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

Antoni Vaello-Sebastià and M. Magdalena Vich-Llompart

Universitat de les Illes Balears

ERFIN, Warsaw 2017

Importance of Risk Neutral Distributions (RND)

- Security Valuation
- Risk Management
- Asset Allocation
- **Forecasting**

Mixed evidence regarding RND forecasting ability (based on Berkowitz test):

- **Bliss & Panigirtzoglou (2004)**
- **Anagnou et al. (2003)**
- **Alonso et al. (2005)**
- **Craig et al. (2003)**

Importance of Risk Neutral Distributions (RND)

- Security Valuation
- Risk Management
- Asset Allocation
- **Forecasting**

Mixed evidence regarding RND forecasting ability (based on Berkowitz test):

- **Bliss & Panigirtzoglou (2004)**
- **Anagnou et al. (2003)**
- **Alonso et al. (2005)**
- **Craig et al. (2003)**

Motivation & Marginal Contribution

- **Are Berkowitz test assumptions true?**
- Data set:
  - Longer and recent sample period: 20 years long (1996 to 2015) including two major crisis (2000 and 2007).
  - Three major indexes: S&P500, Nasdaq100 and Russell2000.
- Three different methodologies to extract RND: One parametric and two non-parametric (applied to the same data set)
- Estimation of the tails considered.

Motivation & Marginal Contribution

- **Are Berkowitz test assumptions true?**
- Data set:
  - Longer and recent sample period: 20 years long (1996 to 2015) including two major crisis (2000 and 2007).
  - Three major indexes: S&P500, Nasdaq100 and Russell2000.
- Three different methodologies to extract RND: One parametric and two non-parametric (applied to the same data set)
- Estimation of the tails considered.

Main approaches

- **PARAMETRIC APPROACH:** Relies on a prior known probability function whose parameters are then adjusted so as to obtain the best fit of the observed prices.
- **NON-PARAMETRIC APPROACH:** Based on Breeden-Litzenberger technique which allows us to obtain the whole state price density from observed option prices by taking the second partial derivative of the option pricing formula

Parametric Techniques

- **Log-Normal Mixture:** Flexibility and non-negativity ensured

$$f_Q(x) = p\Psi(x|F_1, \sigma_1, T) + (1-p)\Psi(x|F_2, \sigma_2, T)$$

Non-parametric Techniques

Based on Breeden-Litzenberger method which requires a continuum of Strike prices encompassing all future payoffs and smooth data.

$$f\left(S_T\right) = e^{r(T-t)} \frac{\partial^2 C\left(S_t, K, T, t\right)}{\partial K^2} \Big|_{K=S_T}$$

However:

- Options trade at discrete prices.
- Observations may contain noise.

→ Data needs to be **interpolated** and **smoothed**:

- Splines technique, Bliss & Panigirtzoglou, 2004
- Kernel smoothing, Aït-Sahalia and Lo, 1998
- Better to fit on (iv/δ) space, Malz, 1997

Non-parametric Techniques

Based on Breeden-Litzenberger method which requires a continuum of Strike prices encompassing all future payoffs and smooth data.

$$f(S_T) = e^{r(T-t)} \frac{\partial^2 C(S_t, K, T, t)}{\partial K^2} \Big|_{K=S_T}$$

However:

- Options trade at discrete prices.
- Observations may contain noise.

→ Data needs to be **interpolated** and **smoothed**:

- Splines technique, Bliss & Panigirtzoglou, 2004
- Kernel smoothing, Aït-Sahalia and Lo, 1998
- Better to fit on (iv/δ) space, Malz, 1997

example:

Non-parametric Techniques: Kernel Regression plus Pareto Tails

- **Kernel Regression:** Nadaraya-Watson

$$\hat{m}_h(x) = \frac{n^{-1}\sum_{i=1}^{n} K_h(x - x_i)Y_i}{n^{-1}\sum_{i=1}^{n} K_h(x - x_i)}$$

where K_h are the weights which determine the spread around x and depend on the smoothing parameter h

