Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 23 Issue 1 (2021) ISSN$_{Print}$: 1816-9171

Using kernel smoothing methods for oil Iraqi prices

Prof. Dr. Mohammed Habeb Al-Sharoot Habeb Kazem Alwan
Al-Qadisiyah University: College of Administration and Economics/ Department of Statistics
Corresponding Author: Habeb Kazem Alwan

Abstract **:** In this paper we use the nonparametric methods to estimate the time series models , which is different of the parametric methods, in which the data is given the opportunity to express itself and the principle of letting the data speak for itself and estimating the time series model. The time series were used representing the monthly final prices of a barrel of Iraqi crude oil in US dollars for the period from January 2003 to June 2020 by 210 observations . we use some non-parametric methods such as the Kernel Smoothing method represented by Nadaraya- Watson (NW) and local polynomial. So, we use some different methods to choose the smoothing Parameter, such as the plug -in method, smoothing cross validation method and Least Squared Cross Validation method , and some precision criteria such as (MSE, MAE, MAPE). We have been calculated to compare between the applied method models. We found that the cubic polynomial estimator (CU) is the best nonparametric method to estimate the monthly final prices of a barrel of Iraqi crude oil time series model

INTRODUCTION: Observations are often successive through time, and future values usually depend on the random behavior of available observations in the past. This dependence makes it useful to predict the future from those values in the past, in addition to the basic dynamic from which the observed data is generated, so the series can be defined. Time Series is define as the data arranged depending on time, this means that the time series is a group of observations related to each other, which are recorded for a certain phenomenon in previous time periods and are arranged sequentially according to time, and one of the benefits of time series is to understand the underlying dynamic that is create existing data, predict future events, and control future events through intervention[3] . Nonparametric ideas are being applied since a long time is smoothing and decomposing seasonal time series. Local polynomial regression can be traced back to 1931 (R.R. Macaulay). A. Fisher (1937) and H.L. Jones (1943) discussed a local least squares fit under the side condition that a locally constant periodic function (for modeling seasonal fluctuations) be annihilated and already in 1960 J. Bongard developed a unified principle for treating the interior and the boundary part (with and without seasonal variations) of a time series derived from a local regression approach[7]. Nonparametric regression has become an area with an abundance in new methodological proposals and developments in recent years .In this paper we discuss the application of some nonparametric techniques to time series .There is indeed a long tradition in applying nonparametric methods in time series analysis, In contrast to that in nonparametric regression no assumption is made about the form of the regression function. Only some smoothness conditions are required. The complexity of the model will be determined completely by the data. One lets the data speak for themselves. one avoids subjectivity in selecting a specific parametric model. But the gain in exibility has a price. Besides this, a higher complexity in the mathematical argumentation is involved.

2. Non-Parametric Smoothing

The nonparametric smooth method can provide a flexible tool for analyzing unknown regression relationships, This tool is used to describe the direction and effect of the explanatory variable on the response variable, and this tool can be represented as a function in terms of one or more explanatory variables. A typical situation for an application to a time series $\{z_t\}$ is that the regressive vector$\{X\}$ consists of past time series values

$X_t = (z_{t-1},\ldots\ldots , z_{t-p})$

which leads to the general nonparametric auto regressive model

$z_t = m(z_{t-1},\ldots\ldots , z_{t-p}) + at \quad , t = p + 1,p + 2 ,\ldots..$ (1)

with $\{a_t\}$ a white noise sequence. Of course $\{x_t\}$ might also include time series values of other predictive variables like leading indicators[7] .The function (m) in the formula (1) is the nonparametric regression function as it relies on the data to determine the space of the function(m) and that the researcher assumes only the preamble properties, which are (continuity, derivability, or derivative with the square integral of the second derivative) that you have the function m. The result of using nonparametric methods in estimating is to obtain a smoothed function ($\hat{m}$). Therefore, we will deal with two types of nonparametric methods which are Kernel

