis of the pre and post crisis RNDs, it is clear that the (moneyness) domain has spread out.

3 The tests

In order to verify whether RNDs accurately forecast realized ex-post returns, we rely on two tests. First, we analyze the performance of the Berkowitz (2001) test that jointly tests independence and uniformity and which has been used by Bliss and Panigirtzoglou (2004) and Alonso et al. (2006), among others.[3]. Second, we focus on the tails using the Brier Score, which is based on the realized frequency for a certain (extreme) quantile (i.e. 5%,

[3]According to Bliss and Panigirtzoglou (2004), this test performs better than several non-parametric tests, such as, Kolmogorov-Smirnov, Chi-squared or Kupier tests.

10%, 90%, 95%). Finally, in order to verify the reliability of these tests, we compute the bootstrap distribution of the test statistics.

For the Berkowitz test, given a set of implied RNDs for each date t_i with a specific τ-horizon, $\hat{f}_{t_i,\tau}(S_{t_i+\tau})$, where $S_{t_i+\tau}$ are the values at expiration; the Berkowitz test first transforms $S_{t_i+\tau}$ into a new variable z_{t_i} using the probability integral transform,

$$z_{t_i,\tau} = \Phi^{-1}\left(\int_{-\infty}^{S_{t_i+\tau}} \hat{f}_{t_i,\tau}(u)\, du\right) \tag{4}$$

where $\Phi^{-1}(\ldots)$ stands for the inverse of the standard Normal distribution function.

Under the null hypothesis that $\hat{f}_{t_i,\tau}(\ldots) = f_{t_i,\tau}(\ldots)$ and the assumption that $S_{t_i+\tau}$ are independent, the new variable $z_{t_i,\tau} \sim iid\ N(0,1)$. In this test, independence and normality of $z_{t_i,\tau}$ are tested by estimating by maximum likelihood the following AR(1) model[4],

$$z_{t_i,\tau} - \mu = \rho\left(z_{t_i-1,\tau} - \mu\right) + \epsilon_{t_i,\tau}, \quad \epsilon_{t_i,\tau}\ iid\, N(0,\sigma_\epsilon^2) \tag{5}$$

Under the null hypothesis, the estimated parameters should be $\left[\mu,\, \sigma_\epsilon^2,\, \rho,\right] = [0,\, 1,\, 0]$. Therefore, the likelihood ratio test

$$LR_3 = -2\left[L(0,1,0) - L\left(\hat{\mu}, \hat{\sigma}_\epsilon^2, \hat{\rho}\right)\right] \tag{6}$$

is asymptotically distributed as $\chi^2(3)$ under the null hypothesis.

The presence of overlapping or non-overlapping but serially correlated data may lead to a false rejection of the null hypothesis. For that, Berkowitz (2001) suggests testing the independence assumption separately as follows,

$$LR_1 = -2\left[L\left(\hat{\mu}, \hat{\sigma}^2, 0\right) - L\left(\hat{\mu}, \hat{\sigma}^2, \hat{\rho}\right)\right] \tag{7}$$

which under the null hypothesis is asymptotically distributed as $\chi^2(1)$. As per the previous,

[4]Even though dependency can arise from a more complex structure than an AR(1), this dependence structure is the most evident and intuitive, specially in overlapped data.

LR_3 results will be more reliable when LR_1 fails to reject. Should LR_1 reject, we cannot ascertain whether the reason is lack of predictability of the RNDs or the presence of serial correlation in the data.[5]

Following Anagnou et al. (2005) and Alonso et al. (2006), we also test the goodness of fit of the tails separately. They proposed the statistic suggested by Seillier-Moiseiwisch and Dawid (1993) to test whether Brier Score departs from its expected value. Brier Score is defined as

$$B = \frac{1}{T}\sum_{t=1}^{T} 2\left(\hat{F}_{t,\tau}^{tail} - R_{t,\tau}\right)^2$$

and measures the accuracy of the probabilistic predictions based on the distance between a selected probability mass in the tail, $\hat{F}_{t,\tau}^{tail}$, and a binary variable, $R_{t,\tau}$, which takes value 1 if the true realization of the underlying falls into the tail being tested, or 0 otherwise.

$$Y = \frac{\sum_{t=1}^{T}\left(1-2\hat{F}_{t,\tau}^{tail}\right)\left(R_{t,\tau}-\hat{F}_{t,\tau}^{tail}\right)}{\left[\sum_{t=1}^{T}\left(1-2\hat{F}_{t,\tau}^{tail}\right)^2 \hat{F}_{t,\tau}^{tail}\left(1-\hat{F}_{t,\tau}^{tail}\right)\right]^{\frac{1}{2}}} \quad (8)$$

which is asymptotically distributed as a Standard Normal.

Due to the features of the data (short samples and dependence), the empirical distribution of the statistics in equation (6) may differ from the asymptotic ones, yielding to different critical values, and thus wrong decisions about rejection of the null hypothesis may be taken. To overcome this problem, we compute bootstrap-based critical values. Since it is of interest to maintain the structure present in the data, we use block-bootstrap.[6] This method was first introduced by Künsch (1989) and it divides the sample into different blocks (which may be overlapped) of b consecutive observations. Then, bootstrap samples are built by randomly concatenating blocks to match the original sample size.[7] Once m

[5]Note that failure to reject does not necessary imply that the null hypothesis is true.

[6]When re-sampling, note that the outcome will be only as good as the ability of the data generating process (bootstrap simulations) to fairly mimic the actual data and their structure.

[7]Künsch (1989) proposed that a reasonable block length would be $n^{1/3}$, where n is the length of the original sample. We have also tried with $n^{1/3}+3$, $n^{1/3}+8$, $n^{1/3}+13$ and $n^{1/3}-1$ observations. These values generate blocks of length 10, 15, 20 and 6 observations, respectively, for the S&P500 case. In our analysis, results are very similar and lead to the same conclusions regardless the length of the block.

bootstrap samples have been generated, the statistics of interest are calculated for each sample. Then, we can empirically approximate the desired percentile (critical value).

4 The Data

We have a set of European call and put options written on three of the major and widely traded indexes, S&P500, Nasdaq100 and Russell2000, from the OptionMetrics database. We have observations ranging from January 1996 until October 2015. We use daily closing prices for all the indexes and calculate the mid-point of the bid and ask price of the options.

Because extreme observations are considered to be very-far-away-from-the-money and therefore illiquid and fairly unreliable, following Panigirtzoglou and Skiadopoulos (2004) we discard observations with delta, Δ, values beyond the range $[0.01;\ 0.99]$.

We calculate a risk-neutral distribution for those days of the sample with options maturing in 15, 30, 45 and 60 days. Nonetheless, the fit at time t may be discarded due to the reasons exposed later in this section; being such the case, we try to fit the RND from the previous day, $t-1$ (in this case the options will mature in 16, 31, 46 or 61 days), should this second fit be also discarded, we try to fit data from the following day, $t+1$ (in this case the options will mature in 14, 29, 44, or 59 days). Therefore, for each maturity, should the fit for the corresponding day be discarded, we allow to include the fit on the previous or the following day. We proceed in this way in order to increase the sample size to run the tests explained in section 3.

