Modeling Tax Revenues Using Kernel Approach Case Study: North Kivu Province (Democratic Republic of Congo) Tax Revenues Time Series Forecasting

Byamungu Wanguwabo David

Jomo Kenyatta University of Agriculture and Technology/ Master Program
Lecturer at Université Libre des Pays de Grands-Lacs / Goma

Abstract: ***Forecasting the distribution of tax revenues in the Democratic Republic of Congo has been an uphill task.The recent past of the country has been dominated by economic uncertainty, particularly in mining and in agricultural products (cash crops) meant for export. This factor alone has greatly contributed to high volatility of tax revenues collected by the custom officers. The fuzzy characteristic of Tax Revenues has made it quite impossible for researcher to detect or distinguish from randomness the three well known components of a classical time series, precisely Trends, Seasonality, and cyclical phenomena. Hence parametric methods, however rich appear not to be suitable at all to produce reliable forecasts. The current project focuses on modeling tax revenues time series using nonparametric method, mainly the kernel approach. Several kernels have been discussed in literature. In this project, due to its optimal property, the Epanechnikov kernel is used as an index kernel to model the time series under investigation. Other commonly kernels, including the Parzen, Gaussian, Biweight, cosine, rectangle, triangle and the alternative Epanechnikov (epan2) kernels have been used to fit the dataset and their performance compared with the index kernel. By default, the Gaussian and the alternative Epanechnikov kernels performed very close to the index kernel. Having chosen to use, for comparison purposes, the Epanechnikov, the Gaussian and the alternative Epanechnikov kernels, an optimal choice of the bandwidth has been discussed through the kernel weighted polynomial smoothing setup. Two crucial aspects of the problem were evaluated, including the degree of the polynomial that precisely fit the data points and the level of the bandwidth that is required to achieve bell-fit. To this end, the performance of the Epanechnikov, the Gaussian and the alternative Epanechnikov (epan2) kernel using kernel weighted polynomial of degrees 1, 3 and 7 for different values of the bandwidth, precisely for h = {1,5,7,10} has been examined. As expected, findings suggest unequivocally that the higher the degree of the kernel weighted local polynomial smoothing combined with the smallest value of the bandwidth, the better is the fit of the kernel used to the tax revenues data. Hence, to predict or forecast tax revenues, either the Epanechnikov, the Gaussian or the alternative Epanechnikov (epan2) kernel can be used, with a careful choice of the pair (p,h) where p is the degree of the polynomial which is assumed to be reasonably high and h is the optimal bandwidth.***

1. Introduction

In statistics, precisely in inferential statistics, density estimation plays a crucial role, if not the main role; since it builds an estimate of some underlying probability function using sample observations. This density estimate is often used to forecast future values of the observed data in general. Moreover, this density estimate can either be parametric from known distribution or nonparametric.

The field of parametric forecast of a time series is well documented in several time series textbooks. Interested reader can visit (Scott, 1992). Gramack(2017), Tsay (2014) and Fan and Yao (2003). A robust list of forecasting methods for stationary and non-stationary time series with application in finances, economics (housing expenditure), insurance, etc, are broadly discussed and both univariate and multivariate time series covered.

But most researchers or Time series users in finance, insurance or economics or in any field where time series are used are open to rather using forecasting methods that enable them to access a robust information characterizing the time series under study. When a time series displays exclusively random walk, as it is the case with the DRC tax revenue, usual methods fail to provide the best forecast. This factor has motivated the researcher enormously, given that data he will be processing belongs to a region with higher volatile economy.

The focus of this project is rather on nonparametric density function that can be used to forecast the Democratic Republic of Congo (DRC) tax revenues in time. Precisely, the researcher will use kernel density estimator to forecast future values of tax revenues collected in Kivu province, DRC. The main references the researcher will consider on the concept of Kernel estimation include (Bowman and Azzalini, 1997) and (Silverman, 1989), among many others.

To introduce the concept, Elliot and Timmermann (2016) emphasizes that density forecast is about forecasting the likelihood of different outcomes which must avail the overall information on the uncertainty that goes with any forecast. Due to their generality, these density estimates are often employed by various users who might utilize different loss functions to produce any point forecast of their choice (Yuyan, 2017). Common methods that are widely used include the histograms, kernel methods and penalized approaches. If interested in an extensive coverage of these methods, one can visit (Scott, 1992), (Gramack, 2017) and (Silverman, 1989).

Gramacki (2018) offers an extensive in-depth on kernels and their utilization in various fields, including both in univariate and multivariate statistical data analysis. Three areas of applications have attracted the attention from the scientific community, namely in Kernel discriminant analysis, kernel cluster analysis and in kernel regression. J. (2004) provides a short and fast-paced description of the concepts underlying discriminant analysis. Then kernel discriminant analysis is simply a method used for solving a general classification task and is based on analyzing probability density functions in classes k, say. The cluster analysis technique however, belongs to the group of methods known as unsupervised classification. It focuses in the mean-shift algorithm which is itself an example of a density-based clustering as discussed in (Ester et al., 1996), the like of the well-known k-means method (Hartigan and Wong, 1979). Kernel regression analysis is simply anvariant of the usual regression analysis in nonparametric settings. Early works in kernel regressionanalysis include (Nadaraya, 1964) and (Watson, 1964). Ever since the era of early paper, kernel density estimation has been applied in many fields, particularly in applied statistics and in time series.

Yuyan (2017) covered an asymptotic theory on kernel estimation for time series. He established an asymptotic normality and uniform rates of convergence of kernel estimators under mild regularity conditions.

In the field of applied statistics, Trosset (2008) discussed the kernel density estimation and its implementation in the software R. In a case study, the forearm lengths were tested whether they were Normally Distributed or not using kernel density estimation methods. In the area of time series, there is a huge library containing the use of the kernel density estimation in many aspects. Joanqing and Qiwei (2003) offers a complete survey on the concepts. In one of the early papers, Robinson (1984) discussed kernel estimation alongside interpolation method for time series containing missing observations. All along, one can see that kernel estimation method is simply one of the many smoothing methods that are used in various fields. Mohamed and Claude (2009) studied the efficiency of the Paris Stock Exchange market using the kernel methods. Ana (2012) proposes a new method for estimating the whole changing distribution of a time series using nonparametric kernel estimation and exponentially weighted moving average filters (EWMA).

Andrew and Vitaliy (2012) estimated a time-varying probability density function or the corresponding cumulative distribution function using a kernel and did weight the observations using schemes derived from time series modeling. Krisp and Stefan (2011) investigated the density calculation and representation of spatially and temporally highly dynamic point data sets. In this paper, he suggested an approach to explore point patterns that have a temporal dimension and therefore introduced an incremental development of the traditional kernel density estimation processes. Jeffrey (1991) discussed the kernel regression estimation with time series errors. In his paper, he addressed the problem of objectively choosing the band width of a kernel estimate for a function f. In this paper, he showed that both theoretically and by simulation, cross-validation produces extremely rough kernel estimates when the data are sufficiently positively correlated. Derryberry (2014), in the context of time series smoothing method, discussed about kernel smoothers in as a method that can be used in the software R to produce smoothed periodigrams that pictures clearly the bias-variance trade-off. Jeffrey (1991) discussed the kernel regression estimation with time series errors. He showed that when one incorporates the estimated covariances into a risk estimation procedure, results in more efficient smoothing of positively correlated data.

In all these literatures, it can be observed that no authors applied the kernel density estimation on taxes for a volatile economy. Therefore, the focus of this paper is to model Tax revenues of DRC using kernel density estimation and use this to predict taxes from the main sources. This paper is organized as follows. Section two is devoted to the review of the kernel density estimation methods. Section 3 presents the empirical application of Kernel density on tax revenue for the DRC and section 4 is the conclusion and recommendation.

2. Kernel density estimation method

In this section, kernel density function is presented alongside its properties, and other kernel characteristics with respect to the optimal bandwidth selection and kernel smoothing procedure.