2. Kernel estimation in time series

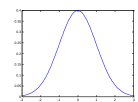
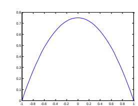
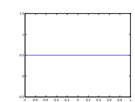
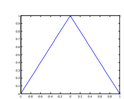
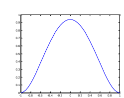
The kernel smoothing provides a statistical method for estimating the nonparametric regression function or for estimating the conditional prediction function, as it aims to find a nonlinear relationship between pairs of random variables as well as to find a structure or pattern of data that enables us to find the structure of a set of data without need a parameterized model, as this smoothing represents estimation of the regression function shown in formula (1) at a given point based on locally. When a kernel estimator is applied to dependent data, as it is the case in time series,

then it is effected only by the dependence among the observations in a small window and not by that between all data. This fact reduces the dependence between the estimates ,so that many of the techniques developed for independent data can be applied in these cases as well. This fact was called the whitening by windowing principle by Hart (1996).
The first step to go is therefore to look at nonparametric estimation of densities and conditional densities. Let $x_j \in IR^p$ be vector a random variable whose distribution has a density (f)and let z_{t-1}, ……..,z_{t-p} be a random sample from x . Then a kernel density estimator for(f) in time series applications frequently product kernels are applied is given by[7]

$$f_n(X) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{p}\frac{1}{n_j}k_j\left(\frac{z_{ij}-z_j}{h_j}\right) \dots\dots\dots\dots..(2)$$

where K is a so-called kernel function, i.e. a symmetric density assigning weights to the observations z_{ij} which decrease with the distance between z_i and z_j. Some popular kernel functions are listed in Table (1). h_j is the bandwidth which drives the size of the local neighborhood being included in the estimation of (f)at z_j . A very small bandwidth will lead to a wiggly course of the estimated density, whereas a large bandwidth yields a smooth course but will possibly atten out interesting details.

Table (1) some of the kernel functions

Function Name	Mathematical Formula	Interval	Graph
Gaussian	$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$	$I\{\lvert x\rvert \le \infty\}$	
Epanechnikov	$K(x) = \frac{3}{4}(1-x^2)$	$I\{\lvert x\rvert \le 1\}$	
Uniform	$K(x) = \frac{1}{2}$	$I\{\lvert x\rvert \le 1\}$	
Triangular	$K(x) = (1-\lvert x\rvert)$	$I\{\lvert x\rvert \le 1\}$	
Quartic	$K(x) = \frac{15}{16}(1-x^2)^2$	$I\{\lvert x\rvert \le 1\}$	

2.1 Nadaraya-Watson Estimator

We will just give a very brief introduction to the most well-known local weighted average estimator, the Nadaraya-Watson estimator. The estimator can be also be though of as local constant estimator which is a special case included in the local polynomial estimator that we will introduce later.[7]
In general let the random vector be (z, x), zϵR, xϵR ^ p

$g(z/x)=\frac{f(z,x)}{fx(x)}$

$k=R^{p+1} \to R, \quad K(z,x)=K_1(z)K(x)$

$$g(z/x)=h\frac{\sum_{i=1}^{n} K1\left(\frac{z_i-z}{h_1}\right)K\left(\frac{x_i-x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)} \quad \dots\dots\dots..(2)$$

Our goal is to predict a point by estimating m (x) and using conditional expectation of equation (1). The conditional expectation.

m (x)=E (Z/x)

$$m(x) = \int_{-\infty}^{\infty} zg(^{z}/_{x})dz \dots\dots\dots\dots..(3)$$

Substituting equation (2) into equation (3), we get The Nadaraya-Watson estimator is defined as:

$m_n(x)=\frac{\sum_{i=1}^{n} z_i \; K(\frac{x_i-x}{h})}{\sum_{i=1}^{n} K(\frac{x_i-x}{h})}$

It is easy to verify that the estimator is the weighted sum of z_i, that

$m_n(x) = \sum_{i=1}^{n} z_i w_{n,i}(x: x_1, \dots x_n)$ ……………(4)

The estimator can handle both fixed design and random design with non-uniform distributions. It also can be easily extended to multivariate cases.[7]

2.2 Local Polynomial Regression

Local polynomial regression was introduced into the statistical literature by Stone (1977) and Cleveland (1979). The statistical properties were investigated by Tsybakov (1986), Fan (1993), Fan and Gijbels (1992,1995), Ruppert and Wand (1994) and many others. A detailed description may be found in the book of Fan and Gijbels (1996).[7]