Data can present some anomalies, and therefore a filtering is required before the implementation of the different models. Under the assumption of complete markets, those options which do not satisfy the arbitrage conditions are discarded from the sample. Those options which are very-far-away-from-the-money are also dropped from the sample since they are poorly traded and thus illiquid, so the information embedded in their prices can be unreliable and of no use. Therefore, following the literature, we keep only in the sample those observations whose moneyness lies within 0.75 and 1.25. We also require a minimum of 8 observations to perform any estimation.

When working with options we need to deal with the presence of the dividends. Such variables are unobservable and difficult to estimate. We will follow in this study the approach proposed by Aït-Sahalia and Lo (2000), in which they work with forward quotes of the underlying instead, therefore dividends go out from the formulas. Since the assumption of complete markets holds, we can infer the forward prices, F, for the underlying from the put-call parity formula,

$$F = (C - P)\, e^{r\tau} + X \tag{9}$$

where C and P are the prices for the call and put options respectively, r is the risk-free rate, τ is the time left to maturity of the option and X is the strike price

Given a certain day and maturity, there exist a call and a put option for each exercise price. Following Aït-Sahalia and Lo (2000), we remove those call and put contracts that are in-the-money (ITM), which are less liquid. Out-of-the-money put options are translated to their counterpart ITM call options by using the put-call parity, being these put options removed from the sample afterwards. By doing this, all the options kept in the sample are OTM. We consider call options to be ITM when their moneyness ratio F/X is higher than 1.03, while puts are ITM when their F/X is below 0.97.

For the non-parametric cases, once RNDs have been estimated and before appending tails, we discard those RNDs which account for less than the 70% of the probability mass. In case no RND is successful in matching the above criteria, we try to fit the RND on the previous or the following day instead as described above. Should the method fail to obtain a successful distribution, then that specific day is discarded from the sample.

The risk-free rate used in our analysis is the zero-coupon yield provided by OptionMetrics.[8].

[8]Bliss and Panigirtzoglou studied the effect of the risk-free proxy and concluded that a change of 100 basis points in the risk-free rate leads to a two basis points change in the measured implied volatility for a one-month horizon, and this change will be up to 5 basis points for the six-months horizon. Therefore,

5 Results and discussion

Risk-neutral distributions have been estimated using different parametric and non-parametric methods from options on 3 different indexes, S&P500, Nasdaq100 and Russell2000. Due to space limitations, we present the results for the S&P500 index only. Nevertheless, the findings are similar yielding to the same conclusions. [9]

In figure 3 we plot the volatility, skewness and kurtosis implied for the estimated RNDs for the S&P500 index options with 30 days to maturity. The figure compares the different moments across methodologies. In general, we can appreciate that moments for all methods yield to very similar results and that the estimated risk-neutral skewness is negative and the kurtosis is higher than 3 for all methods, confirming that the RNDs are not normal.

Once RNDs have been estimated under each of the different methods for the whole sample period, both Berkowitz and Brier tests are performed on all of them. Note that during this period financial markets have been hit by two major financial crisis, one in 2000, and one in 2008. The fact that such periods present anomalies and extreme movements in stock prices might have some impact and thus mislead the results of the tests. In order to check that, both tests have been run on a restricted set of data which excludes the above turmoil periods.

Both LR_3 and LR_1 Berkowitz statistics reject the null hypothesis. This is the case for all the different RNDs, regardless the index, the maturity and the methodology used. Because both statistics reject the null hypothesis, we cannot ascertain the reason of rejection. Table 1 shows the p-values of the Berkowitz test for the different indexes, maturities and methodologies. Similarly, Table 2 shows the p-values for the Berkowitz test performed on the restricted data set which does not contain those crisis periods. Both analysis yield to the same conclusions about rejection of the null hypothesis therefore crisis are not

the proxy used will have little impact on the results

[9] Tables similar to those discussed in this section for options on Nasdaq100 and Russell2000 indexes are available from the authors upon request.

responsible for the rejection of the null hypothesis.

Brier Score test has been performed to test how accurate is the tail fitting based on a given probability mass level which determines the beginning of the tail being tested. In this analysis we test 5% and 10% probability mass levels in both left and right tails separately. The test concludes good fitting of the right tail in general across methodologies, indexes and maturities. However, the null hypothesis is mainly rejected for the left tail. Results of Brier test are presented in table 3. In this table we can see a column which contains the frequency in which the true realization of the underlying falls into the tail being tested. We see that in general observed frequencies are lower than those predicted, which is due to the volatility skew.

In general right tail is performing better than the left tail,. For the right tail, we can also see that in general the 10% significance level performs slightly better than the 5%. One reason for this would be because 5% is more extreme and so statistically there are fewer observations falling into this area. Therefore the indicator is more volatilte. One more observation falling into this tail, can have a bigger impact.

Brier test is also performed on the restricted data set, that is that data without crisis periods. Results are presented in table 4 where we can see that results barely change from those calculated on the whole data set. Therefore, as before, we can conclude that the crisis periods are not the cause of such results.

From Berkowitz results, we have seen that LR_1 statistic reject the null hypothesis of independence. Further more, due to the nature of the data, one may suspect that the observations indeed present some kind of auto-correlation structure. Should this be the case, then Berkowitz assumptions would not be accurate.

In order to check whether LR_3 statistic is indeed distributed following a χ^2_3 as assumed by Berkowitz, we apply the block-bootstrap technique. Block-bootstrap is based on $5,000$ simulations of the actual data in blocks, over which we calculate the LR_3 statistic as in the Berkowitz test, obtaining a series of $5,000$ LR_3 values which provide a distribution of the statistic itself.

Recall that Berkowitz test, under the assumption that it is distributed following a χ^2_3, rejects RNDs as good forecasters of future realizations; being this the case for all indexes, methodologies and maturities used in this study. On the other hand, block-bootstrap results suggest that Berkowitz LR_3 is biased and that the distribution of the statistic is not a χ^2 with 3 degrees of freedom as previously assumed. We compute the 95th and 90th percentile of the empirical distribution of the statistic, which values will be our threshold of reference when assessing the rejection of the null hypothesis. We observe that for any methodology and time horizon studied, this value is higher than the Berkowitz LR_3 statistic. Therefore, block-bootstrap fails to reject the null hypothesis which states that RNDs are good forecasters. A comparison of such values with the Berkowitz LR_3 statistic is provided in table 5. The reported block-bootstrap of table 5 is based on blocks of length $n^{1/3}$ as per Künsch (1989), which in our data set is 6 to 7 observations per block. [10] Block-bootstrap has been also applied to the restricted data set which does not consider the crisis periods, and the conclusions reached are exactly the same, proving once again that crisis periods are not responsible for the obtained results.

6 Conclusions

Risk-neutral distributions are of great importance for portfolio and risk managers. Many studies have focused on their ability of forecasting future underlying realizations. There is a vast literature about risk-neutral distributions implied from option prices and they mainly use parametric and non-parametric techniques. Although, most of the authors have compared different methodologies, there is no consensus on which method is the best to use.