2.1 Kernel density function

The starting point is as follows: Given T data points $X_1, \cdots, X_T$, how do we estimate their empirical distribution function? If one considers the mass 1/T at each observed data point, then it follows that the required empirical distribution function is of the form

$$\hat{F}(x) = \frac{1}{T}\sum_{t=1}^{T} I(X_t \leq x) \qquad (2.1)$$

which is a nondecreasing function. However, this is not helpful since it cannot be used to examine the overall distribution structure of the data at hand. Moreover, the density of (2.1) does not exist (Jianqing and Qiwei, 2003, pg 195). To improve (2.1), one introduces the kernel function K which is used to smoothly redistribute the mass 1/T at each data point.

Referring to David W. Scott, (), the basic kernel estimator is of the form

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) \quad (2.2)$$

Where $k_h(t) = K(t/h)/h$, a notation introduced by Rosenblatt (1956). Kernels that are used commonly include the Gaussian kernel $K(u) = \left(\sqrt{2\pi}\right)^{-1} exp(-u^2/2)$ and the symmetric Beta family $K_\gamma(u) = \frac{1}{Beta\left(\frac{1}{2},\gamma+1\right)}(1 -$ $u2\gamma Iu\leq 1$ where the choice of $\gamma=0, 1, 2, 3$ corresponds respectively to the uniform, the Epanechnikov, the biweight and the triweight kernel functions. It is well known that the choice of the kernel is not really a problem, but the choice of the bandwidth, h. When this is optimally chosen, any kernel selected for use will perform as good as any other kernel left out.

2.2 Properties of Kernel function

In statistical analysis of kernel estimators, one focuses on the measure of efficiency, that is, MISE analysis and the selection of the bandwidth.

(i) Mean Integrated Squared Error (MISE) analysis

A close examination of equation (2.2) reveals that

$$K_h(x, X_i) = \frac{1}{h} K\left(\frac{x-X_i}{h}\right) \quad (2.3)$$

is simply the arithmetic mean of the T i.i.d. random variables. Thus, it follows that

$$E\left[\hat{F}(x)\right] = E[K_h(x, X)] \quad (2.4)$$

and that

$$var\left(\hat{F}(x)\right) = \frac{1}{n} var(K_h(x, X)) \quad (2.5)$$

In terms of computations of (2.3) and (2.4), one computes the right hand side as

$$E[K_h(x,X)] = \int \frac{1}{h} K\left(\frac{x-t}{h}\right) f(t)\, dt$$
$$= \int K(\omega) f(x - h\omega)\, d\omega; \quad \left(\omega = \frac{x-t}{h}\right)$$
$$= f(x) \int K(\omega)\, d\omega - hf'(x) \int \omega\, K(\omega)\, d\omega + \frac{1}{2} h^2 f''(x)$$
$$\int \omega^2 K(\omega)\, d\omega + \ldots \approx f(x) \int K(\omega)\, d\omega + \quad (2.6)$$
$$var(K_h(x, X)) = E[K_h(x,X)]^2 - (E[K_h(x,X)])^2$$
$$= E\left[\frac{1}{h} K\left(\frac{x-X_i}{h}\right)\right]^2 - \left(E\left[\frac{1}{h} K\left(\frac{x-X_i}{h}\right)\right]\right)^2 \quad (2.7)$$

For the first term in (2.9), one computes

$$\int \frac{1}{h^2} K\left(\frac{x-t}{h}\right)^2 f(t)\, dt = \int \frac{1}{h} K(\omega)^2 f(x-h\omega)\, d\omega \approx \frac{f(x)R(K)}{h} \quad (2.8)$$

Having set

$$\omega = \frac{x-t}{h} \Rightarrow d\omega = -\frac{1}{h} dt; t = x - h\omega \quad (2.9)$$

Since

$$\int \omega^2 K(\omega)\, d\omega \approx \sigma^2{}_K > 0 \text{ and}$$
$$\int \omega K(\omega)\, d\omega = 0 \quad (2.10)$$

it follows from (2.9) that the expectation of $E\left[\hat{F}(x)\right]$ simply equals f (x) onto the order $O(h^2)$.

Clearly, the induced bias measure is given by

$$Bias = \frac{1}{2}\sigma_K^2 f''(x) + O(h^4) \Rightarrow ISB = \frac{1}{4}\sigma_K^2 h^4 R(f''(x)) + O(h^6) \quad (2.11)$$

where ISB stands for 'Integrated Squared Bias'.

In similar manner, a close examination of (2.10), having (2.7) and (2.8), one deduces that

$$var(x) = \frac{f(x)R(K)}{nh} - \frac{f(x)^2}{h} + O\left(\frac{h}{n}\right) \Rightarrow$$
$$IV = \frac{R(K)}{nh} - \frac{R(K)}{h} + \cdots \quad (2.12)$$

where IV stands for 'Integrated Variance'. To summarize these results, one has the following

Theorem 3.1 Let (2.3) define a nonnegative univariate kernel density estimator. Then, it holds that

$$AMISE = \frac{R(K)}{nh} - \frac{1}{4}\sigma_K^4 h^4 \frac{R(K)}{n} \quad (2.13)$$

and for optimum smoothness

$$AMISE^* = \frac{4}{5}\left(\frac{\sigma_k R(K)}{h}\right)^{4/5} R(f'')^{1/5} \quad (2.14)$$

for the optimal bandwidth

$$h^* = \left(\frac{1}{n}\frac{R(K)}{\sigma_K^4 R(f'')}\right)^{1/5} \quad (2.15)$$

Among many other scholars, Parzen (1962) covered a set of condition under which the above theory holds.

2.3 Optimal selection of the bandwidth

Since the choice of the bandwidth **h** is crucial for a robust MISE, the lines below present briefly how optimal selection of h is carried out. Let X_t denote a realization from a stationary time series. Thus, the mean square error (MSE) of the kernel density estimator, by definition, is given by

$$MSE(x) = E\left(\hat{f}_h(x) - f(x)\right)^2$$
$$\approx \frac{1}{4}\left(\int_{-\infty}^{\infty} u^2 K(u)\, du\right)^2 \left(f''(x)\right)^2 h^4 + \quad (2.16)$$
$$\int_{-\infty}^{\infty} K^2(u)\, du\, \frac{f(x)}{Th}$$

where x is a member of the interior support off. A global measure referred to as the MISE, is given by

$$MISE(x) = E\left(\hat{f}_h(x) - f(x)\right)^2$$
$$\approx \frac{1}{4}\left(\int_{-\infty}^{\infty} u^2 K(u)\, du\right)^2 \int_{-\infty}^{\infty} h^4 \left(f''(x)\right)^2 dx + \quad (2.17)$$
$$\int_{-\infty}^{\infty} K^2(u)\, du\, \frac{f(x)}{Th}$$

From (Fan and Yao, 2003), one learn that minimizing (2.19) with respect to h results in the optimal bandwidth given by

$$h_{opt} = R(K)\, \|f''\|_2^{-2/5} T^{-1/5} \quad (2.18)$$

Where $\|g\|_2^2 = \int_{-\infty}^{\infty} g(u)du$ is the L2-norm, $\mu_2(K) = \int_{-\infty}^{\infty} u^2\ K(u)^2\ du$ is the variance of K and

$$R(K) = \frac{\|K\|_2^{2/5}}{\mu_2(K)^{2/5}} \quad (2.19)$$

is a known constant. Given (2.18), the optimal MISE will be given by

$$\frac{5}{4}\beta(K)\, \|f''\|_2^{2/5} T^{-4/5} \quad (2.20)$$

where $\beta(K) = \mu_2(K)^{2/5} \|K\|_2^{8/5}$.

