Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v9n2p400

Kernel density smoothing using probability density functions and orthogonal polynomials
By Ouamaliche, Sayah

Published: 14 October 2016

Electronic Journal of Applied Statistical Analysis
Vol. 09, Issue 02, October 2016, 400-416
DOI: 10.1285/i20705948v9n2p400

Kernel density smoothing using probability density functions and orthogonal polynomials

Ouamaliche Soufiane*and Sayah Awatef

Mohammed V University in Rabat
Faculty of Sciences
Laboratory of Mathematics, Computing and Applications
BP1014RP, Rabat, Morocco

Published: 14 October 2016

This article is the first of a series devoted to providing a way to correctly explore stock market data through kernel smoothing methods. Here, we are mainly interested in kernel density smoothing, our approach revolves around introducing and testing the goodness of fit of some non-classical kernels based on probability density functions and orthogonal polynomials, the latter ones are of interest to us when they are of order two and above. For each kernel, we use a modified version of the "rules of thumb" principle in order to compute a smoothing parameter that would offer optimal smoothing for a reasonable computational cost. Compared to the Gaussian kernel, some of the tested kernels have provided a better Chi-Square statistic, especially the kernels of order 2 based on Hermite and Laguerre polynomials. These results are illustrated using data from the Moroccan stock market.

keywords: kernel smoothing, higher order kernels, probabilistic kernels, orthogonal polynomials, goodness of fit test, MASI index.

1 Introduction

1.1 General Introduction

Classical statistical methods are based upon parametric assumptions: it can be assumed that the data obey a certain known distribution defined by one or more parameters, at

*Corresponding author: ouamalich@hotmail.fr

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

least one of which is presumed to be unknown and must be estimated. However, the studied process or population might not fit into any known distribution and by adopting a parametric approach we can be led into making premises that are not coherent with the data at hand. These wrong assumptions might cause inaccurate approximations of the underlying statistical structure. Therefore, statistical methods which can be applied without any regards to the actual distribution of data are needed; these techniques consist in relying on the observations in order to estimate an unknown function belonging to a certain class of functions.

In the next section we describe rapidly the problem that we are addressing through nonparametric estimation then we give a short overview of how the studied approach can be useful in the exploration, presentation and analysis of data.

In section 2, we discuss briefly the foundation of kernel density estimation then we introduce the probabilistic and higher order kernels that we plan on using. Section 3 is devoted to the bandwidth/smoothing parameter, we present a classical estimation technique before stating the formula that builds on the "rules of thumb" principle. Finally in section 4, we proceed to some numerical experiments using some of the suggested kernels then we use the Chi-Square test in order to assess the goodness-of-fit for each kernel.

1.2 Nonparametric Density Estimation

Probability density estimation is one of the most fundamental problems in statistics. Consider any identically distributed random variables whose common distribution is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$, the problem of density estimation is to estimate f, an unknown function from $\mathbb{R}$ to $[0, +\infty[$. An estimator of f is a function $x \longrightarrow f_n(x, X_1, X_2, ..., X_n)$ measurable with respect to the observations $X = (X_1, X_2, ..., X_n)$.

On one hand, the parametric approach to this problem is assuming that f belongs to a known family $\{g(x, \theta) : \theta \in \Theta\}$, where $g(.,.)$ is a given function and Θ is a subset of $\mathbb{R}^k$ with a fixed dimension k independent of n. Then estimating f is equivalent to estimating θ. On the other hand if such assumptions about f are ignored then we would be dealing with a nonparametric approach where we usually assume that f belongs to a massive class $\mathbb{F}$ of densities such as all the continuous probability densities on $\mathbb{R}$. The main point of the latter approach is to get rid of all the restrictions imposed by the parametric methods, thus letting the data speak for themselves.

1.3 Kernel Smoothing In The Exploration And Presentation Of Data

In many contexts it is common to have at our disposal an important quantity of data presenting a wild dynamic that needs to be explained or analyzed.

Kernel smoothing provides a reliable way of displaying data without any prior opinions, its most simple use consists in graphically illustrating conclusions that might've been reached by other means. It may as well be used for analyzing data in an informal way in order to investigate important features such as multimodality, skewness, kurtosis,

stochastic effect...This preliminary analysis can lead to some valuable conclusions or point the way to some new paths of research. However, the real purpose of kernel smoothing lies in the framework of simulations and forecasting for it makes a strong forecasting tool that contributes toward improving in a significant way the predictability of erratic processes behavior.

2 Kernel Density Estimation

2.1 General Overview And Error Measurements

General overview:

Let X be a random variable on $\mathbb{R}$ and for all Borel sets of $\mathbb{R}$ we can write $\int_{\Omega} f(x)\,dx = P(x \in \Omega)$. As stated earlier, the task of nonparametric density estimation is to estimate f from a random $i.i.d$ sample $\{X_1, X_2, ..., X_n\}$.

