Ioana Andreea DUCA, PhD Candidate
Professor Gheorghe RUXANDA, PhD
E-mail: ghrux@ase.ro
The Bucharest Academy of Economic Studies

CHOOSING THE APPROPRIATE WINDOW LENGTH FOR ESTIMATING PHYSICAL DENSITIES

Abstract: *This paper focuses on the estimation of physical densities using historical returns of index prices. Specifically, we suggest using a kernel estimator for estimating the objective densities and Berkowitz test for determining the appropriate window length. We apply the proposed methodology to Romanian BET Index data. The test manages to capture the jump in the data and a recommendation is done with respect to the length of the window to be used in the estimation procedure.*

Key words: *physical density, risk neutral density, kernel estimators*.

JEL Classification: C13, C14, G12

1. Introduction

Whether stock returns are predictable or not represents a question tackled by researchers for a long time. With time, interest has slid from point estimates of returns to density estimates which can further be used for asset pricing. The literature distinguishes between physical or objective density and risk neutral density. The physical density represents an assessment of the agent acting on the financial market over the possibility of a certain state of nature to occur. The risk neutral density or state price density, which is the term generally used in preference based equilibrium models, is related to the concept of Arrow-Debreu Securities. The Arrow-Debreu securities pay one unit if a particular state of nature occurs and zero otherwise. As stated in Jackwerth (2000), there is a link between the two densities: the risk neutral density is equal to the physical density corrected by a risk adjustment factor.

Initially, this relationship was used to evaluate the risk aversion of the agent, but more recently it has been used for the computation of physical density as well. For instance, Bliss and Panigirtzoglu (2004) and Kang and Kim (2006) consider that the ratio between risk neutral density and physical density is equal to a pricing kernel, which is calculated based on utility functions. The risk neutral density is calculated based on the pricing formula of an European call option. Having the kernel and the risk neutral density they compute the physical density. Garcia et al (2011) link the two densities through an integral martingale using high frequency data available on an underlying asset together with option data.

Unlike risk neutral densities which can typically be estimated using option data, we have the possibility to estimate physical density from historical asset price data, which is much easier to obtain. Jackwerth (2000) estimates the density through a kernel density estimator using 31 day non-overlapping returns over a four year sample. Prasanna and Vause (2007) use a threshold-GARCH model of the returns of the S&P500 index.

In this paper, we will use kernel density estimators to estimate the physical density, but the main focus will lie on determining an appropriate window length which can be used for density estimation. More and more often we confront with situations where we need to deal with jumps or spikes present in the data, which generally cause a change in the estimated parameters. Local adaptive methods and rolling window techniques have been developed to deal with this kind of problems. In this case, we suggest using a simpler particular solution by using the Berkowitz (2001) test. This test was developed to evaluate whether the estimated density equals the true density. In the following it will be shown how it can be used to determine the appropriate window length to be employed in estimation. Romanian BET index data is utilized.

The next section presents several methodologies that can be used in physical density estimation. The third section describes the Berkowitz test and the way it can be employed to serve our purpose. Section four presents an application and the last section concludes.

2. Estimation of physical densities of stock returns

Current literature offers two main approaches in terms of estimating the physical (objective) distribution of asset returns. The first and perhaps the simplest one is based on historical prices. This is why physical density is sometimes called historical density. The second one consists in behavioral based models via risk neutral density corrected by pricing kernels. Both methods will be referred to in this section.

The method of kernel density estimation can be applied to stock prices returns in order to obtain a smooth, continuous probability density function (pdf). Härdle et al (2004) mention that the idea behind kernel density estimation lies in finding an interval around x, not an interval including x. Thus, the estimated pdf can be written as:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1}$$

where K is the kernel function, h is the bandwidth, x is the point at which the function is estimated, while n is the sample size. Considering $u = (x - X_i)/h$ several forms of the kernel function can be found in Table 1.