The local polynomial regression estimator that has gained wide acceptance as an attractive method for estimating the regression function and its derivatives. Some of its advantages are the better boundary behaviour, its adaptation to estimate regression derivatives, its easy computation and its good minimax properties ,among others. This estimator is obtained by fitting locally to the data a polynomial of degree p, using weighted least squares. More specifically, m is assumed to be smooth in the sense that the (p+1)th derivative exists at x, so that it can be expanded in a Taylor series around x . The local polynomial estimator of m (x) is defined as the value, β0, that minimizes

$\text{m(x)}=\sum_{i=1}^{n}(z_i - \sum_{j=0}^{p} \beta j (Xi-x)\;)^2\; K(\frac{Xi-x}{h})$……………..(5)

With the design matrix X having the n rows [1; x_i – x,….. , $(xi - x)^r$], the diagonal weight matrix W = diag K(xi-x/h) and the vector z = $(z_{1\dots\dots}, z_n)^T$ the solutions at

x is given by

$\widehat{m}(\;x)=\hat{\beta}_0=\text{e}^T{}_1(X^T\text{WX})^{-1}X^T WZ$ …………(6)

In the particular case of p = 1,2,3 the local linear ,quadratic and cubic polynomial estimator is obtained.[12]

3.Parameter selection

The problem of selecting the smoothing parameter, which controls the bandwidth, is one of the main problems facing the researcher when estimating the nonparametric regression function as well as in the prediction process for the time series. It is the vital component of the estimation process for the function m and controls the neighbor width of point x at which the estimate is [x-h, x + h] .It is well known that in practice the choice of the kernel is not very important compared to the choice of the bandwidth. The most important task in kernel smoothing is the bandwidth selection. It is very well-known that a large bandwidth would give over smoothed estimations, with a large bias. On the other hand, if the bandwidth is too small, the estimation becomes under smoothed and the its variance gets large. An optimal bandwidth is achieved when the changes in bias and variance balance .There are plenty of papers that have dealt with the problem of bandwidth selection for independent data, but, under dependence, this problem has been much less studied. In general, there are three different types of selecting methods.[3]

3.1 Plug-in method

They are based on the idea of obtaining the bandwidth that minimizes some estimation of the asymptotic mean integrated squared error of the estimator (or some other global or local error measure). Their performance is good in the fixed design case but much worse in the random design case under dependence for more details see Ruppert , Sheather and Wand (1995). In time series applications we are mainly interested in a constant, global bandwidth, for which the integrated mean squared error (IMSE)[3]

$$IMSE = \int [Bias(\widehat{m}^{(j)}(x))^2 + Var(\widehat{m}^{j}(x))]w(x)dx \dots\dots\dots\dots\dots\dots\dots\dots\dots (7\;)$$

By minimizes the above criterion, we get the constant optimal bandwidth.

$$h_n^* = c_0(\frac{p}{n})^{\frac{1}{(p+4)}} \dots\dots\dots\dots\dots\dots\dots\dots (\,8\,)$$

$$C_0 = [\frac{v_0}{\mu_2^2}\,\frac{\sigma^2(x)}{f(x)tr\{H_m(x)\}}]^{\frac{1}{(p+4)}} \quad \dots\dots\dots\dots\dots\dots . (9\,)$$

And H_m is the second derivative matrix of the function (m)

3.2 cross-validation method

When using the cross-validation bandwidth selector it is very important to know if the aim is to estimate the regression function in a whole region. The bandwidth ($h = h_{cv}$) is chosen as it depends on minimizing the cross-validation function, i.e. finding the least squares.[12]

$$CV(h)=n^{-1}\sum_{j=1}^{n}(z_j - \widehat{m}_{h,j})^2 w(x_j) \qquad \dots\dots\dots\dots\dots\dots\dots\dots..(\,10\,)$$

Where as when estimating the function (m) each time it leaves one value (Leave one out) of(m) at x_j, meaning that the observation (z_j, x_j) is not used in the estimation procedure .