Given two kernel functions, K_1 and K_2 each having its optimal bandwidth, $h_{opt}(K_1)$ and $h_{opt}(K_2)$ respectively. Then, the optimal bandwidth (2.18) satisfies the relation

$$h_{opt}(K_1) = \frac{R(K_1)}{R(K_2)} h_{opt}(K_{12}) \quad (2.21)$$

In terms of implementation of (2.18), observe that it depends of the knowledge of the parameter $\|f''\|_2$ which obviously is not known. When f is a Gaussian density with standard deviation, σ, from (2.18), one deduces

$$h_{opt,T}(K) = \left(\frac{8\sqrt{\pi}}{3}\right)^{1/5} R(K)\sigma\, T^{-1/5} \quad (2.22)$$

When the standard deviation is not available, this is replaced by the sample standard deviation, s. Numerically, from (Fan and Yao, 2003), optimal bandwidth selector has the form

$$h_{opt,n}(K) = 1.06\, s\, T^{-1/5} \quad (2.23)$$

for the Gaussian kernel and

$$h_{opt,n}(K) = 2.34\, s\, T^{-1/5} \quad (2.24)$$

for the Epanechnikov kernel.

(i) Improvement of the optimal bandwidth

Equations or the bandwidth selections (2.25) and (2.26) are very useful, particularly when the observed data are nearly normal. Otherwise, they may lead to oversmoothing particularly when the observed data are asymmetric or multimodal. An improvement suggested by Hjort and Jones (1996) is to consider the optimal bandwidth, h of the form

$$\hat{h}^*_{opt,T} = \hat{h}_{opt,T}\left(1+\frac{35}{48}\hat{\gamma}_4+\frac{35}{32}\hat{\gamma}_3+\frac{385}{1024}\hat{\gamma}_4^{\ 2}\right)^{-1/5} \quad (2.27)$$

Where

$$\hat{\gamma}_3 = (T-1)^{-1}\sum_{t=1}^{T}\left(\frac{X_t-\bar{x}}{s}\right)^3 \ and$$

$$\hat{\gamma}_4 = (T-1)^{-1}\sum_{t=1}^{T}\left(\frac{X_t-\bar{x}}{s}\right)^4 - 3 \quad (2.28)$$

are respectively the skewness and kurtosis of the observed data.

The extension of Kernel density estimation to a multivariate case is discussed in many textbooks. Interested reader to visit, among many other textbook ArturGramacki (2018).

2.4 Kernel smoothing method

An equivalent but improved version of the moving average scheme in time series smoothing methods is what is referred to as the kernel regression estimator. Formally, this is defined by

$$\hat{f}_{t0} = \frac{\sum_{t=1}^{T} Y_t K\left(\frac{t-t_0}{h}\right)}{\sum_{t=0}^{T} K\left(\frac{t-t_0}{h}\right)} \quad (2.27)$$

Which is also referred to as the "Nadaraya-Watson estimator" (see Nadaraya (1964) and Watson (1964)). Equation (2.29) reduces to a (2h +1) data points moving average when the uniform kernel, K(u) = 0.5I(|u| ≤ 1). Jianqing and Qiwei (2003) have shown that the bias of the kernel (2.29) and its variance are respectively

$$E\hat{f}_{t0} - f_{t0} = \frac{\sum_{t=1}^{T}(f_{t0}-f_{t0})K\left(\frac{t-t_0}{h}\right)}{\sum_{t=0}^{T} K\left(\frac{t-t_0}{h}\right)} \quad (2.28)$$

$$Var\left(\hat{f}_{t0}\right) = \sum_{t=1}^{T}\sum_{j=1}^{T}\gamma x(|i-j|)\omega_i\omega_j \quad (2.29)$$

Where $\omega_t = \frac{K\left(\frac{t-t_0}{h}\right)}{\sum_{t=0}^{T} K\left(\frac{t-t0}{h}\right)}$ is the weight and γx(t) is the autocovariance function of the process X(t).

In the section that follows, the implementation of kernel density estimation on RDC tax revenue time series is examined.

3. Research Methodology

In this section, the researcher presents the methodology he is using to fit kernel density model and how he measures the efficiency of the kernels.

3.1 Data and Sampling method.

To implement this project, secondary data on tax revenues were be used. Data set have been retrieved from the Ministry of Finance, Department of Contributions, Kivu province. The set of data to be modeled covers six years and nine months, precisely from January 2010 to September 2016. This makes a total of 81 months or 2463 days. Hence the size of the sample under scrutiny is of 2463 data points. In the process, data collected were cleaned using Excel suite.

3.2 Target population

The population targeted in this project was tax revenues received by custom departments from products that, by certain measure, contributed the most on DRC annual financial budget. These products includecoltan and scoriestanique mineral, cassiterite mineral, gold mineral, coffee Arabica and machine spare parts. The list above reveals that both mineral and agricultural products exported by DRC provide, each year, a reliable source of income that the country usesto fund its developmental projects across the regions.

3.3 Statistical model

Let X_{ij} ,i = Jan, Feb,…., Dec, and j = 2010, 2011,…, 2016 denote monthly tax revenue data recorded by the custom officers, ministry of finance, department of contribution, Kivu province. For this project, the researcher intends to use a kernel function that has been proved to be optimal, among many others. In literature, it has been already shown that the optimal kernel that minimizes the Mean square Error (MSE) over a class of kernel functions is precisely the Epanechnikov kernel. Formally, this is given by

$$K(x) = \frac{3}{4}(1-x^2)I_{|x|<1} \quad (3.1)$$

This kernel will be used alongside the optimal bandwidth given as

$\hat{h}_{opt,n} = \frac{2.34\ s}{T^{1/5}}$associated with the Epanechnikov kernel.

For comparison of performance purposes, the following kernel will also be used.

Normal	$(2\pi)^{-1/2}exp\left(-\frac{1}{2}x^2\right)$	(3.2)
Triangle	$\frac{1}{2}I_{\|x\|<1}$	(3.3)
Epanechnikov	$\frac{3}{4}(1-x^2)I_{\|x\|<1}$	(3.4)
Biweight	$\frac{15}{416}(1-x^2)I_{\|x\|<1}$	(3.5)
Triweight	$\frac{35}{36}(1-x^2)I_{\|x\|<1}$	(3.6)
Triangular	$(1-\|x\|)I_{\|x\|<1}$	(3.7)

are used and comparison with 3.1 will be made in terms of performance.

4. Empirical application of the Kernel density on tax revenues for RCD

4.1 Data Visualization

In this section, it is of interest to represent graphically the time series under study and to motivate the reason why kernel density estimation method has been preferred to be used in this study. The figure below is the graphic of the time series under investigation. Clearly, one observes

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

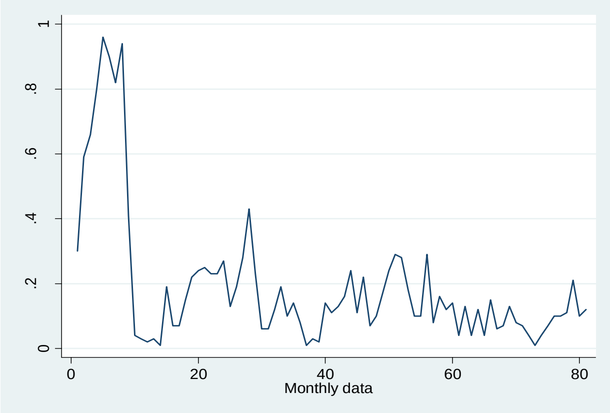

Figure 4.1: Displays the time series under study.

That some of the classic components of a time series are hardly spotted. Trend, seasonal, cycles components are absent, but one can only conclude that the random components are definitely present.