Kernel density estimation is a way of carrying out this task, the simplest idea of a kernel estimator is to retrieve the cardinal of the observations that fall into a certain interval centered at the point where we wish to estimate the density:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2h} \mathbb{1}(x - h \leq X_i \leq x + h) \tag{1}$$

We can rewrite (1) as:

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K_0\left(\frac{X_i - x}{h}\right) \tag{2}$$

Where

$$K_0(u) = \frac{1}{2}\mathbb{1}(-1 \leq u \leq 1) \tag{3}$$

The function K_0 is called the kernel function. Nevertheless, kernels are not restricted to the example given in (3) since all functions $K : \mathbb{R} \longrightarrow \mathbb{R}$ satisfying the condition $\int K(u)\,du = 1$ can be considered as kernels even though we can add some restrictions concerning symmetry and non-negativity by imposing some conditions on the kernel moments μ_i, these moments can be expressed as:

$$\mu_i = \int u^i K(u)\,du \tag{4}$$

Some common kernels are listed below:

Silverman kernel: $K(u) = \frac{1}{2}\exp\left(\frac{-|u|}{\sqrt{2}}\right)$

Epanechnikov kernel: $K(u) = \frac{3}{4}\left(1 - |u^2|\right)\mathbb{1}(|u| \leq 1)$

Biweight kernel: $K(u) = \frac{3}{4}\left(1 - |u^2|\right) \mathbb{1}(|u| \leq 1)$

Gaussian kernel: $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$

Let us note that many of the common kernels are special cases of the polynomial family:

$$K_s(u) = \frac{(2s+1)!!}{2^{s+1}s!}\left(1 + u^2\right)^s \mathbb{1}(|x| \leq 1) \tag{5}$$

Measuring the error:

A well-known way of measuring the error in the estimation process of kernel density estimation is the L_2 norm or Integrated Squared Error (*ISE*), the *ISE* can be averaged over the support of the density in order to get the Mean Integrated Squared Error (*MISE*):

$$\begin{array}{rcl} MISE\left(\hat{f}_n\right) & = & \frac{1}{nh}\int K^2(x)\,dx + \left(1 - \frac{1}{x}\right)\int (K_h * f)^2(x)\,dx \\ & & -2\int (K_h * f)(x) f(x)\,dx + \int f^2(x)\,dx \end{array} \tag{6}$$

Where $K_h(u) = h^{-1}K\left(\frac{u}{h}\right)$ and $*$ denotes the convolution such that:

$$(f * g)(x) = \int f(x-y)\, g(y)\, dy$$

There has been much controversy over the suitability of *MISE* as a measure of error (Grund et al., 1994), but it has been shown that it allows very deep analysis and thus it is a reliable measure (Marron and Wand, 1992).

However (6) makes it relatively hard to evaluate the influence of the bandwidth on the performance of the estimator, and since the *MISE* is mainly used for finding an optimal smoothing parameter as we'll see in an upcoming section, we need a more tractable formula.

Usually for large samples and under standard technical assumptions (Silverman, 1986) asymptotic analysis provides a less complicated formula which is the Asymptotic Mean Integrated Squared Error (*AMISE*):

$$AMISE\left(\hat{f}_n\right) = (nh)^{-1} R(K) + \left(\frac{1}{4}\right) h^4 R\left(f''\right) \mu_2(K)^2 \tag{7}$$

We adopt the notations $\mu_2(K) = \int u^2 K(u)\,du$ and $R(K) = \int K(u)^2\,du$ here and in the following.

2.2 Probabilistic Kernels

Building on the idea that has led to the Gaussian kernel we introduce two kernels that are based on probability laws. The choice of these laws has been guided by previous knowledge of the studied phenomenon and since we are trying to estimate the density

of stock market data, our first approach consists in using a classical result of financial mathematics which states that under some quite restrictive assumptions stock prices are governed by a lognormal law, for more details one can consult (Di Nunno and Øksendal, 2011). The second approach follows from the Pareto principle which, in the case of stock markets, could be seen as if 80% of the variations of an index are being caused by 20% of its constituents (Newman, 2006).

Lognormal kernel:

Let X be a random variable on $\mathbb{R}$, the lognormal law $LN\left(\mu,\sigma^2\right)$ is the distribution of $\exp\left(X\right)$ when $X \sim N(\mu,\sigma^2)$ where $N(\mu,\sigma^2)$ denotes the normal law with mean μ and variance σ^2.

The probability density function of the lognormal law $LN\left(\mu,\sigma^2\right)$ is given by:

$$f_X\left(x\right) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-\left(\log\left(x\right)-\mu\right)^2}{2\sigma^2}\right) \tag{8}$$

It's clear that the Gaussian kernel is directly extracted from a standard normal distribution, when adopting the same approach we could write the lognormal kernel as:

$$K\left(u\right) = \frac{1}{u\sqrt{2\pi}} \exp\left(\frac{-\left(\log\left(u\right)\right)^2}{2}\right) \tag{9}$$

One could eventually keep the parameters of the law and estimate those using parametric methods such as the maximum likelihood method or the method of moments.