I.This global cross-validation bandwidth, h_{GCV} , is obtained by minimizing the function CV (h), using the weights ω (x_J) = 1, for every j.
II.The value h_{LCV} will be obtained as the minimizer of CV (h) with weight function
$W(x_j)=\prod_{t=1}^{n}\Phi(\frac{x_{j,t'}-x_t}{0.2\sigma})$

4. Comparative criteria

There are several criteria for comparison between prediction methods and for the same time series in order to assess the accuracy of prediction within the sample and from these standards (MSE, MAE, MAPE) and therefore the best method that can be adopted is the one that gives the least error among those criteria.[12]

- **Mean Square Error**

$$MSE=\frac{1}{k}\sum_{i=1}^{k}(\hat{m}_n(L)-z_{n+L})^2 \ldots\ldots\ldots\ldots\ldots(11)$$

- **Mean Absolute Error**

$$MAE=\frac{1}{k}\sum_{i=1}^{k}|\hat{m}_n(L)-z_{n+L}| \ldots\ldots\ldots\ldots\ldots(12)$$

- **Mean Absolute Percentage Error**

$$\text{MAPE}=\frac{1}{k}\sum_{i=1}^{k}|\frac{\hat{m}_n(L)-z_{n+L}}{z_{n+l}}|\ 100 \ldots\ldots\ldots\ldots\ldots(13)$$

5. A practical side

The practical side will include the application and analysis of the methods presented in the theoretical side to the final price time series data for oil in Iraq in US dollars The data provided by OPEC greatly facilitated the task of data analysis. The studied data was analyzed on the ready-made software package R. To estimate the appropriate model for the original oil price series using first some nonparametric methods such as the Kernel Smoothing method represented by Nadaraya Watson (NW) and local polynomials to different degrees and depending on some different methods of selecting the smoothing parameter, whereas the plug -in method, Smoothing Cross Validation method and the Least Squared Cross Validation method . Some of the accuracy criteria such as (MSE, MAE, MAPE) were calculated to make a comparison between the estimated models to choose the best estimated model . The time series has been drawn as shown in Figure (1), which represents the time series of monthly oil prices for the period studied. It is clear from the graph that the monthly oil price series in Iraq is a non-linear time series and contains clear fluctuations in its historical behavior, which suggests to it is an non stationary series in mean and variance.

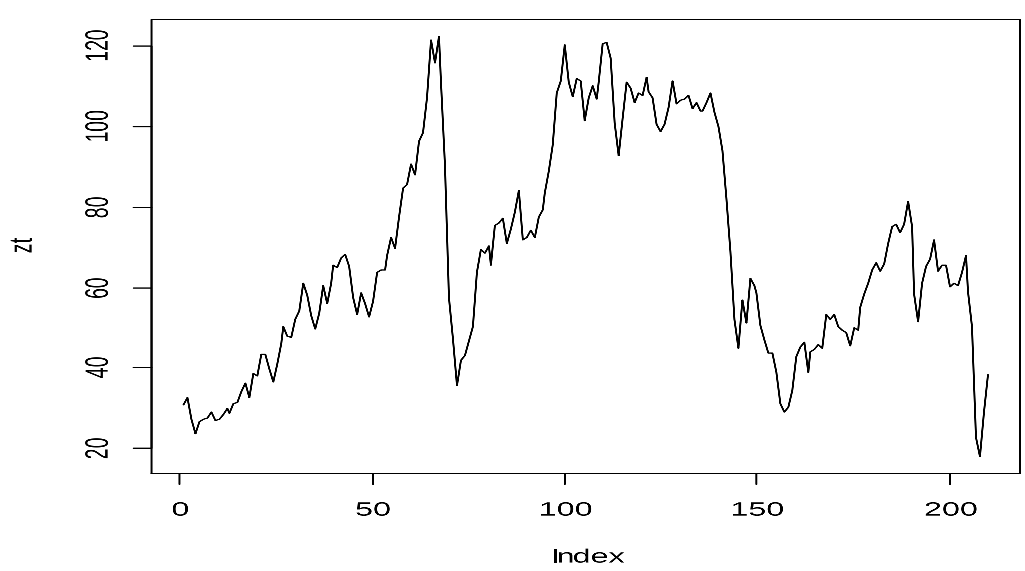


Figure (1) represents the time series of monthly oil prices for the period studied, represented by the period from January 2003 to June 2020.