In this paper we extract the RNDs using a parametric technique (Log-Normal mix-

[10] As we have mentioned earlier, this is a rather arbitrary choice. In order to provide more robustness to the analysis, we have also tried different lengths, increasing it to 10, 15 and 20 observations per block; as well as decreasing it to 6. Analysis yields to similar results.

ture distribution) and two non-parametric (Kernel regression and Spline). With the non-parametric methodologies we are faced with the issue of the tails. We do not have data, or it is scarce, concerning the tails area and so we need to estimate them. Following Bliss and Panigirtzoglou (2004), for the Spline methodology we extrapolate the spline outside the available range of data in order to obtain the tails. Another approach proposed by Birru and Figlewski (2012) is to append tails drawn from a Pareto Distribution. We also adopt this approach in this analysis and append pareto tails to the kernel and spline RNDs.

Different to previous literature, we apply four different methodologies on the same data set, which conatins data for a longer and more recent period of time (observations range from 1996 until 2015). Based on the moments we see that they are similar across methodologies presenting a negative skewness and kurtosis higher than 3, which demonstrate that RNDs depart from normality.

In order to check the ability of RNDs to accurately forecast the future movements of the underlying, previous literature has based their analysis on Berkowitz test and concluded that RNDs do lack of such ability. In this paper we have run Berkowitz on the different RNDs extracted using the different methodologies and for the different indexes and maturities and we have also concluded rejection of the null hypothesis, which states good forecasting ability. Nevertheless, Berkowitz test is based on the assumption that the statistic is distributed as χ^2 with 3 degrees of freedom and data is independent. But, due to the nature of the data, observations can present autocorrelation, therefore: can we really reject their lack of forecasting ability? Is Berkowitz test accurate when rejects the null hypothesis? In this paper we propose to run block-bootstrap simulations in order to check the empirical distribution of the Berkowitz LR_3 statistic. Block-bootstrap captures the auto-correlation that might be present in the data since it simulates blocks of consecutive observations instead of individual simulations, therefore the structure in the data is maintained.

Block-bootstrap suggests that indeed Berkowitz assumption does not hold for our data set since the statistic is not distributed following a χ^2 with 3 degrees of freedom. Furthermore, block-bootstrap results fail to reject the null hypothesis, therefore the forecasting ability of RNDs cannot be rejected, as suggested by the Berkowitz test. Block-bootstrap has been applied to RNDs extracted using four different methodologies, being one of them parametric and the rest non-parametric. The analysis has been run on four different major indexes, S&P500, Nasdaq100 and Russell2000, and on RNDs with four different time horizons, 15, 30, 45 and 60 days. Block-bootstrap has provided consistent results across all methodologies, indexes and maturities; and even for different lengths of the blocks. Further more, the analysis have been run on a data set which does not contain the crisis periods, for which same results and conclusions have been reached, suggesting that crisis are not responsible for the results obtained. Therefore, we can conclude that rejection of the forecasting ability of the RNDs is not as obvious as Berkowitz test suggests.

APPENDIX

Adding Generalized Pareto tails

GPD is the density of the observations beyond a specific threshold c which determines the amount of probability contained in the missing tail; this is, in the case in which the left 5th percentile of the distribution is to be fitted by the GPD, then c takes the value 0.05. The GPD is given by

$$F\left(S_T | S_T \geq c\right) = \begin{cases} 1 - \left(1 + \xi\left(\dfrac{S_T - c}{\sigma}\right)\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\left(\dfrac{S_T - c}{\sigma}\right)\right) & \text{if } \xi = 0 \end{cases} \tag{10}$$

The GPD is a density itself, therefore the area under the curve is 1. Because we want to estimate the cth percentile, we need to multiply the whole density function by the value of $c\%$.

Following Birru and Figlewski (2012), we use the GPD to approximate the tails of our kernel estimated distributions. As mentioned previously, with the kernel technique we are only capable to estimate the central part of the distribution where all the available observations lay, being the probability at the extremes sometimes impossible to be explained by the non-parametric methods and therefore we are left with an amount of probability α which is missing from the analysis. With the GPD we try to fit this missing amount of probability in order to complete our RNDs. We denote α_{0R} and α_{0L} the amount of probability missing in the right and left tails respectively; and $X_{\alpha_{0R}}$ and $X_{\alpha_{0L}}$ those strike prices which leave α_{0R} and α_{0L} probability at their right and left respectively. Being these points where the pareto tails are to begin.

Following Figlewski (2008), we define an inner second point for each of the tails called $X_{\alpha_{1R}}$ and $X_{\alpha_{1L}}$, which are the strike prices that leave a probability α_{1R} at the right and

α_{1L} at the left, where

$$\begin{aligned} \alpha_{1R} &= \alpha_{0R} - p \\ \alpha_{1L} &= \alpha_{0L} + p \end{aligned} \tag{11}$$

being p some amount of probability. In our case p is set to be 2% probability, thus the amount of probability to be fitted by the pareto tail will be the missing amount of probability in each of the tails plus a 2% extra probability. However, to perform the analysis we require at least a missing probability amount in each of the tails of 2%. In case one or both α_{0R} and α_{0L} are smaller than 2%, we will manually set such α values to be 2%, and therefore their corresponding α_{1R} and α_{1L} will be set to be 3%, which means that the probability amount to be fitted by the tail in such particular cases will be 5%.

We denote $f^{RND}(X_{\alpha_{0R}})$ $(f^{RND}(X_{\alpha_{0L}}))$ and $f^{RND}(X_{\alpha_{1R}})$ $(f^{RND}(X_{\alpha_{1L}}))$ as those values of the estimated RND at $X_{\alpha_{0R}}$ $(X_{\alpha_{0L}})$ and $X_{\alpha_{1R}}$ $(X_{\alpha_{1L}})$, respectively. Pareto tails will be appended with some matching restrictions similar to Figlewski (2008). First we require that the amount of probability contained in each of the GPD tails is the same as the amount contained in the estimated RND tails. And second, we force the new GPD distribution to pass through the exact $f^{RND}(X_{\alpha_{0R}})$ $(f^{RND}(X_{\alpha_{0L}}))$ and $f^{RND}(X_{\alpha_{1R}})$ $(f^{RND}(X_{\alpha_{1L}}))$ points, thus matching the shape of the estimated RNDs. That is, we have both distributions matching values at the following points,

$$\begin{aligned} f^{RND}(X_{\alpha_{0R}}) &= f^{GPD}(X_{\alpha_{0R}}) \\ f^{RND}(X_{\alpha_{0L}}) &= f^{GPD}(X_{\alpha_{0L}}) \\ f^{RND}(X_{\alpha_{1R}}) &= f^{GPD}(X_{\alpha_{1R}}) \\ f^{RND}(X_{\alpha_{1L}}) &= f^{GPD}(X_{\alpha_{1L}}) \end{aligned} \tag{12}$$