To enforce this argument, one examines the behavior of the sample spectral distribution to test whether figure 4.1 may contain cycles which are not detected. Theoretically, the spectral density of a time series is referred to as a plot that depicts the sinusoidal amplitudes versus the frequencies for the sinusoidal decomposition in the time series (StataCorp. 2009). When the sinusoidal amplitudes are computed for a discrete set of natural frequencies, say (1/n, 2/n, …, q/n), one obtains what is referred to as a "periodigram" for the time series. Formally, let $\{x_t\}_{t=1}^{n}$ denote the time series under study. Also let $\omega_j = (j-1)/n$ denote the natural frequencies for the values of j = 1,…,$\left\lfloor \frac{j}{n} \right\rfloor + 1$ where the $\lfloor . \rfloor$ denote the greatest integer function operator. Define also the expression

$$C_j^2 = \frac{1}{n^2}\left|\sum_{t=1}^{n} x_t e^{2\pi i(t-1)\omega_j}\right|^2 \qquad (4.1)$$

A periodigram of the time series described above is obtained by plotting nC_j^2versus ω_j . To define the sample spectral density, one considers$\hat f\,(\omega_j) = nC_j^2$. Let then $\hat f(\omega_1),\cdots,\hat f(\omega_Q)$ to be the sample spectral density function of the time series under investigation being evaluated at ω_j, for *j=1,…, Q*. If one fixes the value of *q*as $q = \lfloor Q/2 \rfloor + 1$, then, it follows that the resulting probability distribution function representing the sample spectral density is of the form

$$\hat F(\omega_j) = \sum_{i=1}^{j}\hat f\,(\omega_k) \Big/ \sum_{i=1}^{q}\hat f\,(\omega_k) \qquad (4.2)$$

This defines what is known as the sample spectral-distribution function of the time series under study. Given the time series under study, one fixes a vertical line at frequency 1/12, since the data are monthly. At this point, one expects to observe a jump signifying existence of an annual cycle in the data.

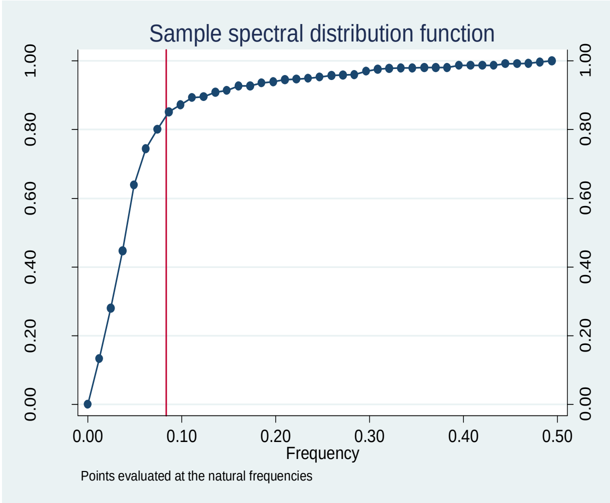

Figure 4.2: Displays the sample spectral distribution function of the time series under study.

In Figure 4.5, at the vertical line, one notices no jump. This indicates absence of annual cycle in the dataset. This observation has been expected, given the nature of the dataset (taxes) records from an unstable milieu.

4.2 Kernel density estimation

In this section, several kernels are considered. A comparison of performance between Epanechnikov kernel on one side versus *Parzen*, *Biweight*, *Epan2* and *Cosine*(PBEC) kernels on the other side as described in 3.2 through 3.7.

4.2.1 Kernel density fitting – Epanechnikov versus PBEC kernels

Recall that a kernel density estimate is built in assuming the weighted values with the kernel function, say K, in the density

$$\hat f_K = \frac{1}{qh}\sum_{i=1}^{n}\omega_i\,K\left(\frac{x-x_i}{h}\right) \qquad (4.5)$$

With $q = \sum_{i=1}^{n}\omega_i$ is the weights are frequencies (or analytic), or q = 1 if weights are importance weights. In practice, analytic weights are rescaled so that $\sum_{i=1}\omega_i = n$. When weights are not used, it is customary to assume w_i = 1 for *i =1,…,n*.

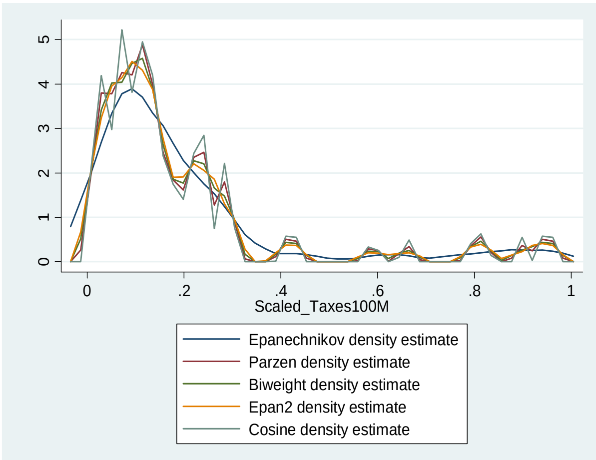

Figure 4.3: Depicts a multi-plots comparison between the epanechnikov density estimate curve versus parzen, biweight, Epan2 and cosine density estimate curves.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

In measuring the performance for each density estimate, the Mean Absolute Deviation (MAD) is considered. Formally, this is defined as

$$MAD_{k,i} = \frac{1}{N}\sum_{j=1}^{N}\left|x_j^{data} - \hat{x}_{k,j}\right| \qquad (4.6)$$

where N is the number of data points, k is the kernel density considered, x_j^{data} are dataset that are subject for density estimation, j=1,…,N and $\hat{x}_{k,j}$ is the data point estimated by the kernel density k.

Let K denote the set of all kernel density estimates considered. That is K={k_e, k_p, k_b,k_{ep2},k_{cos}} where the subscripts denote the initial for each kernel used. Let $MAD_{k.,i}$ denote the mean absolute deviation scored by the kernel density of reference and $MAD_{k(.),i}$ the mean absolute deviation for a specific kernel. To measure efficiency of any reference kernel used, one makes the following comparison, using equation (58): The kernel $\mathbf{k_e}$ is considered outperforming any other kernel, k(.) if

$$MAD_{k_e} < MAD_{k_{(.)}} \qquad (4.7)$$

Otherwise one considered it underperformed it competitor. Prior to searching for an optimal bandwidth, h, the performance of 5 kernel density estimates is examined, including Epanechnikov, Parzen, Biweight, Epanechnikov2 and Cosine kernel density estimates.

Table 4.1: Displays the mean absolute deviation for Epanechnikov, Parzen, Biweight, Epanechnikov2 and Cosine kernel density estimates.

```
summarize D_epan D_parzen D_biweight D_epan2 D_Cosine
```

Variable	Obs	Mean	Std. Dev.	Min	Max
D_epan	50	.7407645	1.02216	.0013504	3.075671
D_parzen	50	.8473191	1.139632	.0045433	3.942162
D_biweight	50	.8170499	1.128722	.0073345	3.641644
D_epan2	50	.8046727	1.122904	.0050461	3.684663
D_Cosine	50	.8912979	1.181603	.01	4.315539

From table 4.2, third column, one observes that the Epanechnikov kernel outperforms its competitors. Its MAD is of 0.7407645 and all other kernels scored higher values. In terms of risks, it does also very well, with the smallest standard deviation amounting to 1.02216. One need to remember that the bandwidth was not specifies. This implies that the optimal bandwidth, precisely h, had not been used.

4.2.2 Kernel density fitting – Epanechnikov versus GTR kernels

In this section, using a default bandwidth h, the performance of Epanechnikov kernel is compared to the ones of its three competitors, including Gaussian, Triangle and Rectangle (GTR) kernels.

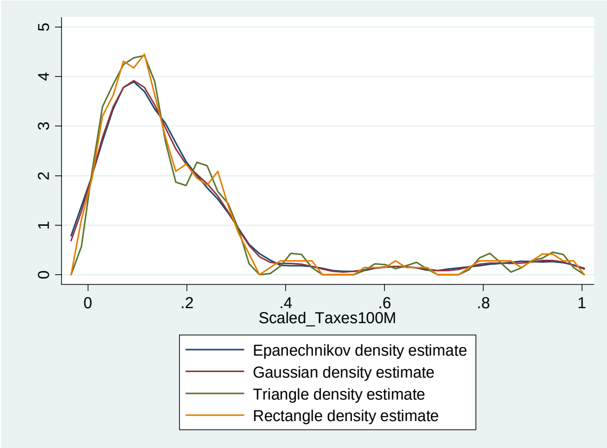

Figure 4.4: Depicts a multi-plots comparison between the Epanechnikov density estimate curve versus Gaussian, triangle and rectangle density estimate curves.