Table 1. Kernel functions

Kernel	$K(u)$
Uniform	$\frac{1}{2} I(\lvert u \rvert \le 1)$
Triangle	$(1 - \lvert u \rvert) I(\lvert u \rvert \le 1)$
Epanechnikov	$\frac{3}{4}(1 - u^2) I(\lvert u \rvert \le 1)$
Quartic (Biweight)	$\frac{15}{16}(1 - u^2)^2 I(\lvert u \rvert \le 1)$
Triweight	$\frac{35}{32}(1 - u^2)^3 I(\lvert u \rvert \le 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$
Cosine	$\frac{\pi}{4}\cos\left(\frac{\pi}{2}u\right) I(\lvert u \rvert \le 1)$

Härdle et al (2004) show that the choice of the kernel function is not so important in empirical studies. Basically, if we want to obtain pdf estimates based on two different functions that would have the same degree of smoothness, we would have to multiply one of the bandwidths with an adjustment factor. On the other hand, the choice of the bandwidth h seems rather crucial. The choice of a bandwidth which is too small may render a non-continuous estimate of the density function. Choosing a bandwidth that exceeds the optimal bandwidth may cause an over smoothing of the probability distribution function. Thus, choosing the optimal bandwidth is very important. One of the criterions that can be used is minimizing the cross-validation criterion:

$$CV(h) = \int \hat{f}_h^2(x)dx - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, i\neq j}^{n} K\left(\frac{X_i - X_j}{h}\right) \qquad (2)$$

To estimate the density more complicated techniques can be used as in Ruxanda and Smeureanu (2012). They present an expected maximization algorithm which can be used in parametric estimation of probability distribution function of a mixture of distributions.

Bliss and Panigirtzoglu (2004) propose a different approach, starting from a very simple equation:

$$\zeta = \frac{q}{p} \qquad (3)$$

where ζ is the pricing kernel, q represents the risk neutral density, while p is the physical density. The idea stated in this paper is that knowing any two of the three permits us to infer over the third one.

The risk neutral density is obtained from option prices. Options based on the same underlying asset provide different prices for different strikes, so that a pdf of the prices of that asset can be estimated. These pdf's have the advantage that they represent forward looking forecasts of the distribution of prices of the underlying prices, but at the same time they convey different information than the physical densities of stock prices. They generally do not correspond to the beliefs of the agents on the market about future price movements. Therefore, the risk neutral densities need to be corrected by a pricing kernel. Bliss and Panigirtzoglu (2004) use a pricing kernel that is based on utility functions, more specifically the power and exponential utility functions. In addition, Kang and Kim (2006) use the HARA utility function, the log plus utility function and the linear plus exponential utility function.

Breeden and Litzenberger (1978) show that the risk neutral pdf can be obtained from the price function of an European call option. Given the European call price option formula:

$$C(X,\tau,r_{t,\tau},\delta_{t,\tau},S_t) = e^{-r_{t,\tau}\tau}\int_0^{\infty}\max(S_T - X,0)\,q(S_T|\tau,r_{t,\tau},\delta_{t,\tau},S_t)\,dS_T \qquad (4)$$

where S_t is the underlying asset price at time t, X is the strike price, τ is the time to maturity, $T = t + \tau$ is the expiration date, $r_{t,\tau}$ is the risk free rate and $\delta_{t,\tau}$ is the corresponding dividend. Therefore, we can write the risk neutral density:

$$q(S_T) = e^{r\tau}\frac{\partial^2 C}{\partial X^2}\Big|_{X=S_T} \qquad (5)$$

It is worth mentioning that option quotes do not provide a continuous price function. As a consequence smoothing techniques need to be applied either in option prices or implied volatilities. More details can be found in Grith et al (2009).

3. Testing density estimates

Another challenge is to test whether the estimated distribution fits the real one. There are several possibilities available in the literature to do so. All of them though, start from the transformation of Rosenblatt (1952) below:

$$y_t = \int_{-\infty}^{x_t}\hat{f}_t(u)du \qquad (6)$$

The essential property is that the return series x_t, which generally follows a heavy tail distribution, is transformed into a series y_t which is iid and uniformly distributed on (0,1). This series is easier to analyze. There are several tests which have been proposed to check the uniformity, but Berkowitz (2001) proposes a test which is able to account for both uniformity and independence. This statistic requires an additional transformation:

$$z_t = \Phi^{-1}(y_t) = \Phi^{-1}\left(\int_{-\infty}^{x_t}\hat{f}_t(u)du\right) \qquad (7)$$