However, between $X_{\alpha_{0R}}$ and $X_{\alpha_{1R}}$, as well as between $X_{\alpha_{0L}}$ and $X_{\alpha_{1L}}$, both the estimated RND and the fitted GPD are overlapping, having different values for each strike price contained within this overlapping zone. In order to approximate the distribution of this overlapping zone and trying to avoid abrupt jumps next to the matching points so to reach a smooth transition between both distributions, we define a weighting function which will give different weights to the strike prices based on their distance to $X_{\alpha_{0R}}$, $X_{\alpha_{0L}}$, $X_{\alpha_{1R}}$ and $X_{\alpha_{1L}}$,

$$w = \frac{f^{RND}\left(X_{\alpha_{0R}}\right) - f^{RND}\left(X_i\right)}{f^{RND}\left(X_{\alpha_{0R}}\right) - f^{RND}\left(X_{\alpha_{1R}}\right)} \tag{13}$$

for those i observations which lay within $X_{\alpha_{1R}}$ and $X_{\alpha_{0R}}$. The corresponding distribution values for each i data point is then calculated by,

$$f_{X_i}^{new} = w_i f_{X_i}^{RND} + \left(1 - w_i\right) f_{X_i}^{GPD}$$

The above calculations are for the overlapping zone at the right tail only. The equivalent equations for the left overlapping zone are,

$$w = \frac{f^{RND}\left(X_{\alpha_{1L}}\right) - f^{RND}\left(X_i\right)}{f^{RND}\left(X_{\alpha_{1L}}\right) - f^{RND}\left(X_{\alpha_{0L}}\right)} \tag{14}$$

and

$$f_{X_i}^{new} = \left(1 - w_i\right) f_{X_i}^{RND} + w_i f_{X_i}^{GPD}$$

Once we have the extracted RNDs calculated by the kernel method, in order to append tails we require to have a missing amount of probability in the tail to be fitted, either the left, the right or both tails, of at least 0.25%; should we have a lesser amount of missing probability, no estimation of the tails is required since almost all the density is explained by the observed data. Therefore, for the right tail, the amount of α_{0R} is the probability value to the right corresponding to the most extreme observation in the right end as long as such value is higher than 1%. In case this amount is lower than 1% we will assign to α_{0R} the value of 1%. In either case, the α_{1R} value will be $\alpha_{0R}+1\%$ of probability.

For the left tail, the procedure is the same. We first assign to the value of α_{0L} the probability amount corresponding to the most extreme observation to the left, which is required to leave a probability amount to the left higher than 1%. Should this observation have a probability amount lower than this 1%, we will assign to our α_{0L} the value of 1%. Consequently, the value of α_{1L} will be equal to the value of α_{0L}+1%.

Figure 1 shows the RND calculated on the S&P500 for a time horizon of 30 days. In this figure we can appreciate the main body of the distribution in blue, which has been calculated using kernel technique; red region which represents the pareto tails which have been appended in each case; and finally the figure depicts in green what we call the overlapping zone, that is the region between α_0 and α_1 which has been approximated using a weighting scheme as per 13 and 14.

References

Alonso, F., Blanco, R., Rubio, G., 2005. Testing the forecasting performance of ibex 35 option-implied risk-neutral densities., working paper No. 0505, Banco de España.

Alonso, F., Blanco, R., Rubio, G., 2006. Option-implied preferences adjustments, density forecasts, and the equity risk premium.

Anagnou, I., Bedendo, M., Hodges, S., Tompkins, R., 2002. The relationship between implied and realised probability density functions, working paper, University of Warwick and the University of Technology, Vienna.

Anagnou, I., Bedendo, M., Hodges, S., Tompkins, R., 2005. Forecasting accuracy of implied and garch-based probability density functions. Review of Futures Markets 11, 41–66.

Arrow, K. J., 1964. The role of securities in the optimal allocation of risk-bearing. Review of Economic Studies 31, 91–96.

Aït-Sahalia, Y., Lo, A. W., April 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. The Journal of Finance 53, 499–547.

Aït-Sahalia, Y., Lo, A. W., 2000. Nonparametric risk management and implied risk aversion. Journal of Econometrics 94, 9–51.

Bahra, B., 1997. Implied risk-neutral probability density functions from option prices: Theory and application., working paper, Bank of England.

Banz, R., Miller, M., 1978. Prices for state-contingent claims: Some estimates and applications. Journal of Business 51 (4), 653–672.

Berkowitz, J., 2001. Testing density forecasts with applications to risk management. Journal of Business and Economic Statistics 19, 465–474.

Birru, J., Figlewski, S., 2012. Anatomy of a meltdown: The risk neutral density for the s&p500 in the fall of 2008. Journal of Financial Markets 15, 151–180.

Bliss, R. R., Panigirtzoglou, N., 2002. Testing the stability of implied probability density functions. Journal of Banking and Finance 26, 381–422.

Bliss, R. R., Panigirtzoglou, N., 2004. Option-implied risk aversion estimates. Journal of Finance 59, 407–446.

Breeden, D. T., Litzenberger, R. H., October 1978. Price of state-contingent claims implicit in option prices. Journal of Business 51 (4), 621–651.

Bu, R., Hadri, K., 2007. Estimating option implied risk-neutral densities using spline and hypergeometric functions. Econometrics Journal 10 (2), 216–244.

Campa, J., Chang, K., Reider, R., February 1998. Implied exchange rate distributions: Evidence from otc option markets. Journal of International Money and Finance 17 (1), 117–160.

Craig, B., Glatzer, E., Keller, J., Scheicher, M., 2003. The forecasting performance of the german stock option densities, discussion Paper 17, Studies of Economic Research Centre, Deutsche Bundesbank.

Debreu, G., 1959. Theory of Value. Wiley.

Figlewski, S., 2008. Estimating the implied risk neutral density for the U.S. market portfolio. Chapter in *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Oxford University Press.

Hamidieh, K., 2010. Recovering the tail shape parameter of the risk neutral density from option prices., working paper, Rice University.

Härdle, W., 1990. Applied Nonparametric Regression.

Jackwerth, J., Rubinstein, M., 1996. Recovering probability distributions from option prices. Journal of Finance 51 (5), 1611–1631.

Jondeau, E., Rockinger, M., 2000. Reading the smile: the message conveyed by methods which infer risk neutral densities. Journal of International Money and Finance 19 (6), 885–915.

Künsch, H. R., 1989. The jackknife and the bootstrap for general stationary observations. The Annals of Statistics 17 (3), 1217–1241.

Liu, X., Shackleton, M. B., Taylor, S. J., Xu, X., 2007. Closed-form transformations from risk-neutral to real-world distributions. Journal of Banking and Finance, 1501–1520.

Lynch, D., Panigirtzoglou, N., 2008. Summary statistics of option-implied probability density functions and their properties., working paper No. 345, Bank of England.

Malz, A., 1997. Estimating the probability distribution of the future exchange rate from option prices. Journal of Derivatives 5 (2), 18–36.

Nadaraya, E. A., 1964. On estimating regression. Theory of Probability and its Applications 10, 186–190.

Panigirtzoglou, N., Skiadopoulos, G., 2004. A new approach to modeling the dynamics of implied distributions: Theory and evidence from the s&p500 options. Journal of Banking and Finance 28, 1499–1520.

Ross, S., February 1976. Options and efficiency. Quarterly Journal of Economics 90 (1), 75–89.

Rubinstein, M., July 1994. Implied binomial trees. Journal of Finance 49 (3), 771–818.

Söderlind, P., Sevensson, L., October 1997. New techniques to extract market expectations from financial instruments. Journal of Monetary Economics 40 (2), 383–429.