Of the three competitors, the Gaussian kernel, appears closer to the reference kernel, the Epanechnikov. In terms of risks, that is MAD described in (4.10q) through (4.11), both the Epanechnikov and the Gaussian kernels exhibit tremendous results. While the Gaussian kernel recorded less MAD as compared to any other kernel considered, it is the Epanechnikov kernel which recorded the smallest standard deviation. This implies that both

Table 4.2: Displays the mean absolute deviation for Epanechnikov, Gaussian, Triangular and Rectangular kernel density estimates

```
. summarize D_epan D_gaussian D_triangle D_rectangle
```

Variable	Obs	Mean	Std. Dev.	Min	Max
D_epan	50	.7407645	1.02216	.0013504	3.075671
D_gaussian	50	.7373619	1.029984	.0028654	3.098926
D_triangle	50	.8113494	1.122531	.0028052	3.559049
D_rectangle	50	.7961119	1.099585	.0007384	3.516372

Kernels can be used for reliable predictions or forecasting exercises. Still, in this evaluation, the default bandwidth, h, had been used. When the bandwidth is not specified by the researcher, conventionally, it is determined by the algorithm in the software used as follows

$$m = min\left(\sqrt{var(x)} - \frac{Interquartile \quad range\ (x)}{1.349}\right); h = \frac{0.9m}{n^{1/5}} \qquad (4.8)$$

To recap these results, one realizes that the Epanechnikov and the Gaussian kernels performed reasonably well as compared to other kernel competitors, when the bandwidth by default is used. In real life, one needs to select an optimal bandwidth for efficiency in forecasting.

4.2.3 Determination of the bandwidth h

In this section, optimal bandwidth selection is considered. Before then, two observations are made. First, it is convenient to precise that optimal bandwidth is selected within the kernel weighted local polynomial smoothing which is described below.

(a) Local polynomial smoothing

In this case, one considers the model

$$y_i = m(x_i) + \sigma(x_i)\varepsilon_i \qquad (4.9)$$

Where **m(.)** and **σ(.)** are respectively unknown mean and variance functions, and ε_i the symmetric errors assumed to have zero mean and unit variance. With no assumption on the functional form of **m(.)**, the idea is to estimate **m(**x_0**) = E[y|X=** x_0**].** Using STATA, the command lpoly generally estimates like the intercept of the regression, weighted by the kernel function that is selected of the dependent variable on the polynomial terms $(x - x_0)$, $(x - x_0)^2, \cdots, (x - x_0)^p$ for each smoothing data point.

The weight matrix in this case is defined as $W = diag\{K_h(x_i - x_0)\}_{nxn}$

where the weights are $K_h(.)$ defined by $K_h(x) = h^{-1}K(x/h)$, with $K(.)$ being the kernel selected and h the bandwidth. In general, as the degree of the polynomial increases, the smoother the kernel density estimate.

In this section, three kernel are considered: Epanechnikov, Gaussian and the alternative Epan2 kernels, for p ={1,3,7} degree.

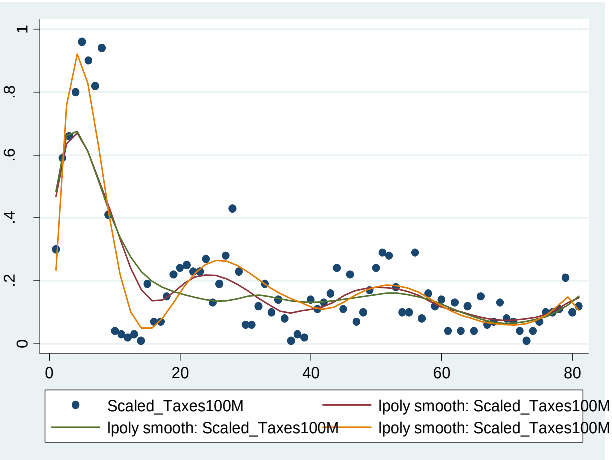

Figure 4.5: Depicts Epanechnilov kernel weighted local polynomial smoothing for p ={1,3,7}.

In the current Figure 4.5, for p = 7, it can be observed that the curve produced has a better approximation of the data than the other two ones. In Figure 4.9, the same phenomenon is confirmed once more, the higher the order of the polynomial, the better is the approximation of the kernel density. At p = 7, the Gaussian kernel exhibits similar behavior as observed for the Epanechnikov kernel. However, a slight difference

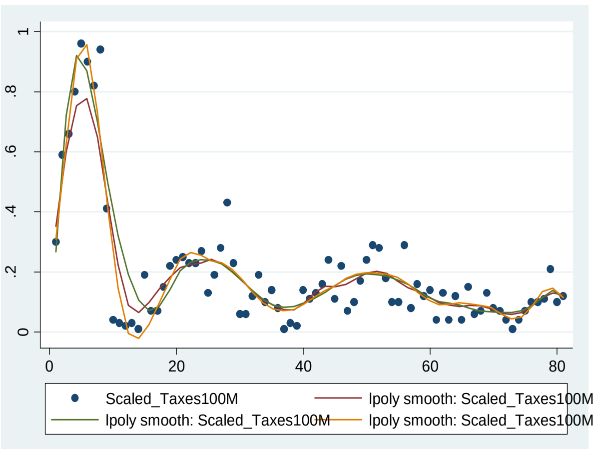

Figure 4.6: Depicts alternative Epanechnikov (epan2) kernel weighted local polynomial smoothing for p ={1,3,7}.

In Figure 4.6, the alternative Epanechnikov (epan2) behaves in similar fashion as the Gaussian kernel. In summary, for the three kernels considered, it came clearly noticeable that the higher the order of the polynomial, the smoother the kernel density estimate.

(b) Selection of the bandwidth

There are two choices that can be explored. One may decide to use a default bandwidth which uses equation (4.8) above, or simply opts for optimal bandwidth in which case kernel weighted local polynomial smoothing is implemented in (4.9).

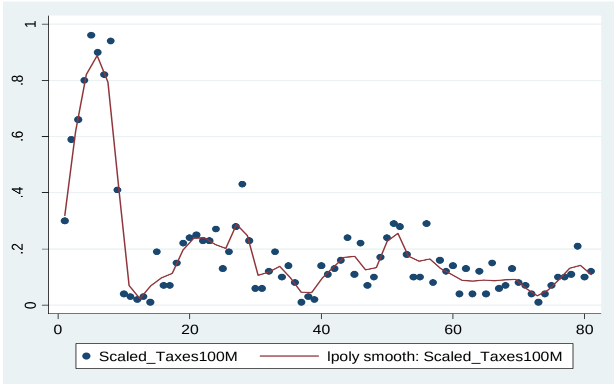

Figure 4.7: Depicts the performance *of the Epanechnikov for p=1, and the bandwidth = 1*

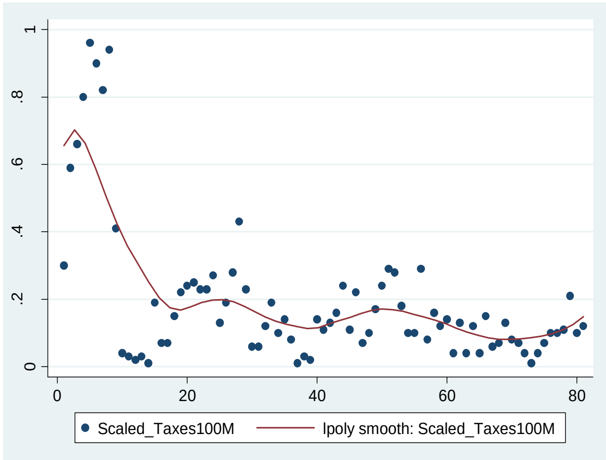

Figure 4.8: Depicts the performance of the Epanechnikov for p = 1 andbandwidth = 5.

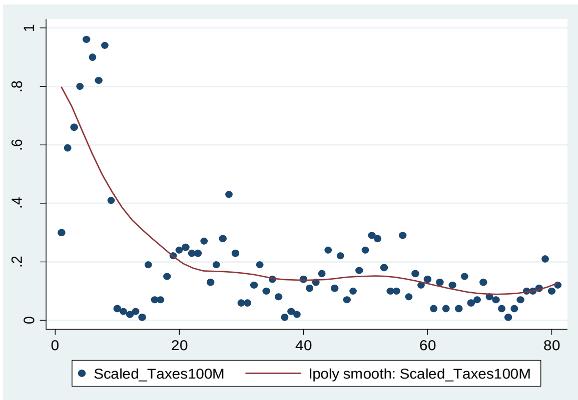

Figure 4.9: Depicts the performance Epanechnikov for p = 1 and the bandwidth = 7.