- **Tails:** Approximated with a Generalised Pareto Distribution (Birru and Figlewski, 2012)

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

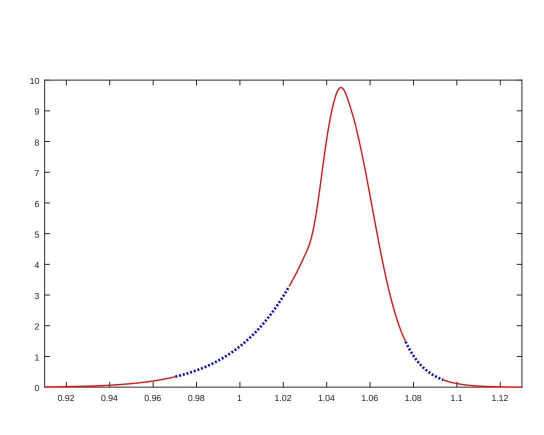
Non-parametric Techniques:Splines

- **Splines:**
  - cubic smoothing splines

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

    where m_i is a weighting value, Y_i is the implied volatility of the ith option observation, $g(\Delta_i, \theta)$ is the fitted implied volatility which is a function of $\Delta_i$and a set of spline parameters, θ. λ is the smoothing parameter, which following Bliss & Panigirtzoglou 2004, takes value 0.99, and $f''(x;\theta)^2$ is the smoothing spline.
  - three pseudo-points added at both ends of the available range (Bliss & Panigirtzoglou, 2004)
- **Tails:**
  - Once the spline is fitted, the iv / δ function is extrapolated
  - Pareto tails appended (as in Kernel approach)

Non-parametric methods: Kernel Regression and the tails



RND for the S&P500 with 30 days to maturity on 17^{th} December 2009.

The tests

- **Are extracted RNDs indeed equal to the true RNDs?**
  Berkowitz and Brier tests.

- **Are extracted RNDs contaminated somehow by the crisis anomalies?**
  Reproduce the analysis for a smaller sample excluding crisis periods:
  - March 2000 to October 2002
  - October 2007 to March 2009

Testing the forecasting ability

- **Berkowitz test:** Tests whether $\hat{f}_{t,\tau}\left(X_{t+\tau}\right) = f_{t,\tau}\left(X_{t+\tau}\right)$

- Berkowitz jointly tests independence and uniformity. It proposes the following auto-regressive model estimated using maximum likelihood estimates,

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

$$z_{t,\tau} - \mu = \rho\left(z_{t-1,\tau} - \mu\right) + \epsilon_{t,\tau}$$

Testing the forecasting ability

- To test for **normality** and **independence** the assumptions $\mu = \rho = 0$ and $\sigma^2(\epsilon_{t,\tau}) = 1$ must hold. The likelihood ratio statistic is given by

$$LR_3 = -2\left[L\left(0,1,0\right) - L\left(\hat{\mu},\hat{\sigma}^2,\hat{\rho}\right)\right] \text{ distributed as } \chi^2\left(3\right)$$

- To test for **independence** $\rightarrow \rho = 0$ and the likelihood ratio statistic is

$$LR_1 = -2\left[L\left(\hat{\mu},\hat{\sigma}^2,0\right) - L\left(\hat{\mu},\hat{\sigma}^2,\hat{\rho}\right)\right] \text{ distributed as } \chi^2\left(1\right)$$

Testing the tails

- **Brier Score:** To test the goodness of fit of the tails separately.

Based on the distance between the forecast probability mass and a binary variable, R, which takes value 1 if the true realization of the underlying falls into the tail being tested, or 0 otherwise.

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

About the tests

Do Berkowitz assumptions hold?

- **Block-Bootstrap simulations:**
  - Re-sampling with replacement
  - Simulation of 5,000 drawings with replacement from original sample (actual data).
  - Re-sampling a block of consecutive observations in order to preserve the dependence structure in the data.
  - Length of the block: $n^{1/3}$, being n the number of observations (Künsch, 1989).