The selection of autoregressive variables is one of the important steps in analyzing the time series and to determine the time-lagging variables affecting the variable in the current time. Observations of the time series (Z_t) are drawn with the autoregressive variables of the time lag with several time shifts starting from the displacement (t-1) , we show from the figure (2) that the relationship between Z_t and Z_{t-1} is a very strong linear relationship and so for the relationship between Z_t and Z_{t-2},Z_{t-3},Z_{t-4},Z_{t-5} , but when drawing the relationship between the time series (Z_t) with the self-regression variable with the time lag (t-6), that is, with the variable (Z_{t-6}), it was shown that there is a nonlinear relationship between them, as shown in Figure (3) which shows the shape of the propagation of the points other than Linear, therefore, we adopted this non-linear relationship between time series observations and time-delay autoregressive variables (t-6) as a basis for building a nonparametric prediction model for the time series and adopting it for the purposes of analysis and estimation by nonparametric methods presented in the theoretical side.

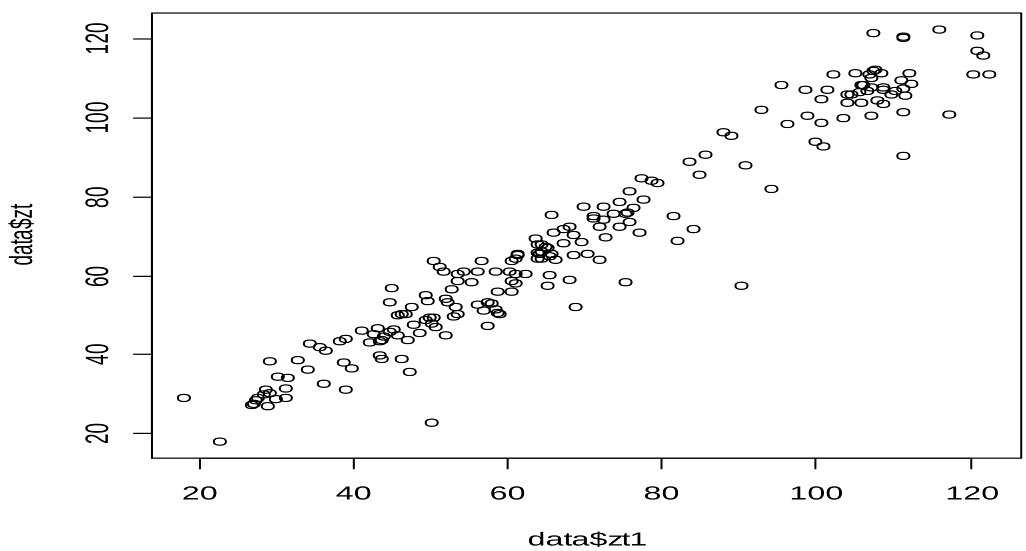


Figure (2) the time series (Z_t) with time-delay autoregressive variables (Z_{t-1}).

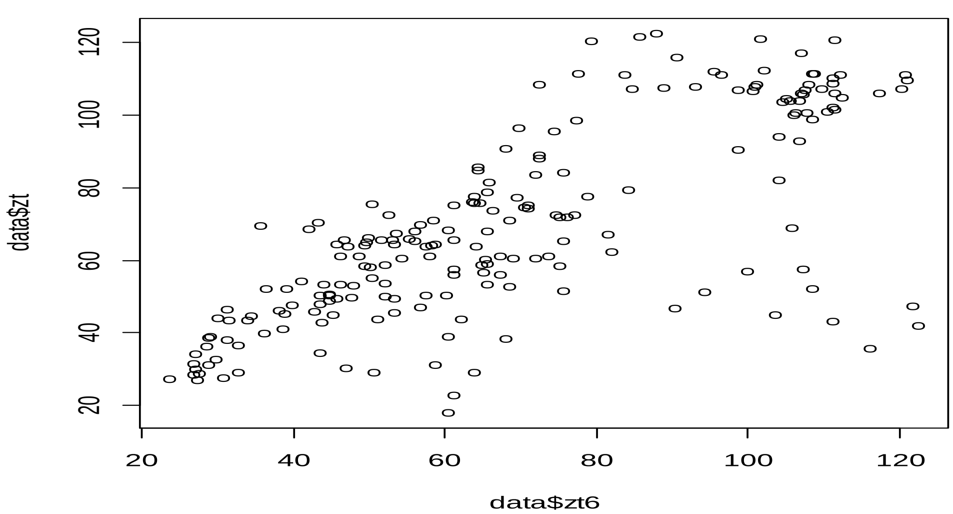


Figure (3) Observations of the time series (Z_t) with time-delayed auto-regressive variables (Z_{t-6}).