Seillier-Moiseiwisch, F., Dawid, P., 1993. On testing the validity of sequential probability forecasts. Journal of the American Statistical Association 88, 355–359.

Silverman, B. W., 1986. Density Estimation for Statistics and Data Analysis. Vol. 26. Chapman & Hall, London.

Taylor, S. J., 2005. Asset Price Dynamics, Volatility, and Prediction. Princeton University Press.

Watson, G. S., 1964. Smooth regression analysis. Shankya Series A 26, 359–372.

Table 1: Berkowitz test P-values

τ	model	S&P500		NASDAQ 100		RUSSELL 2000	
		LR_3	LR_1	LR_3	LR_1	LR_3	LR_1
15 days	LNM	0.0000	0.0000	0.0018	0.0002	0.0000	0.0000
	Kernel	0.0000	0.0000	0.0000	0.0226	0.0000	0.0000
	Spline	0.0000	0.0001	0.0000	0.0369	0.0000	0.0000
	Sp+PT	0.0000	0.0001	0.0000	0.0399	0.0000	0.0000
30 days	LNM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kernel	0.0000	0.0000	0.0000	0.0027	0.0000	0.0001
	Spline	0.0000	0.0000	0.0000	0.0024	0.0000	0.0002
	Sp+PT	0.0000	0.0000	0.0000	0.0027	0.0000	0.0002
45 days	LNM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kernel	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003
	Spline	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003
	Sp+PT	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003
60 days	LNM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kernel	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Spline	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sp+PT	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The table shows the LR_3 and LR_1 Berkowitz test corresponding p-values. The test is run on the RND implied from S&P500, Nasdaq 100 and Russell 2000 indexes for the different maturities (15, 30, 45 and 60 days and for the different methodologies used in the paper: Log-Normal Mixture (LNM), Kernel with Pareto tails appended (Kernel), Splines with extrapolation (Spline) as well as Splines with Pareto tails appended (Sp+PT). Results are run on the whole data set.

Table 2: Berkowitz test P-values

τ	model	S&P500		NASDAQ 100		RUSSELL 2000	
		LR_3	LR_1	LR_3	LR_1	LR_3	LR_1
15 days	LNM	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000
	Kernel	0.0000	0.0000	0.0000	0.0073	0.0000	0.0000
	Spline	0.0000	0.0000	0.0000	0.0148	0.0000	0.0000
	Sp+PT	0.0000	0.0000	0.0000	0.0165	0.0000	0.0000
30 days	LNM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kernel	0.0000	0.0000	0.0000	0.0071	0.0000	0.0001
	Spline	0.0000	0.0000	0.0000	0.0066	0.0000	0.0001
	Sp+PT	0.0000	0.0000	0.0000	0.0075	0.0000	0.0002
45 days	LNM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kernel	0.0000	0.0000	0.0000	0.0002	0.0000	0.0013
	Spline	0.0000	0.0000	0.0000	0.0002	0.0000	0.0014
	Sp+PT	0.0000	0.0000	0.0000	0.0001	0.0000	0.0011
60 days	LNM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kernel	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Spline	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sp+PT	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The table shows the LR_3 and LR_1 Berkowitz test corresponding p-values. The test is run on the RND implied from S&P500, Nasdaq 100 and Russell 2000 indexes for the different maturities (15, 30, 45 and 60 days and for the different methodologies used in the paper: Log-Normal Mixture (LNM), Kernel with Pareto tails appended (Kernel), Splines with extrapolation (Spline) as well as Splines with Pareto tails appended (Sp+PT). Results are run on the restricted data set which excludes those crisis periods.

Table 3: **Brier test results for the RNDs on S&P500**

			Left tail						Right tail					
			5%			10%			5%			10%		
τ	Model	N	Freq.	Stat.	p-value	Freq.	Stat.	p-value	Freq.	Stat.	p-value	Freq.	Stat.	p-value
15 days	LNM	374	0.0080	-3.7249	0.0001	0.0722	-1.7926	0.0365	0.0294	-1.8269	0.0339	0.0936	-0.4137	0.3396
	Kernel	374	0.0080	-3.7249	0.0001	0.0455	-3.5162	0.0002	0.0267	-2.0641	0.0195	0.0963	-0.2413	0.4047
	Spline	374	0.0080	-3.7249	0.0001	0.0374	-4.0333	0.0000	0.0374	-1.1151	0.1324	0.0936	-0.4137	0.3396
	Sp+PT	374	0.0080	-3.7249	0.0001	0.0374	-4.0333	0.0000	0.0374	-1.1151	0.1324	0.0936	-0.4137	0.3396
30 days	LNM	375	0.0213	-2.5471	0.0054	0.0560	-2.8402	0.0023	0.0213	-2.5471	0.0054	0.0800	1.2910	0.0984
	Kernel	375	0.0240	-2.3102	0.0104	0.0480	-3.3566	0.0004	0.0267	-2.0732	0.0191	0.0933	-0.4303	0.3335
	Spline	375	0.0240	-2.3102	0.0104	0.0453	-3.5287	0.0002	0.0320	-1.5993	0.0549	0.0933	-0.4303	0.3335
	Sp+PT	375	0.0213	-2.5471	0.0054	0.0453	-3.5287	0.0002	0.0320	-1.5993	0.0549	0.0933	-0.4303	0.3335
45 days	LNM	319	0.0157	-2.8130	0.0025	0.0627	-2.2209	0.0132	0.0376	-1.0147	0.1551	0.0658	-2.0343	0.0210
	Kernel	319	0.0219	-2.2992	0.0107	0.0408	-3.5273	0.0002	0.0376	-1.0147	0.1551	0.0846	-0.9145	0.1802
	Spline	319	0.0188	-2.5561	0.0053	0.0251	-4.4605	0.0000	0.0408	-0.7578	0.2243	0.0878	-0.7279	0.2333
	Sp+PT	319	0.0188	-2.5561	0.0053	0.0251	-4.4605	0.0000	0.0408	-0.7578	0.2243	0.0878	-0.7279	0.2333
60 days	LNM	295	0.0136	-2.8718	0.0020	0.0475	-3.0081	0.0013	0.0203	-2.3375	0.0097	0.0881	-0.6793	0.2485
	Kernel	295	0.0102	-3.1389	0.0008	0.0407	-3.3963	0.0003	0.0169	-2.6046	0.0046	0.0780	-1.2615	0.1036
	Spline	295	0.0136	-2.8718	0.0020	0.0237	-4.3667	0.0000	0.0271	-1.8032	0.0357	0.0746	-1.4556	0.0728
	Sp+PT	295	0.0102	-3.1389	0.0008	0.0237	-4.3667	0.0000	0.0237	-2.0704	0.0192	0.0746	-1.4556	0.0728

The table shows the frequency in which the actual value of the underlying falls into the specific tail area, the statistic for the Brier test as well as the p-value (columns Freq., Stat. and p-value, respectively). All information is provided for a 5% and 10% probability mass levels for both left and right tails. Results provided in this table are for the Brier test calculated on the RNDs calculated using the different methodologies and maturities analyzed in this paper for the S&P500 index on the complete data set. Column N contains the number of days for which a RND has been implied.