The data

- 3 major indexes: S&P500, Nasdaq100 and Russell2000
- Daily observations from January 1996 until October 2015
- 4 time horizons, T = 15, 30, 45 and 60 days
- European Options
- mid-point bid-ask closing prices
- Discard:
  - bid and ask prices which are 0
  - options which do not satisfy non-arbitrage conditions
  - options with moneyness (K/F) outside the range 0.75 and 1.25
- Only ATM-OTM options.
- OTM puts are translated into their ITM counterpart through the put-call parity formula

$$F = (c - p)\, e^{r_f T} + K$$

The data: Filters

- RNDs must explain at least 70% of the probability,
- In case tails are appended, they must explain at least 99% of the probability,
- We require at least 8 observations per day in order to compute the RND estimation

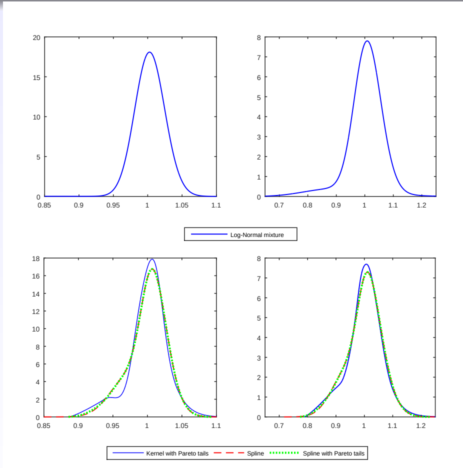
With the aim of having the largest sample to carry out the test, should the previous conditions fail, we allow the RND to have one extra day or one day less to maturity ($T+1$ or $T-1$).

The data: Filters

- RNDs must explain at least 70% of the probability,
- In case tails are appended, they must explain at least 99% of the probability,
- We require at least 8 observations per day in order to compute the RND estimation

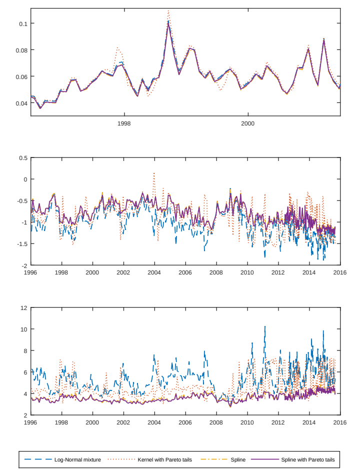
With the aim of having the largest sample to carry out the test, should the previous conditions fail, we allow the RND to have one extra day or one day less to maturity ($T + 1$ or $T - 1$).

Main results



RNDs for the S&P500 with 30 days to maturity on the 21 July 2005 (left) and 23 July 2009 (right)

Main results: the case for the S&P500



Berkowitz test p-values for the S&P500

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	0.0013	0.0007	0.0000	0.0000
	Kernel	0.0000	0.0000	0.0004	0.0003	0.0000	0.0000
	Spline	0.0000	0.0000	0.0004	0.0006	0.0000	0.0000
	Sp+PT	0.0000	0.0000	0.0004	0.0007	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	0.0000	0.0009	0.0001
	Spline	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sp+PT	0.0000	0.0000	0.0000	0.0000	0.0013	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	0.0000
	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
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

Berkowitz test p-values for the S&P500, excluding crisis periods

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	0.0001	0.0001	0.0000	0.0000
	Kernel	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Spline	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
	Sp+PT	0.0000	0.0000	0.0000	0.0001	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	0.0000	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
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	0.0000
	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
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

Berkowitz test results

- Both LR_3 and LR_1 statistics reject the null hypothesis $\rightarrow$ we can not ascertain the reason of rejection: poor forecasting ability or lack of independence of the transformed variable.