Pareto kernel:

The Pareto law is an example of a power law that gives a mathematical formalism to the $80-20$ principle which states that a large part of the effects is caused by a small part of the causes.

The probability density function of a Pareto law of type(I) $P\left(k,x_m\right)$ is given by:

$$f_X\left(x\right) = k\frac{x_m^k}{x^{k+1}}; x \geq x_m \tag{10}$$

Where x_m is the minimal value that X could take and k is the Pareto index.

In order to suggest an adequate kernel we need to properly estimate the Pareto index, we could either use the standard value $k_s = log_4\left(5\right)$ which satisfies the $80-20$ principle or estimate a value $\hat{k}$ through the maximum likelihood method which gives the estimator $\hat{k} = n\left[\sum_{i=1}^{n}\log\left(x_i\right) - n\log\left(x_m\right)\right]^{-1}$, in the latter case the Pareto kernel is given by:

$$K\left(u\right) = \left\{\left[\frac{\sum_{i=1}^{n}\log\left(u_i\right)}{n} - \log\left(u_m\right)\right]u\right\}^{-1}\left(\frac{u_m}{u}\right)^{\left[\frac{\sum_{i=1}^{n}\log(u_i)}{n}-\log(u_m)\right]^{-1}} \tag{11}$$

We can easily notice that both of the suggested probabilistic kernels are directly derived from the corresponding probability density functions, thus we do not need to prove that they satisfy the condition $\int K(u)\,du = 1$.

2.3 Higher Order Kernels Based On Orthogonal Polynomials

Let $l \geq 1$ be a natural number, a function $K : \mathbb{R} \longrightarrow \mathbb{R}$ is said to be a kernel of order l if the functions $u \longrightarrow u^i K(u), i = 0, 1, 2, ..., l$ are integrable and satisfy:

$$\int K(u)\,du = 1\ ;\ \int u^i K(u)\,du = 0\ ;\ i = 1, 2, ..., l \tag{12}$$

There exist some authors that impose for a kernel of order l to satisfy $\int u^{l+1} K(u)\,du \neq 0$, however we don't take this condition into account since the first definition seems more natural and less restrictive (Gasser et al., 1985).

The use of higher order kernels seems to be very useful when one is trying to reduce the bias of the estimator without affecting the value of the variance, especially as the bias is directly affected by the variance-bias tradeoff, so when one is reduced the other inflates (Henderson and Parmeter, 2012)(Osemwenkhae and Odiase, 2007). Nevertheless, the use of these kernels presents a main drawback which is the appearance of negative values near the tails where data are sparse (Hall and Murison, 1993).

An obvious way of bypassing this problem is to consider only positive values $\hat{f}_n^+$, this doesn't affect the error of the estimation since the Mean Squared Error (*MSE*) of the new estimator is lesser than the original *MSE*. A longer discussion of this problem and its repercussions can be found in (Hall and Murison, 1993) (Lee and Young, 1994)(Murison, 1993).

In the following we present a variety of higher order kernels based on normalized orthogonal polynomials (Chihaha, 1978), these kernels have been constructed as described below:

Let $\Pi_i(\xi)_{i=1}^{\infty}$ be an infinite orthogonal polynomial basis. This basis is normalized such that $\langle \Pi_i | \Pi_j \rangle = \Delta_{ij}$ where Δ_{ij} is the kronecker's delta. Thus, for any natural number $m \in \{0, 1, 2, ..., l\}$, u^m could be expressed as:

$$u^m = \sum_{d=0}^{m} a_{dm} \Pi_d(u) \tag{13}$$

Where a_{dm} is a real number.

Now, if K is a function $K : \mathbb{R} \longrightarrow \mathbb{R}$ of the form $K(u) = \sum_{i=0}^{l} \Pi_i(0)\,\Pi_i(u)$ then we can write:

$$\int u^m K(u)\,du = \int \sum_{d=0}^{m} \sum_{i=0}^{l} a_{dm} \Pi_d(u)\,\Pi_i(0)\,\Pi_i(u)\,du = \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \in \{1, 2, ..., l\} \end{cases} \tag{14}$$

Hence, the conditions in (12) are satisfied, which implies that the function K can be considered as a kernel of order l.

The result stated in (14) enables us to create kernels of order l using common orthogonal polynomials. Let us note that some polynomials are orthogonal with weights (Gautschi, 2004) and even though this condition is crucial for orthogonality it has very little effect on the proof and it leads us to the modified kernel:

$$K(u) = \sum_{i=1}^{l} \Pi_i(0)\,\Pi_i(u)\,\omega(u) \tag{15}$$

Where $\omega(.)$ is the weight function, which would be different for each kernel.