Table 4: **Brier test results for the RNDs on S&P500**, excluding crisis periods

			Left tail						Right tail					
			5%			10%			5%			10%		
τ	Model	N	Freq.	Stat.	p-value	Freq.	Stat.	p-value	Freq.	Stat.	p-value	Freq.	Stat.	p-value
15 days	LNM	322	0.0031	-3.8610	0.0001	0.0652	-2.0805	0.0187	0.0311	-1.5598	0.0594	0.1025	0.1486	0.5591
	Kernel	322	0.0062	-3.6053	0.0002	0.0404	-3.5666	0.0002	0.0280	-1.8154	0.0347	0.1056	0.3344	0.6309
	Spline	322	0.0031	-3.8610	0.0001	0.0342	-3.9381	0.0000	0.0404	-0.7927	0.2140	0.1025	0.1486	0.5591
	Sp+PT	322	0.0031	-3.8610	0.0001	0.0342	-3.9381	0.0000	0.0404	-0.7927	0.2140	0.1025	0.1486	0.5591
30 days	LNM	326	0.0092	-3.3798	0.0004	0.0429	-3.4339	0.0003	0.0245	-2.1092	0.0175	0.0828	-1.0339	0.1506
	Kernel	326	0.0153	-2.8716	0.0020	0.0368	-3.8031	0.0001	0.0307	-1.6010	0.0547	0.0982	-0.1108	0.4559
	Spline	326	0.0123	-3.1257	0.0009	0.0337	-3.9877	0.0000	0.0368	-1.0927	0.1373	0.0982	-0.1108	0.4559
	Sp+PT	326	0.0092	-3.3798	0.0004	0.0337	-3.9877	0.0000	0.0368	-1.0927	0.1373	0.0982	-0.1108	0.4559
45 days	LNM	267	0.0075	-3.1871	0.0007	0.0562	-2.3868	0.0085	0.0412	-0.6599	0.2547	0.0749	-1.3668	0.0858
	Kernel	267	0.0150	-2.6255	0.0043	0.0375	-3.4067	0.0003	0.0412	-0.6599	0.2547	0.0936	-0.3468	0.3644
	Spline	267	0.0112	-2.9063	0.0018	0.0187	-4.4267	0.0000	0.0449	-0.3791	0.3523	0.0974	-0.1428	0.4432
	Sp+PT	267	0.0112	-2.9063	0.0018	0.0187	-4.4267	0.0000	0.0449	-0.3791	0.3523	0.0974	-0.1428	0.4432
60 days	LNM	240	0.0083	-2.9617	0.0015	0.0333	-3.4427	0.0003	0.0250	-1.7770	0.0378	0.0958	-0.2152	0.4148
	Kernel	240	0.0083	-2.9617	0.0015	0.0333	-3.4427	0.0003	0.0208	-2.0732	0.0191	0.0833	-0.8607	0.1947
	Spline	240	0.0083	-2.9617	0.0015	0.0167	-4.3033	0.0000	0.0333	-1.1847	0.1181	0.0833	-0.8607	0.1947
	Sp+PT	240	0.0083	-2.9617	0.0015	0.0167	-4.3033	0.0000	0.0292	-1.4809	0.0693	0.0833	-0.8607	0.1947

The table shows the frequency in which the actual value of the underlying falls into the specific tail area, the statistic for the Brier test as well as the p-value (columns Freq., Stat. and p-value, respectively). All information is provided for a 5% and 10% probability mass levels for both left and right tails. Results provided in this table are for the Brier test calculated on the RNDs calculated using the different methodologies and maturities analyzed in this paper for the S&P500 index on the restricted data set (in which crisis periods have been removed). Column N contains the number of days for which a RND has been implied.

Table 5: Berkowitz test statistic and Block-Bootstrap 95th and 90th percentiles for the S&P500

τ	Model	S&P500			NASDAQ 100			RUSSELL 2000		
$\chi^2_3(95\%) = 7.85$		LR_3	B-Boot.(95)	B-Boot.(90)	LR_3	B-Boot. (95)	B-Boot. (90)	LR_3	B-Boot. (95)	B-Boot. (90)
15 days	LMN	293.0455	325.4565	318.6257	185.4234	218.0242	211.8529	276.3819	314.0538	305.2469
	Kernel	290.6029	319.2546	313.1809	186.3965	214.9305	209.0587	277.7435	320.6566	311.5598
	Spline	280.7125	314.7229	307.8585	174.3888	207.4572	200.5460	270.1801	317.4453	308.0356
	Sp+PT	280.7953	316.2132	308.8502	173.0493	206.2505	199.6106	264.1271	319.9015	309.4195
30 days	LMN	330.6138	370.9677	361.6552	200.9971	236.9523	229.7752	257.6417	294.2178	286.0025
	Kernel	323.5491	359.8013	350.1886	191.4513	232.6513	223.8547	259.1183	297.7887	289.5883
	Spline	309.3494	351.3955	341.9296	187.2899	226.5716	218.5921	247.3581	290.1936	281.0755
	Sp+PT	309.5835	352.5865	342.5027	183.2907	225.0551	216.6904	245.2189	291.2501	282.1556
45 days	LMN	288.8683	325.4540	314.3464	179.6266	211.2872	204.2211	187.9370	217.9222	210.6144
	Kernel	277.1626	309.6674	300.4922	172.5721	203.2612	196.3735	190.4343	220.6708	213.7647
	Spline	270.1027	307.9507	296.5326	165.2227	196.6452	189.7925	182.6621	215.7416	208.0452
	Sp+PT	270.5135	307.6347	296.8349	163.8697	195.8229	188.0309	181.5916	214.2552	206.6340
60 days	LMN	270.1590	295.0415	285.9951	182.8043	197.2200	192.2441	205.3565	220.4922	213.6826
	Kernel	276.2661	302.2074	293.3276	176.1827	191.5402	186.4697	205.7683	222.0206	215.4375
	Spline	267.5461	297.8701	287.2185	171.7267	187.2500	181.8214	197.7991	215.6575	208.5868
	Sp+PT	268.0386	297.1935	288.4383	166.7008	183.1990	177.5672	197.5185	214.0516	207.5829

The table shows the Berkowitz test LR_3 statistic values for the different RNDs across maturities and indexes, as well as their corresponding block-bootstrap 95th and 90th percentiles. The 95% critical value of the χ^2 with 3 degrees of freedom is stated on the second row. Both the 95th and 90th percentiles are higher than the Berkowitz LR_3 statistic value, concluding that if the statistic is distributed as per block-bootstrap, both percentiles fail to reject the null hypothesis. The table compares this results across the different methodologies, indexes and maturities studied in this paper. The complete data set has been used to produce these results.