4. Selecting the smoothing parameter (SP)

4.1 Plug-in Method :

The Plug-in Method was used to select the smoothing parameter to estimate the function model using the Kernel methods represented by the Nadaraya Watson (NW) method, local linear polynomial (LL), quadratic polynomial (QU) and the cubic polynomial (CU) , the results of the precision criteria (MSE, MAE, MAPE) for estimation methods are shown in Table (1)

Table No. (1) shows the results of accuracy standards (MSE, MAE, MAPE) from the application of the smoother Nadaraya Watson (NW) and local polynomial in degrees (P = 1,2,3) by plug-in method.

Methods	SP	MSE	MAE	MAPE
NW	Plug-in h_1=6.187525	1003.6671	26.01818	0.4545813
LL	Plug-in h_1=6.187525	948.4967	25.02070	0.4361376
QU	Plug-in h_1=6.187525	951.2486	24.91605	0.4344807
CU	Plug-in h_1=6.187525	940.8777	24.70608	0.4310798

It is noticed that the estimator of the cubic polynomial (CU) was the best among the estimators that used the Plug-in Method because the values of the precision criteria (MSE, MAE, MAPE) were the lowest values

MSE = 940.8777 MAE = 24.70608 MAPE = 0.4310798

The reason is due to an increase in the smoothing due to an increase in the degree of the polynomial.

4.2 Cross-validation method:

The Smoothing Cross-Validation method (SCV) method was used to select the smoothing parameter to estimate the function model using the Kernel methods represented by the Nadaraya Watson (NW) method, local linear polynomial (LL), quadratic polynomial (QU) and the cubic polynomial (CU) , the results of the precision criteria (MSE, MAE, MAPE) for estimation methods are shown in Table (2)

Table (2) illustrates the results of the precision criteria (MSE, MAE, MAPE) by applying the graders of Nadaraya Watson (NW) and the local polynomial of grades(P = 1,2,3) using the (SCV) method

Methods	SP	MSE	MAE	MAPE
NW	SCV h_2=6.129657	1002.9760	26.00414	0.4543394
LL	SCV h_2=6.129657	948.4200	25.01575	0.4360795
QU	SCV h_2=6.129657	950.8190	24.90906	0.4343960
CU	SCV h_2=6.129657	939.9207	24.69279	0.4308521

we note from the results in Table (3) the following :

It is noticed that the estimator of the cubic polynomial (CU) was the best among the estimators that used the smoothing cross-validation method (SCV) h2 because the values of the precision criteria (MSE, MAE, MAPE) were the lowest values

MSE = 939.9207 MAE = 24.69279 MAPE = 0.4308521

The reason is due to the increase in the degree of the polynomial.

4.3 The method of least squares cross-validation:

The least squares cross-validation method (LSCV)h3 was used to select the smoothing parameter (SP), whose value was (h3 = 3.747988) to estimate the function model using the Kernel core methods represented by the Nadaraya Watson method (NW) ,local linear polynomial (LL) , quadratic polynomial (QU) and the cubic polynomial (CU) where the results were the precision criteria (MSE, MAE, MAPE) for the estimation methods as in Table (3)

Table No. (3) illustrates the results of the precision criteria (MSE, MAE, MAPE) by applying the graders of Nadaraya Watson (NW) and the local polynomial of grades (P = 1,2,3) using the (LSCV) method.

Methods	SP	MSE	MAE	MAPE

NW	LSCV h_3=3.747988	965.3520	25.31080	0.4435635
LL	LSCV h_3=3.747988	939.2629	24.73121	0.428840
QU	LSCV h_3=3.747988	911.4567	24.31351	0.4249746
CU	LSCV h_3=3.747988	884.2284	23.86001	0.4125501

By noting the results that were reached in Table (3) from using the Least Squares cross-validation method (L SCV) h3 to choose the smoothing parameter, we note the following

It is noted that the estimator of the cubic polynomial (CU) was the best among the estimators that used the Least Squares cross-validation (LSCV) method because the values of the precision criteria (MSE, MAE, MAPE) were the lowest values where they were equal.