Next, we are going to consider the Rodrigues formula for various orthogonal polynomials, these formulas when combined with (15) would allow us to write kernels of order l in a very compact form.

Orthogonal polynomials and corresponding kernels:

Here and in the following we consider $J^N(.)$, $L^N(.)$, $G^N(.)$, ...as being the normalized versions of the polynomials $J(.)$, $L(.)$, $G(.)$, ...

Jacobi polynomials (Limits of orthogonality: $[-1,1]$**)**

$$\text{Rodrigues formula: } J(k,\alpha,\beta,u) = \frac{(1-u)^{-\alpha}(1+u)^{-\beta}}{(-1)^k 2^k k!} D_{u^k}\left[(1-u)^{\alpha}(1+u)^{\beta}\left(1-u^2\right)^k\right]$$

$$\text{Weight: } \omega(u,\alpha,\beta) = (1-u)^{\alpha}(1+u)^{\beta}$$

$$\text{Norm: } \eta_k = \sqrt{\frac{2^{\alpha+\beta+1}\Gamma(k+\alpha+1)\Gamma(k+\beta+1)}{k!\,(2k+\alpha+\beta+1)\,\Gamma(k+\alpha+\beta+1)}}\;;\; \alpha,\beta > -1$$

$$\text{Kernel: } K(u) = \sum_{k=0}^{l} J^N(k,\alpha,\beta,0)\,J^N(k,\alpha,\beta,u)\,\omega(u,\alpha,\beta) \tag{16}$$

Legendre polynomials (Limits of orthogonality: $[-1,1]$**)**

$$\text{Rodrigues formula: } L(k,u) = \frac{1}{(-1)^k 2^k k!} D_{u^k}\left[\left(1-u^2\right)^k\right]$$

$$\text{Weight: } \omega(u) = 1$$

$$\text{Norm: } \eta_k = \sqrt{\frac{2}{2k+1}}$$

$$\text{Kernel: } K(u) = \sum_{k=0}^{l} L^N(k,0)\,L^N(k,u) \tag{17}$$

Gegenbauer polynomials (Limits of orthogonality: $[-1,1]$**)**

The polynomials of Gegenbauer are given in terms of Jacobi polynomials with the specification $\alpha = \beta = \lambda - \frac{1}{2}$.

$$\text{Rodrigues formula: } G(k,\lambda,u) = \frac{\Gamma(2\lambda+k)\,\Gamma\left(\lambda+\frac{1}{2}\right)}{\Gamma(2\lambda)\,\Gamma\left(\lambda+k+\frac{1}{2}\right)} J\left(k,\lambda-\frac{1}{2},\lambda-\frac{1}{2},u\right) \; ; \; \lambda > -\frac{1}{2}$$

$$\text{Weight: } \omega(u,\lambda) = \left(1-u^2\right)^{\lambda-\frac{1}{2}}$$

$$\text{Norm: } \eta_k = \sqrt{\frac{\pi 2^{1-2\lambda}\Gamma(2\lambda+k)}{(\lambda+k)\,k!\Gamma(\lambda)^2}} \; ; \; \lambda \neq 0$$

$$\text{Kernel: } K(u) = \sum_{k=0}^{l} G^N(k,\lambda,0)\,G^N(k,\lambda,u)\,\omega(u,\lambda) \tag{18}$$

Chebyshev polynomials of the first kind (Limits of orthogonality: $[-1,1]$**)**

The polynomials of Chebyshev of the first kind are given in terms of Jacobi polynomials with the specification $\alpha = \beta = -\frac{1}{2}$.

$$\text{Rodrigues formula: } T(k,u) = \frac{2^{2k}(k!)^2}{(2k)!} J\left(k,-\frac{1}{2},-\frac{1}{2},u\right)$$

$$\text{Weight: } \omega(u) = \left(1-u^2\right)^{-\frac{1}{2}}$$

$$\text{Norm: } \eta = \begin{cases} \sqrt{\frac{\pi}{2}} & \text{if } k \neq 0 \\ \sqrt{\pi} & \text{if } k = 0 \end{cases}$$

$$\text{Kernel: } K(u) = \sum_{k=0}^{l} T^N(k,0)\,T^N(k,u)\,\omega(u) \tag{19}$$

Chebyshev polynomials of the second kind (Limits of orthogonality: $[-1,1]$**)**

The polynomials of Chebyshev of the second kind are given in terms of Jacobi polynomials with the specification $\alpha = \beta = \frac{1}{2}$.

$$\text{Rodrigues formula: } U(k,u) = \frac{2^{2k}(k!)(k+1)!}{(2k+1)!} J\left(k,\frac{1}{2},\frac{1}{2},u\right)$$

$$\text{Weight: } \omega(u) = \left(1-u^2\right)^{\frac{1}{2}}$$

$$\text{Norm: } \eta = \sqrt{\frac{\pi}{2}}$$

$$\text{Kernel: } K(u) = \sum_{k=0}^{l} U^N(k,0)\, U^N(k,u)\, \omega(u) \tag{20}$$