Under the null hypothesis that $\hat{f}_t(\cdot) = f_t(\cdot)$, z_t follows a $N(0,1)$ distribution. Berkowitz (2001) further uses this transformation by looking at the $AR(1)$ model:

$$z_t - \mu = \rho(z_{t-1} - \mu) + \varepsilon_t \qquad (8)$$

The estimated parameters are then employed to compute a likelihood ratio test with the null hypothesis:

$$H_0: \begin{array}{c} \mu = 0 \\ \sigma^2 = \operatorname{var}(z_t) = 1 \\ \rho = 0 \end{array}$$

Knowing that the form of the likelihood function is:

$$L = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left[\sigma^2/(1-\sigma^2)\right] - \frac{(z_1 - \mu/(1-\rho))^2}{2\sigma^2/(1-\rho^2)} - \frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - \sum_{t=2}^{T}\left(\frac{(z_t - \mu - \rho z_{t-1})^2}{2\sigma^2}\right) \quad (9)$$

the statistic test can be defined as:

$$LR = 2(L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho}) - L(0,1,0)) \quad (10)$$

Under the null hypothesis the statistic follows a $\chi^2(3)$ distribution. It can also be reduced as a simple test for independence if:

$$LR_{ind} = 2(L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho}) - L(\hat{\mu}, \hat{\sigma}^2, 1)) \quad (11)$$

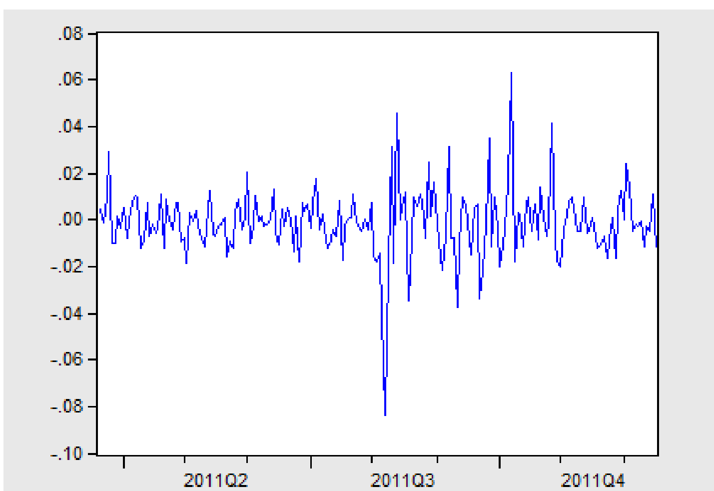
which follows a $\chi^2(1)$ distribution. Bliss and Panigirtzoglu (2004) prove through Monte Carlo simulations that this test is quite reliable in small samples as well. In large samples with autocorrelated data the Berkowitz test rejects with near certainty, while in small samples the authors find that the test rejects slightly more frequently than with uncorrelated data, with the rejection rate increasing in the degree of autocorrelation. In this paper, Berkowitz statistic is used to investigate whether the estimated densities for different window lengths are the same with the true densities. In addition, the p-value of the test is used as criteria for determining which estimated density is closest to the true one. Thus, the most appropriate window length is chosen such that the p-value of the test is maximized.

4. Empirical study on Romanian BET Index

This empirical study is focused on the BET index characterizing Bucharest Stock Exchange. We use data from March, 21st 2011 to December, 26th 2011, that is a period of 200 days. We work with returns of the index. Our final objective is to determine an appropriate physical density estimate for the BET index by using the Berkowitz test, briefly described in the previous section.

The plot in Figure 1 shows the evolution of the returns of BET index. It is immediately observed the spike at the middle of the third quarter. Also, it can be easily noticed that the market becomes significantly more volatile starting with this point in time. As a note, the jump corresponds to August 2011 stock markets' fall across the United States, Middle East, Europe and Asia. The fall reflects concerns of the participants on the stock markets with respect to European sovereign debt crisis associated with decreases in Standard & Poor's rating of United States of America, France, Italy or Spain.