In Figure 4.7 through 4.9, it can be observed that the smaller the bandwidth, the more accurate the approximations of the kernel density estimate. For the same kernel, and for the same order of the polynomial, bandwidth equals to a unit produces a density estimate which captures more data points than the two competitors bandwidths.

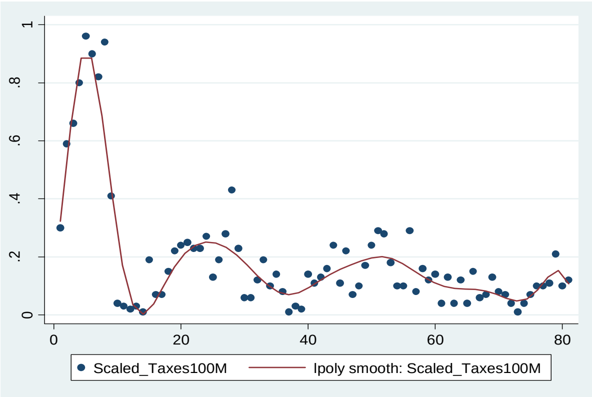

Figure 4.10: Depicts the performance Gaussian kernel for p = 5 and the bandwidth = 5

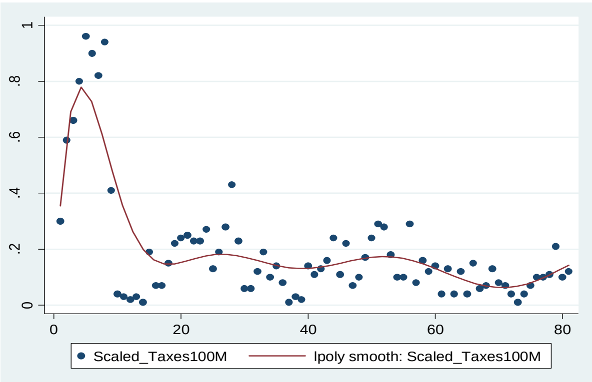

Figure 4.11: depicts the performance Gaussian kernel for p = 5 and the bandwidth = 10

Using the Gaussian kernel, Figure 4.10 through 4.11, one observes that for p=1 all through, with the bandwidth of a unit, one gets a density estimate which captures most data points. This confirms once again that the less is the bandwidth, the better is the approximation, but also the lesser smoother is the density estimate, as in Figure 14.

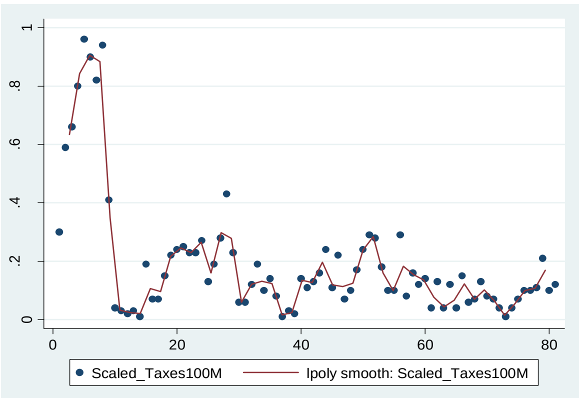

Figure 4.12: Depicts the performance Epanechnikov2 for p = 5 and the bandwidth = 1

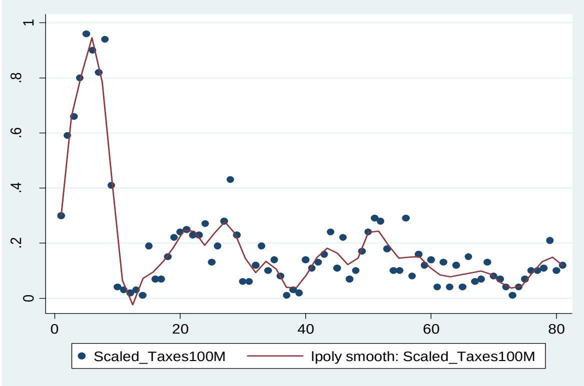

Figure 4.13: Depicts the performance Epanechnikov2 for p = 5 and the bandwidth = 7

In fitting the alternative Epanechnikov kernel weighted local polynomial, from figure 12 through 4.13, one observes the following points: for p = 5, bandwidth = 1, the kernel density estimate appears to have jigsaw features and in the process, it captures most data points (Figure 4.13). However, all the remaining kernel density estimates captured less data points even though smoother than the one depicted in Figure 4.18. Examining the performance of the alternative Epanechnikov (epan2) in terms of bandwidth choice, one makes the following observations relative to Figure 18 through 21. For p = 1 fixing h = {1,5,7,10}, the smaller the bandwidth (h = 1, Figure 4.18) the more accurate the estimates. However, the larger the bandwidth, (for instance h = 10, Figure 4.21), the more errors the estimates contain.

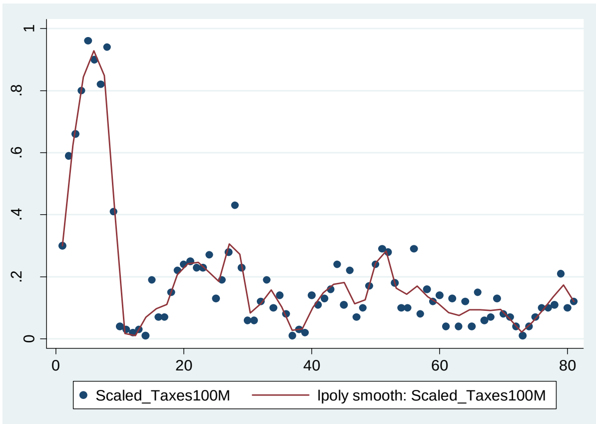

Figure 4.18: Depicts the performance Epanechnikov2 for p = 5 and the bandwidth = 1

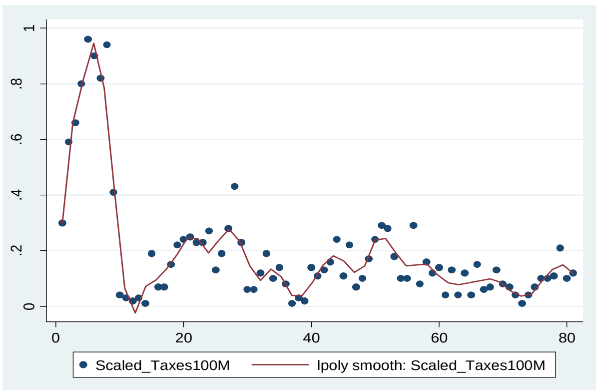

Figure 4.20: Depicts the performance Epanechnikov2 for p = 5 and the bandwidth = 7

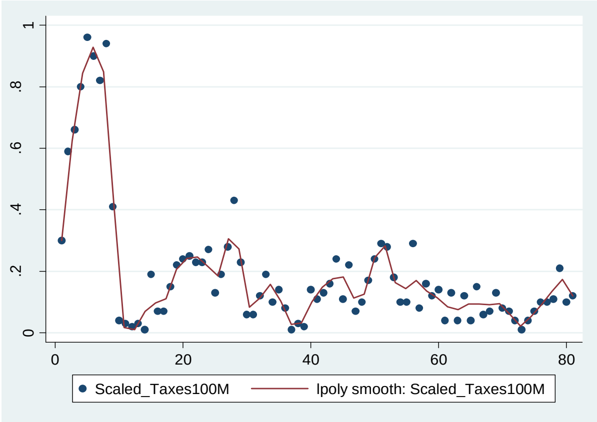

Figure 4.19: Depicts the performance Epanechnikov2 kernel for p = 5 and the bandwidth = 5

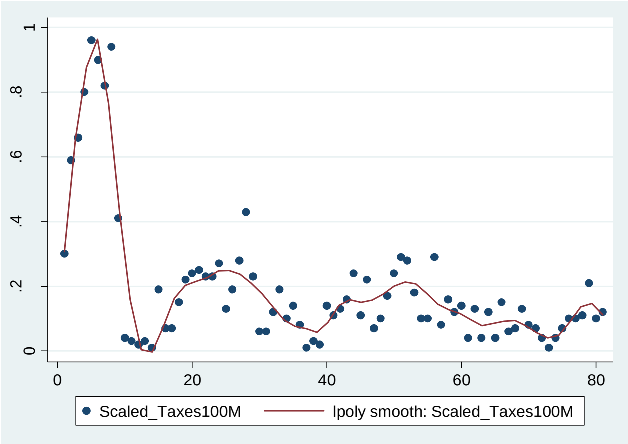

Figure 4.21: depicts the performance Epanechnikov2 for p = 5 and the bandwidth = 10

Comparison between default bandwidth and optimal bandwidth selection.