Hermite polynomials (Limits of orthogonality: $\mathbb{R}$)

$$\text{Rodrigues formula: } H(k,u) = (-1)^k \exp\left(u^2\right) D_{u^k}\left[\exp\left(-u^2\right)\right]$$

$$\text{Weight: } \omega(u) = \exp\left(-u^2\right)$$

$$\text{Norm: } \eta_k = \sqrt{2^k k! \sqrt{\pi}}$$

$$\text{Kernel: } K(u) = \sum_{k=0}^{l} H^N(k,0)\, H^N(k,u)\, \omega(u) \tag{21}$$

Laguerre polynomials (Limits of orthogonality: $\mathbb{R}^+$)

$$\text{Rodrigues formula: } Lag(k,\alpha,u) = \frac{\exp(u)\, u^{-\alpha}}{k!} D_{u^k}\left[\exp(-u)\, u^{k+\alpha}\right] ;\ \alpha > -1$$

$$\text{Weight: } \omega(u,\alpha) = u^{\alpha} \exp(-u)$$

$$\text{Norm: } \eta_k = \sqrt{\frac{\Gamma(k+\alpha+1)}{k!}}$$

$$\text{Kernel: } K(u) = \sum_{k=0}^{l} Lag^N(k,\alpha,0)\, Lag^N(k,\alpha,u)\, \omega(u,\alpha) \tag{22}$$

The figures below [figures 1, 2] illustrate the behavior of two of the suggested kernels over the interval $]-1,1[$. One of these two kernels will be used for the numerical experiments in section 4.

Now, even though we have the proper kernels we are not ready yet to carry out a practical implementation of the kernel density estimation because we are still lacking a correct way to compute the value of a vital parameter which is the bandwidth (i.e. the smoothing parameter). In the next section we will try to provide a tractable formula for computing this parameter.

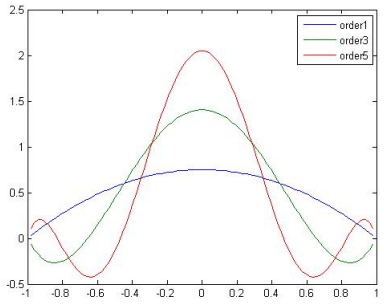

Figure 1: Jacobi kernel of orders 1, 3 and 5.

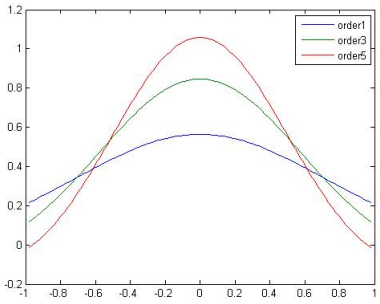

Figure 2: Hermite kernel of orders 1, 3 and 5.

3 Computing The Bandwidth

In the process of estimating a density using a kernel approach, selecting an adequate bandwidth is of the essence, for the most part because this parameter is directly implicated in the computing of both, the density estimation (2) and its error (7). An erroneous choice might yield untrustworthy data, since insufficient smoothing can lead to an estimation that contains spurious features resulting from the sampling process, in the same way that excessive smoothing can cause important features, of the underlying structure, to be smoothed away.

In this work, we consider data-driven bandwidth selectors, which mean that we'll be relying on data to choose a correct smoothing parameter; these selectors have been categorized into two generations, the first and the second. We are mainly interested in introducing a variation of a first generation method namely the "rules of thumb", for an exhaustive treatment of all first generation methods one can consult (Jones et al., 1996).

The classical approach that we are interested in modifying involves h_{AMISE} the minimizer of the *AMISE* (7):

$$h_{AMISE} = \left[\frac{R(K)}{n\mu_2(K)^2 R(f'')}\right]^{\frac{1}{5}} \tag{23}$$

We can easily figure out that this formula cannot be computed due to the unknown part $R(f'')$ that invokes the second derivative of the true density. The "rules of thumb" method requires the replacement of the unknown part by an estimated value based on a parametric family; one recurrent example is the normal scale bandwidth which is the *AMISE*-optimal bandwidth for the normal density.