Table 6: Berkowitz test statistic and Block-Bootstrap 95th and 90th percentiles for the S&P500, excluding crisis periods

τ	Model	S&P500			NASDAQ 100			RUSSELL 2000		
$\chi^2_3(95\%) = 7.85$		LR_3	B-Boot.(95)	B-Boot.(90)	LR_3	B-Boot.(95)	B-Boot.(90)	LR_3	B-Boot.(95)	B-Boot. (90)
15 days	LMN	261.0704	289.0440	282.2690	159.2896	190.4702	184.8310	235.3139	269.9723	262.3063
	Kernel	260.2260	281.8389	276.9479	160.9476	188.9350	183.2188	235.9622	276.5424	268.4681
	Spline	249.2416	278.9527	271.5839	149.2765	180.9111	174.8985	227.6088	273.4686	264.7315
	Sp+PT	249.3369	277.4157	271.3729	147.9143	179.8141	174.1139	221.7908	272.3620	263.4449
30 days	LMN	281.1466	317.8839	309.0520	168.9042	202.5211	195.3909	217.5441	250.8509	244.1965
	Kernel	276.0939	309.3409	301.3133	160.7445	199.9242	191.5948	218.8134	253.1017	245.6558
	Spline	262.1469	300.7101	292.1207	156.7330	192.4020	185.2688	207.0850	245.9339	237.2866
	Sp+PT	262.3871	301.5724	292.7567	153.2677	190.4625	183.1035	205.0723	244.2736	236.2414
45 days	LMN	240.5731	277.2092	267.0114	152.3103	180.7984	174.6603	148.2819	171.1953	165.9782
	Kernel	231.3595	264.9043	255.6758	147.9226	177.2156	170.8150	150.9166	175.2659	169.7407
	Spline	224.1282	261.3162	250.9638	142.2700	170.0332	164.1404	143.8152	168.8273	163.0793
	Sp+PT	224.5074	258.9355	250.1648	141.0963	169.4100	163.4882	142.9454	167.8381	162.4578
60 days	LMN	216.0344	237.1528	229.8808	140.5529	155.5354	150.5096	156.0565	169.8769	164.4650
	Kernel	221.0851	243.1724	235.5457	137.1190	152.6372	147.5020	156.7342	167.1324	162.1388
	Spline	212.6918	237.9665	229.5559	133.5388	149.0193	143.8523	150.4351	160.8767	155.7161
	Sp+PT	213.1123	239.1166	229.9197	130.0278	146.1716	141.4404	150.2366	160.9419	155.7264

The table shows the Berkowitz test LR_3 statistic values for the different RNDs across maturities and indexes, as well as their corresponding block-bootstrap 95th and 90th percentiles. The 95% critical value of the χ^2 with 3 degrees of freedom is stated on the second row. Both the 95th and 90th percentiles are higher than the Berkowitz LR_3 statistic value, concluding that if the statistic is distributed as per block-bootstrap, both percentiles fail to reject the null hypothesis. The table compares this results across the different methodologies, indexes and maturities studied in this paper. The restricted data set which excludes the crisis periods has been used to produce these results.

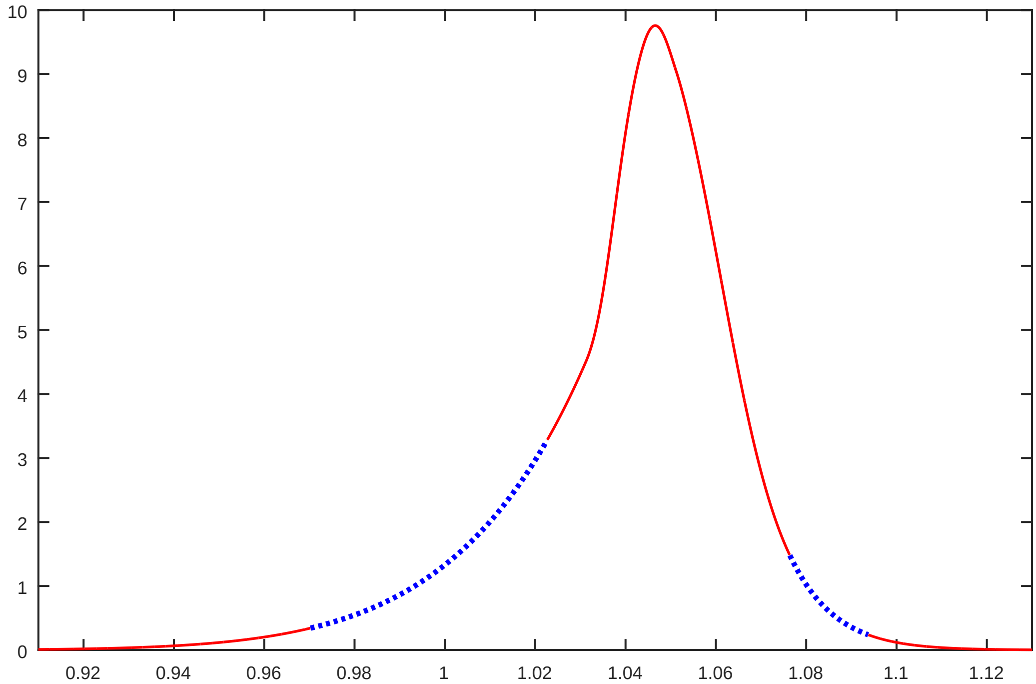

Figure 1: Kernel RND with pareto tails appended

The graph shows a RND calculated on the S&P500 for a 30 days time horizon using kernel technique with pareto tails appended. In the figure we can distinguish the central part represented with a solid line which is the main body of the distribution and it has been calculated using kernel method from the observed range of data. The most extreme regions also depicted in solid line show the pareto tails appended to the main body of the distribution. Finally the graph depicts with a dotted line the overlapping zone between α_0 and α_1 which has been estimated using a weighting scheme. The RND is for 17^{th} December 2009.