LR_1 rejects the null hypothesis of independence. One may suspect that observations indeed present some kind of **auto-correlation structure**. Should this be the case: Berkowitz assumptions would not be accurate

Berkowitz test statistic and Block-Bootstrap for the S&P500

τ	Model	S&P500			NASDAQ 100			RUSSELL 2000		
$\chi^2_{3,95\%} = 7.85$		LR_3	$BB_{95\%}$	$BB_{90\%}$	LR_3	$BB_{95\%}$	$BB_{90\%}$	LR_3	$BB_{95\%}$	$BB_{90\%}$
15 days	LMN	37.089	56.791	51.106	15.734	30.353	26.225	46.798	59.388	54.462
	Kernel	40.830	61.197	55.954	18.376	33.637	28.886	48.104	62.778	57.030
	Spline	42.435	62.207	57.062	18.055	33.366	29.355	50.621	66.590	60.923
	Sp+PT	42.610	63.273	57.503	18.275	33.372	29.363	36.592	63.438	56.236
30 days	LMN	58.840	87.300	78.329	26.951	46.834	41.516	27.186	46.592	41.123
	Kernel	63.959	93.908	85.125	30.550	51.410	45.324	16.507	44.535	37.015
	Spline	58.815	90.228	81.248	27.751	49.166	42.889	26.706	47.867	42.218
	Sp+PT	59.706	90.443	82.155	28.253	49.687	43.257	15.686	44.950	36.780
45 days	LMN	74.691	88.563	82.066	47.966	60.312	55.157	33.151	41.623	37.892
	Kernel	63.792	85.459	77.744	49.855	64.174	58.565	34.284	44.908	40.445
	Spline	75.398	89.657	83.564	50.389	65.868	60.164	34.032	43.919	39.874
	Sp+PT	75.361	91.390	84.146	51.376	67.337	61.035	33.878	44.215	40.283
60 days	LMN	112.039	114.398	107.881	73.926	82.252	76.774	71.856	74.598	69.106
	Kernel	114.031	118.013	110.929	74.820	85.662	79.086	70.178	75.586	70.561
	Spline	117.228	120.290	113.944	76.999	88.106	81.938	72.294	77.799	72.927
	Sp+PT	118.137	120.406	114.285	78.316	90.811	83.791	72.180	78.176	72.974

Berkowitz test statistic and Block-Bootstrap for the S&P500, rest. sample

τ	Model	S&P500			NASDAQ 100			RUSSELL 2000		
$\chi^2_{3,95\%} = 7.85$		LR_3	$BB_{95\%}$	$BB_{90\%}$	LR_3	$BB_{95\%}$	$BB_{90\%}$	LR_3	$BB_{95\%}$	$BB_{90\%}$
15 days	LMN	58.728	72.940	66.907	22.274	39.496	35.042	54.467	66.865	61.705
	Kernel	63.196	76.458	70.912	25.361	42.382	37.786	57.626	70.487	65.111
	Spline	64.158	77.117	72.317	24.189	41.926	37.568	58.938	73.456	67.916
	Sp+PT	64.749	78.354	73.457	24.329	43.341	38.305	57.829	73.205	67.418
30 days	LMN	65.347	94.723	85.247	37.419	58.479	51.960	44.097	59.526	54.620
	Kernel	71.704	100.386	91.721	40.078	60.522	54.363	47.158	64.341	57.969
	Spline	67.081	97.828	89.676	38.393	60.599	54.223	45.037	62.686	56.872
	Sp+PT	67.466	100.500	91.147	38.396	61.092	54.459	45.438	62.744	56.94
45 days	LMN	68.997	86.848	80.047	48.223	63.920	58.269	30.899	42.057	37.551
	Kernel	57.365	84.374	76.511	50.545	68.945	62.748	32.024	44.208	40.077
	Spline	71.066	88.472	82.772	52.255	71.565	65.284	32.826	45.046	40.886
	Sp+PT	70.657	90.921	83.505	53.000	72.416	66.267	33.284	46.050	41.673
60 days	LMN	100.802	106.240	100.591	60.959	73.298	67.178	57.443	62.628	57.735
	Kernel	103.383	109.882	103.162	63.768	78.509	72.115	55.901	62.573	57.062
	Spline	106.194	111.730	106.123	65.682	80.561	74.223	58.402	65.406	60.306
	Sp+PT	107.123	113.412	106.820	67.286	83.366	76.998	58.282	66.260	60.931