The variation we suggest is to use the second derivative of a common kernel such as the Gaussian in order to estimate the unknown density and its derivatives (Bhattacharya, 1967). Given that we have a kernel that is differentiable r times, then the r^{th} derivative of a density estimator based on such a kernel may take the form:

$$\hat{f}_n^{(r)}(x) = \frac{1}{nh^{r+1}} \sum_{i=1}^{n} K^{(r)}\left(\frac{x - X_i}{h}\right) \tag{24}$$

As stated above we'll be using the Gaussian kernel in order to estimate the r^{th} density derivative which, in this case, would be written as:

$$K^{(r)}(u) = (-1)^r H(r, u) K(u) \tag{25}$$

Where $H(r, u)$ is the Hermite polynomial of order r used for the construction of the kernel (21), from here we can explicitly write our density derivative estimator:

$$\hat{f}_n^{(r)}(x) = \frac{(-1)^r}{nh^{r+1}\sqrt{2\pi}} \sum_{i=1}^{n} H\left(r, \frac{x - X_i}{h}\right) \exp\left[-\left(\frac{x - X_i}{h\sqrt{2}}\right)^2\right] \tag{26}$$

For computational purposes we'll be using the normal scale smoother for the density derivative estimator, however in (Wand and Jones, 1985) a similar analysis that led to (23) led to another formula for computing the density derivative bandwidth:

$$h^r_{AMISE} = \left[\frac{R(K)(2r+1)}{n\mu_2(K)^2 R\left(f^{(r+2)}\right)}\right]^{\frac{1}{2r+5}} \tag{27}$$

One computational drawback of this approach is the need to estimate another density derivative of order $r + 2$, either based on a parametric family or by using a kernel approach. Therefore, we have chosen to use a normal scale smoother and not to go deeper than the first derivative estimation.

Now, we aim at finding a proper estimator for the integrated squared density derivative functional noted by $R\left(f^{(r)}\right) = \int \left(f^{(r)}(x)\right)^2 dx$. Integration by parts formula gives us:

$$R\left(f^{(r)}\right) = (-1)^r \int f^{(2r)}(x) f(x)\, dx \tag{28}$$

Next, we are interested in estimating this quantity for an even r, this steers us toward a density functional of the form:

$$\rho_r = \int f^{(r)}(x) f(x)\, dx = E\left(f^{(r)}(x)\right) \tag{29}$$

Where ρ_r can be estimated by:

$$\hat{\rho}_r = \frac{1}{n} \sum_{i=1}^{n} \hat{f}^{(r)}(x_i) \tag{30}$$

Using (24) we can write:

$$\hat{\rho}_r = \frac{1}{n^2 h^{r+1}} \sum_{i=1}^{n} \sum_{j=1}^{n} K^{(r)}\left(\frac{x_i - x_j}{h}\right) \tag{31}$$

Now, by combining equations (23), (26) and (31) we can have the following estimator of the *AMISE*-optimal bandwidth, the very estimator we will be using on our numerical experiments in the next section:

$$\hat{h}_{AMISE} = \left[\frac{R(K)}{n\mu_2(K)^2\,\hat{\rho}_4}\right] \tag{32}$$

4 Numerical Experiments

In this section we will be presenting some examples that illustrate, in the context of stock market data, the use of kernel density smoothing through some of the kernels suggested above.

Considering the representativeness of the data we have chosen the *MASI*(Moroccan All Shares Index), which is a good tracking device that helps assessing the health and performance of the Moroccan stock market. The data sample covers the index values from November the 7^{th} 2014 to November the 4^{th} 2015.

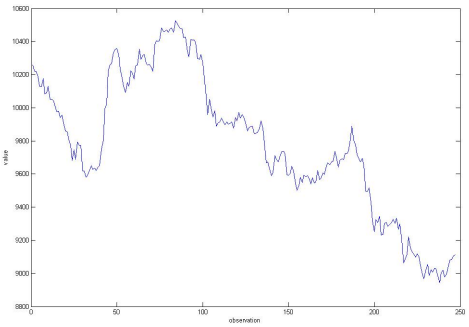

Figure 3: Nonstationary MASI variations.

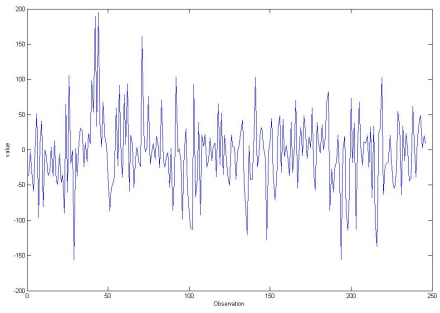

Figure 4: Stationary MASI variations.

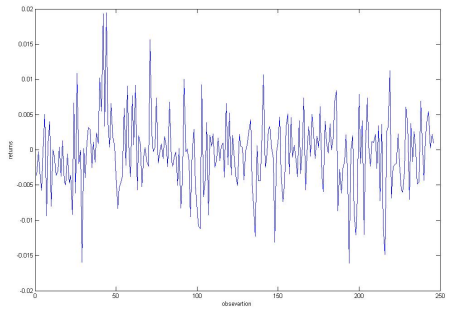

Figure 5: Stationary MASI returns.

First, we have stationarized the index values, as to the returns they have been tested positive for stationarity without any manipulations [figures 3, 4, 5], the tests have been conducted using the autocorrelation function and the augmented dickey-fuller unit root test.

It's worth noting that beside the difference of the scale, we can firmly state that variations of the stationarized index values and the returns are a perfect match, thus we will only be estimating the density of returns since the non-stationarity of the observed values stresses the fact they are not independent and identically distributed which calls into question the relevance and the efficiency of the kernel density estimation as we present it.