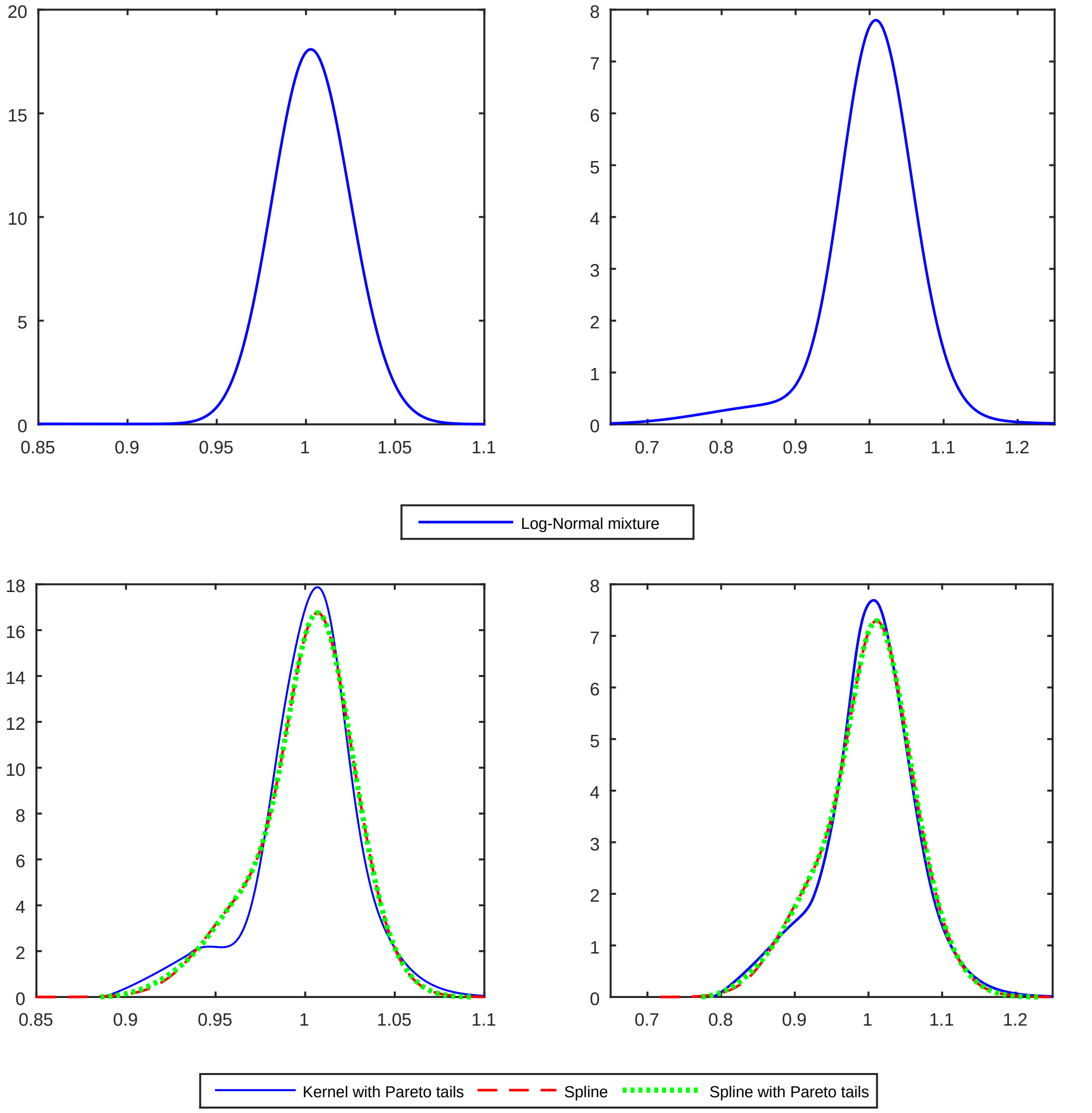

Figure 2: RNDs before and after the crisis

The figure compares a pre-crisis (21 July 2005) and a post-crisis (23 July 2009) RND for the $S\&P500$ at 30 days maturity for each of the different methodologies proposed. Both top panels depict the RNDs implied using the parametric Log-Normal mixture. The bottom plots show the non-parametric RNDs: Kernel with pareto tails appended (solid line), Spline with extrapolation (dashed line) and Spline with Pareto tails appended (dotted line).

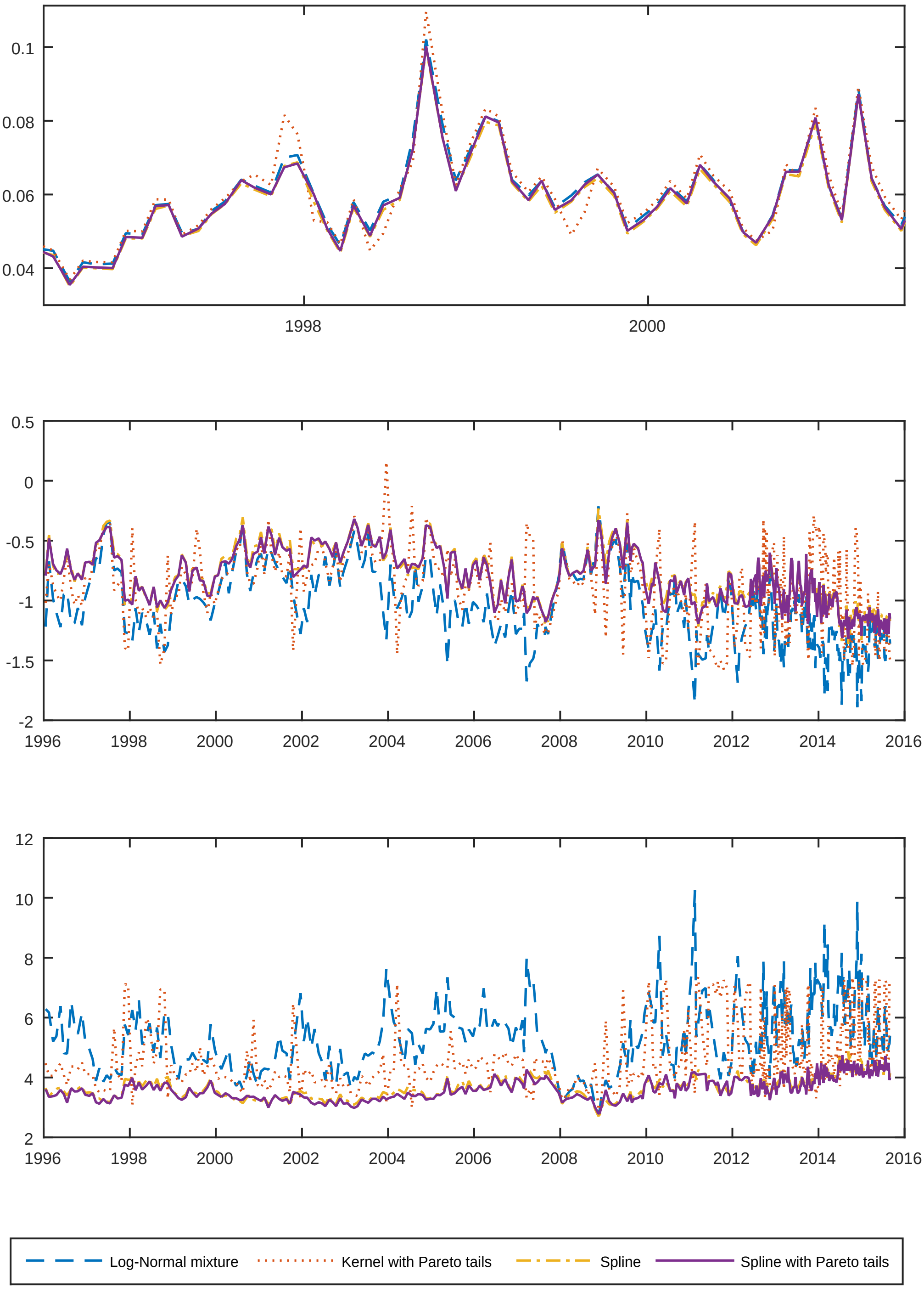

Figure 3: Standard deviation, Skewness and Kurtosis for the RNDs extracted using all different methodologies

The figure compares the 2nd, 3rd and 4th moments of the RNDs calculated on the S&P500 for a 30 days time horizon using all methodologies used in this study: Log-Normal mixture (dashed line), Kernel with Pareto tails (dotted line), Spline with extrapolation (dash-dot line) and Splines with Pareto tails (solid line). Standard deviation is depicted in panel 1, skewness in panel 2 and kurtosis in panel 3. They are represented along the whole sample period, from 1996 until 2015. We can see that all different methodologies yield to very similar results which confirm that RNDs are negatively skewed and present kurtosis higher than 3, concluding that RNDs depart from normality.