Figure 1. BET index stock returns



The normal kernel density estimator was selected to be used in this study. The choice of the kernel estimator however, does not have a significant impact on the probability distribution of the data as it can be seen in Figure 2, which shows a comparison among the normal, triangle and Epanechnikov kernel. There is a slight difference in the peak of the distribution, but the tails are almost identically represented. Choosing the optimal bandwidth is more important. For a 100 days estimation window the optimal bandwidth is $h = 0.0039$. In Figure 3 it can be observed that choosing an underestimated or an overestimated bandwidth leads to significant differences in peak estimation and also in some tails estimation differences.

In the remaining of the paper, we will use only the normal kernel estimator. To continue with, we stress over the importance of choosing an appropriate window length for estimating the probability distribution function. Figure 4, shows the estimated pdf's for a window of 50 days (f1), 100 days (f2), 200 days (f3). There is an obvious difference in the mean and skewness. The 50 day window length density

Figure 2. Normal, Triangle, Epanechnikov density estimators

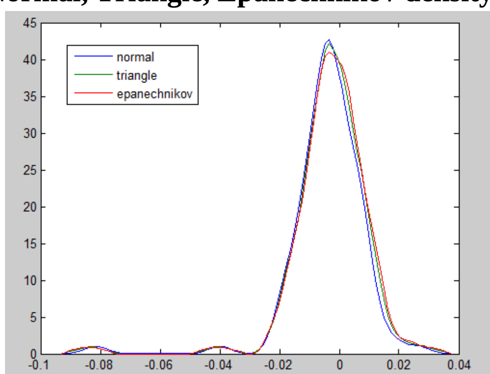


Figure 3. Normal kernel estimator with different bandwidths

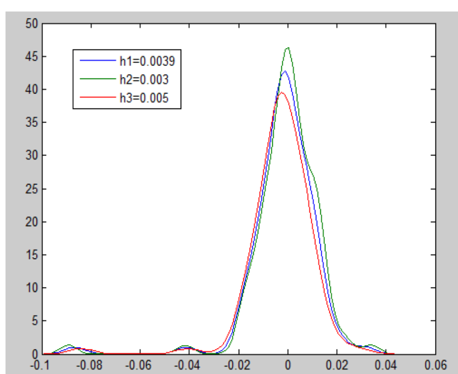
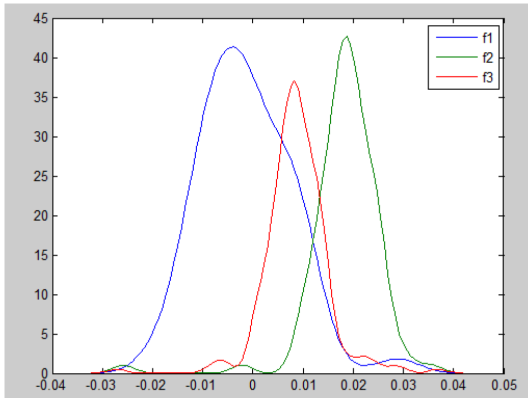


Figure 4. **Density estimates for different window lengths**



estimator is skewed to the right, while the other two estimators are skewed to the left.

In the following we have computed p-values of the Bekowitz test for different window lengths (Table 2). The first window length is of 10 days and at each step we go behind with another 5 days. One can observe that the p-values associated with the first 100 days are significantly higher than the p-values for the rest of 95 days. This corresponds to the two period volatility clustering that has been identified before. We mention though that for the smallest window lengths the test may suffer some sample size bias and therefore accept the null hypothesis too often. Note that the test does not reject for any of the window lengths, but we have to specify that we have used in sample data.

Further, notice the p-value that was obtained for the 100 days window length. It represents the highest value and it is very close to 1. Thus the test statistic does not reject the null hypothesis with a probability close to 1. What is more striking is that the p-value for the next period suffers a significant fall to 0.1654. The explanation for this behavior is the event that we have identified for the third quarter 2011 data. It is clear that data before and after this event have different distributions and therefore our methodology recommends choosing for estimation the 100 days window available before after August 2011.

5. Conclusions

In this paper, several possibilities for estimating physical or objective probability distribution functions were introduced. It has been shown that apart from choosing an appropriate kernel density estimator and an optimal bandwidth, it is critical to use a window length that is most suited for the set of data to be studied. The Berkowitz test is used to evaluate the fit of the estimated density to the data. The p-value of the statistic is taken as criteria to find the most appropriate window length. Romanian BET Index is used for an application. The test manages to capture the fall in the returns series as well as the change in volatility of the data, therefore proving the usefulness of such a methodology for determining the appropriate window length.