The table below [Table1] gives a basic statistical description of the dataset, a quick analysis of this table provides a preliminary insight concerning the distribution of data and the effect of non-stationarity on the values. We can rely on the skewness and the kurtosis measures in order to have a general idea concerning the shape and the behavior of the density, we can notice that the index values are left-skewed adversely the returns are right-skewed, both of the kurtosis measures are different from 3 leading to the index values being platykuric with thin tails and the returns being leptokurtic with fat tails. In view of these observations we can assert that the distribution of both the values and the returns are far from being Gaussian this statement has been confirmed by the Jarque-Bera test that has returned two p-values lesser than 5%, which was the critical value adopted, leading to the rejection of the normality of the distribution.

Table 1: Basic statistical description of MASI values and returns.

	values	returns
Mean	9783.3	$-4.6691e-4$
Median	9773.6	$-1.6314e-4$
Maximum	10527	0.0195
Minimum	8944.1	-0.0161
Std. dev	421.1494	0.0054
Skewness	-0.1835	0.1745
Kurtosis	2.2126	4.4512
J-B test (P-values)	0.0259	0.0015

Now, let's try and estimate the empirical density for the returns. Seeing that the standard Gaussian kernel is one of the most reliable common kernels (Wand and Jones, 1985) we have chosen to use it as a reference to evaluate the performance of the suggested kernels. We will be using the Lognormal, Pareto, Laguerre and Hermite kernels to conduct the density estimations and after each one of these estimations we will run a nonparametric goodness-of-fit test (Albers and Schaafsma, 2008)(Li and Racine, 2007),

namely the Chi-Square test, in order to assess the quality of the fitting and compare it to the performance of the Gaussian kernel. This test can lead us toward a way of improving the estimation either by varying the smoothing parameter or even the kernel when none of the used bandwidths gives satisfying results.

Index returns density estimation:

As stated earlier, in addition to the standard Gaussian kernel, four suggested kernels will be used for the estimation of the index returns density. The results, superimposed on the histograms, are given in the figures below:

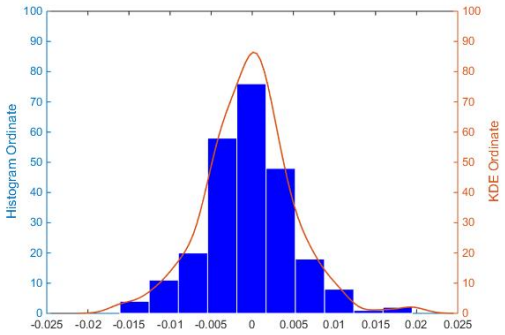

Figure 6: Density estimate for MASI returns using standard Gaussian kernel.

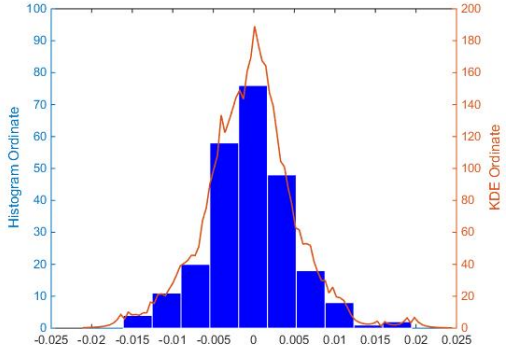

Figure 7: Density estimate for MASI returns using Lognormal kernel.

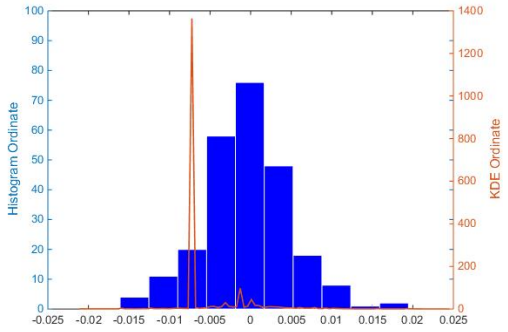

Figure 8: Density estimate for MASI returns using Pareto kernel.

Now, except for [Figure 8] that illustrates the use of the Pareto kernel and looks too rough to reflect a convincing density estimate, all the other figures have exposed some features, that were smoothed away by the Gaussian kernel, whilst keeping a fairly satisfying shape of the density when compared to the conclusions drawn from [Table 1].

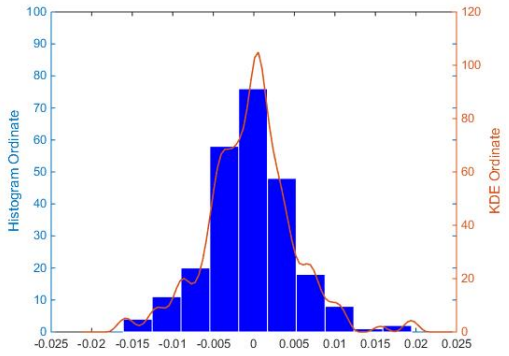

Figure 9: Density estimate for MASI returns using Hermite kernel of order 2.