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
15d	LNM	374	0.008	-3.725	0.000	0.072	-1.793	0.073	0.029	-1.827	0.068	0.094	-0.414	0.679
	Kernel	374	0.008	-3.725	0.000	0.045	-3.516	0.000	0.027	-2.064	0.039	0.096	-0.241	0.809
	Spline	374	0.008	-3.725	0.000	0.037	-4.033	0.000	0.037	-1.115	0.265	0.094	-0.414	0.679
	Sp+PT	374	0.008	-3.725	0.000	0.037	-4.033	0.000	0.037	-1.115	0.265	0.094	-0.414	0.679
30d	LNM	375	0.021	-2.547	0.011	0.056	-2.840	0.005	0.021	-2.547	0.011	0.080	-1.291	0.197
	Kernel	375	0.024	-2.310	0.021	0.048	-3.357	0.001	0.027	-2.073	0.038	0.093	-0.430	0.667
	Spline	375	0.024	-2.310	0.021	0.045	-3.529	0.000	0.032	-1.599	0.110	0.093	-0.430	0.667
	Sp+PT	375	0.021	-2.547	0.011	0.045	-3.529	0.000	0.032	-1.599	0.110	0.093	-0.430	0.667
45d	LNM	319	0.016	-2.813	0.005	0.063	-2.221	0.026	0.038	-1.015	0.310	0.066	-2.034	0.042
	Kernel	319	0.022	-2.299	0.021	0.041	-3.527	0.000	0.038	-1.015	0.310	0.085	-0.914	0.360
	Spline	319	0.019	-2.556	0.011	0.025	-4.460	0.000	0.041	-0.758	0.449	0.088	-0.728	0.467
	Sp+PT	319	0.019	-2.556	0.011	0.025	-4.460	0.000	0.041	-0.758	0.449	0.088	-0.728	0.467
60d	LNM	295	0.014	-2.872	0.004	0.047	-3.008	0.003	0.020	-2.337	0.019	0.088	-0.679	0.497
	Kernel	295	0.010	-3.139	0.002	0.041	-3.396	0.001	0.017	-2.605	0.009	0.078	-1.261	0.207
	Spline	295	0.014	-2.872	0.004	0.024	-4.367	0.000	0.027	-1.803	0.071	0.075	-1.456	0.146
	Sp+PT	295	0.010	-3.139	0.002	0.024	-4.367	0.000	0.024	-2.070	0.038	0.075	-1.456	0.146

Brier test results

- **Left tail:** Bad fit.
  - RNDs overestimate the frequency: observed frequency lower than the predicted one.
- **Right tail:** Good fitting in general.
  - 10% significance level performs slightly better than the 5% $\rightarrow$ more extreme and less observations
- Crisis are not responsible for the results obtained
- Results consistent across the different methodologies, indexes and maturities

Conclusions

- All methods yield to very similar results.
- Berkowitz test rejects RNDs as being good forecasters: for both whole sample and non-crisis sample.
- Brier test concludes:
    - good fitting for the right tail
    - general rejection of the null hypthesis for the left tail

Conclusions

- Block-bootstrap suggests that Berkowitz main assumption is violated by the data, and so **rejection is not that obvious.**
- **Robustness:** Berkowitz and Block-Bootstrap reach the same conclusions
  - for any methodologies proposed to extract RNDs (parametric and different non-parametrics)
  - for any of the four maturity periods checked (15, 30, 45 and 60 days)
  - for any of the indexes contemplated (S&P500, Nasdaq100 and Russell2000)
  - for different lengths of blocks
- Possible extreme observations from crisis are not responsible for the results obtained

The end

Thank you for your attention.