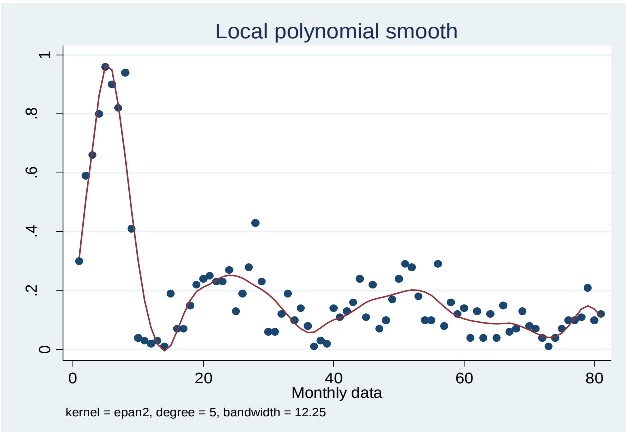

Figure 4.22: Depicts default bandwidth h = 12.25

4.3 Summary

The current chapter dealt with the implementation of the kernel density estimation. Data used have been visualized and different kernel density used.In the spectrum of data used, (see figure 4.2) trends, seasonality and cycles are not tractable, but random components of the time series. Time series analysis findings suggest strongly significant autocorrelation (Table 4.1).

To verify absence of the cycles, sample spectral density has been carried out. Findings reveal the expected; that is, no annual cycles in the time series under study. As mentioned above, a time series with no trends, seasonality and cycles need different analytic treatment, particularly, the use of kernel density estimation methods.

Having defined different kernel densities that can be used, the research compared, first, the performance of the Epanechnikov kernel against the performance of Parzen, Biweight, Epan2 (the alternative Epanechnikov) and the cosine kernels. Findings in this comparison suggest that only Epan2 is similar to the performance of the index kernel. All kernels considered display higher values of MAD as compared to the index kernel, which is the Epanechnikov. Secondly, a comparison of the performance of the index kernel versus Gaussian, Triangle and Rectangle kernels was made. Findings suggest strongly that only the Gaussian performs as good as the Epanechnikov kernel, but the two others exhibited under performance.

The determination of the optimal bandwidth has been carried out under kernel weighted local polynomial smoothing setup (figure 4.8 through 4.10). Two aspects of the problem were examined, namely the degree of the polynomial that adequately fit the data points and the magnitude of the bandwidth required to attain bell fit. In this respect, the research examined the performance of the Epanechnikov, the Gaussian and the alternative Epanechnikov (epan2) kernel using kernel weighted polynomial of degrees 1, 3 and 7 for different values of the bandwidth, precisely for h = {1,5,7,10}. Findings suggest strongly that when the degree of the polynomial is great, and the measure of the bandwidth used small, the likelihood of getting bell-fit to the data is certainly very high.

5. Conclusion and Recommendation

5.1 Conclusion

This intent of this chapter is to lay down conclusions based on research findings and eventually presents some recommendations either to decision makers and/or to the scientific community.

In the current study, the main object has been to model tax revenues time series using kernel density estimation methods. The reason prompting the selection of this method has been that, upon visual inspection of the time series under study, the trend, seasonality and cycles were all not tractable, but random components (Figure 4.2). A preliminary time series analysis under Portmarteau test statistic and Bartlett's test reveals respectively, at lag 12, existence of autocorrelation and of partial autocorrelation (Table 4.1, Figure 4.3).

The attempt to examine presence of cycles in the time series under study has been carried out through sample spectral distribution analysis. Findings in this respect suggest eventually that the dataset under examination does not contain annual cycles. Since in figure 4.5, no jump is observed, this is a clear indication that no cycles are present in the data under study. A time series such as the one under examination which is subjected to a regime with absence of trend, seasonality and cycles definitely need different analytical treatment than the usual time series analysis proposed in most time series and econometrics manuals.

To model tax revenues, kernel density estimation method is proposed. Common or basic kernels that are often on the table in probability and statistics have been presented as well as their individual support. These include the epanechnikov, the alternative Epanechnikov (epan2), the biweight, the cosine, the Gaussian, the parzen, the rectangle and triangle kernel functions. Two main issues have been cleared in this respect, mainly, the choice of best kernel that has the potential to fit the data points and the determination of the bandwidth to use for that matter.

Using the mean absolute deviation (MAD) to measure the performance of each kernel against the kernel index, precisely the epanechnikov kernel, and a comparison has been observed in that direction. The examination of the performance of Parzen, biweight, epan2 and cosine kernel against Epanechnikov singled out the epan2 displaying like performance as the index kernel. In similar evaluation of the performance of Gaussian, Triangle and Rectangle kernel versus the Epanechnikov, the Gaussian kernel displays greater similarities with the index kernel.

Hence, to recap on the choice of the best kernel, by default, the Epanechnikov kernel remain the index kernel; the Gaussian and the alternative Epanechnikov can also be used to model tax revenues.

It is important to state, on the basis of literature review that the question which kernel to use (to model tax revenue) is not often given much weight, but rather the determination of the bandwidth which regulate the smoothness of the kernel density that is being used. To this end, having chosen to use, for comparison purposes, the Epanechnikov, the Gaussian and the alternative Epanechnikov kernels, an optimal choice of the band with has been discussed through the kernel weighted polynomial smoothing setup. Two crucial aspects of the problem were evaluated, including the degree of the polynomial that precisely fit the data points and the level of the bandwidth that is required to achieve bell-fit.

To this end, the performance of the Epanechnikov, the Gaussian and the alternative Epanechnikov (epan2) kernel using kernel weighted polynomial of degrees 1, 3 and 7 for different values of the bandwidth, precisely for $h = \{1,5,7,10\}$ has been examined. As expected, findings suggest unequivocally that when the degree of the polynomial is high, and the bandwidth small, any of the three kernels attains a desirable bell-fit to the data under study.

Hence, to predict or forecast tax revenues, either the Epanechnikov, the Gaussian or the alternative Epanechnikov (epan2) kernel can be used, with a careful choice of the pair (p,h) where p is the degree of the polynomial which is assumed to be reasonably high and h is the optimal bandwidth.

5.2 Recommendations

On the basis of findings in the current study, two recommendations addressed to decision makers and to scientific community, are plausible;

1) If dataset whose distribution displays random walks characteristic are to be used for prediction or forecasting purposes, then kernel density method can be used to model the data set in question.
2) Of all basic kernels, use any of the following kernels to predict future (Tax Revenues): the epanechnikov as the index kernel, the Gaussian and the alternative Epanechnikov as best two competitors capable of producing bell-fit of the data.

Select optimal bandwidth and use of higher degree of polynomial enable desirable accuracy in the prediction.