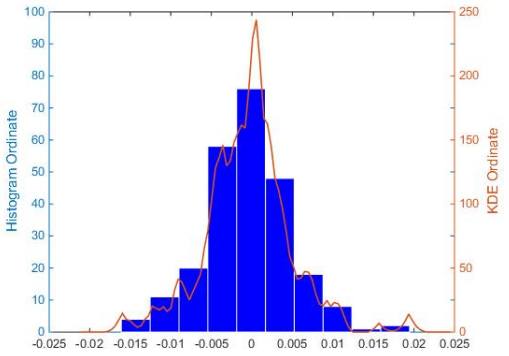

Figure 10: Density estimate for MASI returns using Laguerre kernel of order 2.

Nevertheless, in order to give a more discerning and concise judgment of which kernels fits best the data we conduct a Chi-Square goodness of fit test that tests two hypothesis:

$$H_0 : f = \hat{f}_n \text{ Vs } H_1 : f \neq \hat{f}_n$$

The table below [Table2] lists the Chi-Square statistic computed for each of the fittings with three degrees of freedom ($df = 3$) and a five percent level of significance ($\alpha = 0.05$), let us note that for the given parameters the critical values for the test is 7.81.

Table 2: Chi-Square test statistic for suggested kernels in the case of index returns density estimation.

	Gaussian	Lognormal	Pareto	Hermite	Laguerre
χ^2-Statistic	0.55	0.21	351.9844	0.1692	0.1272

The only statistic leading to the rejection of H_0 is unsurprisingly the one relative to the Pareto kernel, meaning that all the other fittings are acceptable, at different levels of course. One remarkable thing is that all the acceptable fittings have generated a better Chi-Square statistic than the Gaussian kernel meaning that the suggested kernels give a better fitting for the data at hand.

5 Conclusion

This paper was aiming at suggesting and testing some unconventional kernels based on probability densities and orthogonal kernels. Through these kernels we have been

able to generate statistically correct density estimates for the MASI returns, the density estimates of the Gaussian kernel have been outperformed in the sense of the Chi-Square statistic by three of the tested kernels. Still, at this point the use of the results was limited to a graphic illustration; further work needs to be done in order to exploit the data and the kernels in more interesting ways in a financial context such as kernel regression for forecasting data or Monte-Carlo simulations for pricing financial products.

Acknowledgement

The authors are very grateful to the anonymous reviewers for their helpful comments and remarks that were greatly beneficial in improving the material presented in this paper.

References

Albers, C. J. and Schaafsma, W. (2008). Goodness of fit testing using specific density estimate. *Statistics and Decisions*, 26:3–23.

Bhattacharya, P. K. (1967). Estimation of a probability density function and its derivatives. *Sankhya: Series A*, 29:373–382.

Chihaha, T. S. (1978). *An introduction to orthogonal polynomials*. Gordon and Breach.

Di Nunno, G. and Øksendal, B. (2011). *Advanced mathematical methods for finance*. Springer-Verlag.

Gasser, T., Müller, H. G., and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of Royal Statistical Society: Series B*, 47:238–252.

Gautschi, W. (2004). *Orthogonal polynomials: computation and approximation*. Oxford University Press.

Grund, B., Hall, P., and Marron, J. (1994). Loss and risk in smoothing parameter selection. *Journal of Nonparametric Statistics*, 4:107–132.

Hall, P. and Murison, R. D. (1993). Correcting the negativity of high-order kernel density estimators. *Journal of Multivariate Analysis*, 47:103–122.

Henderson, D. J. and Parmeter, C. F. (2012). Canonical higher order kernel for density derivative estimation. *Statistics and Probability Letters*, 82:1383–1387.

Jones, M., Marron, J. S., and Sheater, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of The American Statistical Association*, 91:401–407.

Lee, S. and Young, G. A. (1994). Practical higher-order smoothing of the bootstrap. *Statistica Sinica*, 4:445–459.

Li, Q. and Racine, J. S., editors (2007). *nonparametric econometrics: theory and practice*. Princeton University Press.

Marron, J. and Wand, M. (1992). Exact mean integrated squared error. *The Annals of Statistics*, 20:712–736.

Murison, R. D. (1993). *Problems in curve estimation for independent and dependent data*. PhD thesis.

Newman, M. E. J. (2006). Power laws, pareto distributions and zipf's law. Unpublished paper.

Osemwenkhae, J. E. and Odiase, J. I. (2007). Improving the choice of higher order univariate kernels through bias reduction technique. *Journal of Science and Technology*, 26:19–23.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall.

Wand, M. P. and Jones, M. C. (1985). *Kernel smoothing*. Chapman and Hall.