Table 2. Window lengths and associated p-values of the Berkowitz test

Window Length	P-value	Window length	P-Value
10	0.0966	105	0.1654
15	0.1758	110	0.1455
20	0.8727	115	0.1287
25	0.9664	120	0.1834
30	0.623	125	0.131
35	0.7066	130	0.1194
40	0.3952	135	0.1996
45	0.1426	140	0.2003
50	0.3631	145	0.2117
55	0.6254	150	0.3028
60	0.8825	155	0.3006
65	0.8849	160	0.2124
70	0.8287	165	0.2154
75	0.7861	170	0.1791
80	0.6524	175	0.1575
85	0.4684	180	0.2759
90	0.494	185	0.2973
95	0.5475	190	0.2532
100	*0.9949*	195	0.275

Acknowledgements

This work was cofinaced from the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/107/1.5/S/77213 „Ph.D. for a career in interdisciplinary economic research at the European standards"

REFERENCES

[1] **Berkowitz, J. (2001),** ***Testing Density Forecasts with Application to Risk Management***. *Journal of Business and Economics Statistics* 19, 465-474;
[2] **Bliss, R.R. and Panigirtzoglou, N. (2004),** ***Option-Implied Risk Aversion Estimates***. *Journal of Finance* 59, no.1 : 407-446;
[3] **Breeden, D. and Litzenberger, R.H. (1978),** ***Prices in State Contingent Claims Implicit in Option Prices*** . *Journal of Business* 51, 621-651;
[4] **Garcia, R., Lewis, M.A., Pastorello, S., Renault, E. (2011),** ***Estimation of Objective and Risk Neutral Distributions Based on Moments of Integrated Volatility*** *Journal of Econometrics*, 160, 22-32;
[5] **Grith, M., Härdle, W. K. and Park, J. (2009),** ***Shape Invariant Modelling Pricing Kernels and Risk Aversion*** **.** *Invited resubmission to Journal of Financial Econometrics on 12 April 2012*;
[6] **Härdle, W., Müller, M., Sperlich, S., Werwatz, A. (2004) ,** ***Nonparametric and Semiparametric Models*** . *Springer Series in Statistics;*
[7] **Jackwerth, J.C. (2000)** , ***Recovering Risk Aversion from Option Prices and Realized Returns.*** *The Review of Financial Studies*, 13, no. 2, 433-451;
[8] **Kang, B.J. and Kim, T.S. (2006) ,** ***Option Implied Risk Preferences: An Extension to Wider Classes of Utility Functions*** . *Journal of Financial Markets,* 9, 180-198;
[9]**Lakštutien A., Krušinskas R., Rumšaitė D. (2011),** ***The Influence of Deposits Insurance on the Stability of the Baltic States Banking System***, Volume 12, Issue 3, *Journal of Business Economics and Management, Taylor & Francis Ed.*, Co-Published with Vilnius Gediminas Technical University;
[10] **Prassana, G., Vause, N. (2007),** ***Measuring Investors' Risk Appetite***. *Munich Personal RePEc Archive*,818, 1-18;
[11]**Ruxanda, G. and Smeureanu, I. (2012),** ***Unsupervised Learning with Expected Maximization Algorithm***; *Economic Computation and Economic Cybernetics Studies and Research,* 1, *ASE Publishing*, Bucharest;

[12]**Smeureanu I., Dioşteanu A., Delcea C., Cotfas L. (2011), *Ontologii de afaceri pentru evaluarea responsabilității corporațiilor***, Amfiteatru Economic, no. 29 *Academia de Studii Economice din Bucuresti;*
[13]**Stankevičienė J., Mencaitė E. (2012), *The Evaluation of Bank Performance Using a Multicriteria Decision Making Model: A Case Study on Lithuanian Commercial Banks*** ; *Technological and Economic Development of Economy* , Volume 18, Issue 1; *Taylor & Francis Ed.* , Co-Published with Vilnius Gediminas Technical University.