References

[1] Perez Ana. Comment on kernel density estimation for time series data. International Journal of Forecasting, 28:15–19, 2012.
[2] Harvey Andrew and OryshchenkoVitaliy. Kernel density estimation for time series data. International Journal of Forecasting, 28:3–14, 2012.
[3] M. S. Bartlett. Smoothing periodograms from time series with continuous spectra. Nature, (161):686– 687, 1948.
[4] M.S. Bartlett. Periodograms analysis and continuous spectra. Biometrika, (37):1 – 16, 1950.
[5] Bruce L. Bowerman and Richard T. O'Connell. Forecasting and Time Series: An Applied Approach, volume Third edition. Duxbury Press, California, 1993.
[6] A.W. Bowman and A. Azzalini. Applied Smoothing Techniques for Data Analysis: The Kernel Approach. Oxford University Press, Oxford Statistical Science Series, London, 1997.
[7] M. Charytanowicz, J. Niewczas, P. Kulczycki, P.A. Kowalski, S. Lukasik, and S. Zak. A complete gradient clustering algorithm for features analysis of x-ray images. Springer, Berlin, http://dx.doi.org/10.1007/978-3-642-13105-9-2/jss.v027.i07, pages 15 – 24, 2010.
[8] Deng and H. Wickham. Density estimation in r. had.co.nz, 2011.
[9] DeWayne R. Derryberry. Basic Data Analysis for Time Series with R. John Wiley and Sons, Inc., Hoboken, New Jersey, 2014.
[10] T. Duong. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. Journal of Statistical Software, http://dx.doi.org/10.1863/jss.v021.i07, (21):1 – 16, 2007.
[11] T. Duong. R package 'ks' version 1.9.2. http://CRAN.R-project.org/package=ks, 2014.
[12] T. Duong. Kernel smoothing (r package version 1.10.6). http://CRAN.R-project.org/package=ks, 2017.
[13] T. Duong, A. Cowling, I. Koch, and M.P. Wand. Feature significance for multivariate kernel density estimation. Journal of Computational Statistical Data Analysis, http://doi.org/10.1016/j.csda.2008.02.035, (52):4225 – 4242, 2008.
[14] G. Elliot and A. Timmermann. Economic Forecasting. Princeton University Press, US, 2016.
[15] M Ester, H. P Kriegel, Sander J, and Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. in Proceedings of 2nd International

Conference on Knowledge Discovery and Data Mining (KDD-96), pp. 226-231, 1996.
[16] Joaqing Fan and Qiwei Yao. Nonparametric Time Series: Nonparametric and Parametric Methods. Springer Verlag, New York, 2003.
[17] ArturGramack. Nonparametric Kernel density estimation and its computational aspects, volume 37. Springer, Studies in Big Data, New York, 2017.
[18] Arthur Gramacki. Nonparametric Kernel Density Estimation and its computational aspects, volume 37. Springer International Publishing AG 2018, New York, 2018.
[19] J. A Hartigan and M. A Wong. Algorithm as 136: A k-means clustering algorithm. J. R. Stat. Soc. Ser. C (Appl. Stat.)., http://www.jstor.org/stable/2346830, 28(1):100 – 108, 1979.
[20] Tristen Hayfield and Jeffrey S. Racine. Non parametric econometrics: The np package. Journal of Statistical Software, http://dx.doi.org/10.18637/jss.v027.i07, 27(5):1 – 32, July 2008.
[21] McLachlan G. J. Discriminant Analysis and statistical Pattern Recognition. Wiley International Science, New York, 2004.
[22] Hart D. Jeffrey. Kernel regression estimation with time series errors. J.R. Statist. Soc., 53(1):173–187, September 1991 1991.
[23] Fan Joanqing and Yao Qiwei. Time Series Nonparametric and Parametric Methods. Springer-Verlag, New York, 2003.
[24] Matthias JukkaKrisp and Peters Stefan. Directed kernel density estimation (dkde)for time series visualization. Annals of GIS, 17:155–162, September 2011 2011.
[25] Liu and Y. Wang. Hypothesis testing in smoothing spline models. Journal of Statistical Computation and Simulation, (74):581 – 597, 2004.
[26] Liu, T. Tong, and Y. Wang. Smoothing spline estimation of variance functions. Journal of Computational and Graphical Statistics, (16):312 – 329, 2007.
[27] Liuand W. Meiring and Y. Wang. Testing generalized linear models using smoothing spline methods. StatisticaSinica, (15):235 – 256, 2005.
[28] Chikhi Mohamed and Diebolt Claude. Nonparametric analysis of financial time series by the kernel methodology. Qual Quant., 44:865–880, May 12 2009.
[29] Douglas C. Montgomery, Cheryl L. Jennings, and Murat Kulahci. Introduction to Time Series Analysis and Forecasting (Wiley Series in Probability). JohnWiley and Sons, Inc., Hoboken, New Jersey, 2008.
[30] E. A. Nadaraya. On estimating regression. Theory Probability Applications, http://dx.doi.org/10.1137/1109020, 9:141 – 142, 1964.
[31] E. Parzen. On estimation of probability density function and mode. Annals Math. Statist. (33):1065 – 1076, 1962.
[32] P.M Robinson. Kernel estimation and interpolation for time series containing missing observations. Ann. Inst. Statist. Math., 36:403–417, March 12 1984.
[33] David W. Scott. Multivariate Density Estimation: Theory and Practice and Visualization. Wiley, London, 1992.
[34] B.W. Silverman. Density Estimation for statistics and Data analysis. Chapman and Hall/CRC, London, 1989.
[35] G. Terrell and D. Scott. Variable kernel density estimation. The Annals of Statistics, 3(30): 1236 –1265, 1992.
[36] Michael W. Trosset. An introduction to statistical inference and its application with r. November 10, 2008.
[37] Ruey S. Tsay. Multivariate Time Series Analysis with R and Financial Applications. John Wiley and Sons, Inc.,, Canada, 2014.
[38] M.P. Wand and M.C. Jones. Kernel Smoothing, Monographs on Statistics and Applied Probability 60. Springer Science Business Media, B.V., London, 1992.
[39] M.P. Wand and B. Ripley. Functions for kernel smoothing supporting wand and jones (1995), (r package version 2.23-15). http://CRAN.R-project.org/package=KernSmooth, 2015.
[40] M.P. Wand, G. Luta, and U. Naumann. curvhdr: Filtering of flow cytometry samples, (r package version 1.1-0). http://CRAN.R-project.org/package=curvHDR, 2016.
[41] Y. Wang and C. Ke. Smoothing spline semi-parametric nonlinear regression models. Journal of Computational and Graphical Statistics, (18):165 – 183, 2009.
[42] YuedongWang. Smoothing Splines, Methods and Applications. Monographs on Statistics and Applied Probability 121, CRC Press, Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742, 2011.
[43] G. S. Watson. Smooth regression analysis. Sankhya: Indian J. Stat.Ser. A., http://www.jstor.org/stable/25049340, 26(4):141–142, 1964.
[44] M. Yuan and GWahba. Doubly penalized likelihood estimator in heteroscedastic regression. Statistics and Probability Letters, (69):11 – 20, 2004.
[45] Wei Yuyan. An introduction to density forecasting for time series. Overview: What is Density Forecast?, pages 1 – 10, April 9 2017.
[46] K. Zychaluk and P. N. Patil. A cross-validation method for data with ties in kernel density estimation. Ann. Inst. Stat. Math., http://doi.org/10.1007/s10463-006-0077-1, (60):21 – 44, 2015.
[47] StataCorp. 2009. Stata: Release 11. Statistical Software. College Station, TX: StataCorp LP.
[48] Robert H. Shumway, David S. Stoffer, (2006), Time series analysis and its applications With R Examples, Second edition, Springer Science plus business Media, LLC, USA.
[49] Jianqing Fan, Qiwei Yao, Nonlinear Time Series: Nonparametric and parametric methods, Springers, New York, 2003.
[50] Hjort, N.L. and Jones, M.C. (1996b). Better rules of thumb for choosing bandwidth in density estimation. Statistical Research Report, Department of Mathematics, University of Oslo, Oslo, Norway.
[51] Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and Its Applications*, **9**, 141–142.
[52] Watson, G.S. (1964). Smooth regression analysis. *Sankhya, Series A*, **26**,359–372.
[53] Jeffrey D. Hart, Kernel regression estimation with time series errors, J.R. Statist. Soc. B (1991), 53, No 1, pp. 173 - 187