MSE = 884.2284 MAE = 23.86001 MAPE = 0.4125501

The reason is due to the increase in the degree of the polynomial.

Through the results of Tables No. (1), Table No. (2) and Table No. (3), we conclude that the estimator of the cubic polynomial (CU) was the best smoothing method using the Kernel function and for all methods of selecting the smoothing parameter because the values The comparison criteria were the lowest values, and this indicates that the best estimate depends mainly on the smoother first method in addition to the method of selecting the smoothing parameter.

The cubic polynomial calibrator (CU) can be shown using the smoothing parameter selection method (LSCV) as shown in Figure(4)

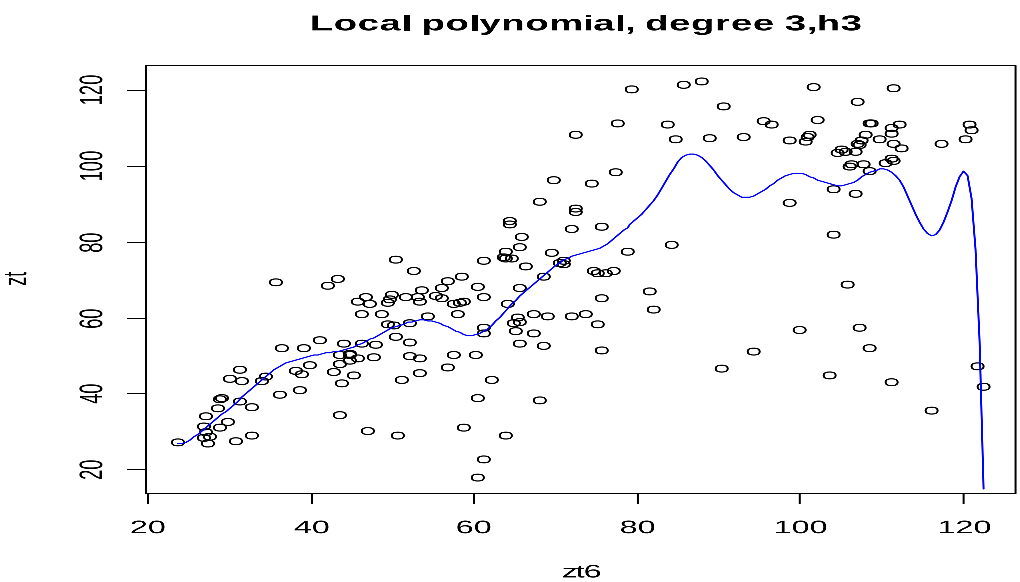


Figure (4) a cubic polynomial curve (CU) using the LSCV method for bandwidth selection.

6-Conclusions:

1.The Cubic Polynomial (CU) estimator was the best among the estimators that used the Plug-in method because the values of the precision criteria (MSE, MAE, MAPE) were the lowest.

2. The cubic polynomial regulator (CU) was superior to the other estimators that used the smoothing cross-validation method (SCV) h2 because the values of the precision criteria (MSE, MAE, MAPE) were the lowest.

3 The cubic polynomial (CU) estimator was the best one that used the Least Squares cross-validation(LSCV) method because the values of the precision criteria (MSE, MAE, MAPE) were the lowest.

4. Through the results of paragraphs 1,2 and 3, we conclude that the cubic polynomial estimator (CU) was the best method of smoothing using the Kernel function and for all methods of selecting the preamble parameter because the values of the precision criteria (MSE, MAE, MAPE) were the lowest values and this indicates that the best estimation using The kernel function depends mainly on the bootstrap method in addition to the smoothing parameter selection method.

References:

1- Bongard, J. (1960). Some Remarks on Moving Averages. In: O.E.C.D. (editor),Seasonal Adjustment on Electronic Computers. Proceedings of an international conference held in Paris, 361-387.
2- Fan, J. and GIJBELS, I. (1996). Local Polynomial Modelling and Applications. Chapman & Hall, London.
3-Fan,J.,and Y., Qiwei.(2003) .Nonlinear Time Series Nonparametric and Parametric Methods, "Springer, New York,USA".
4- Fisher, A. (1937). A Brief Note on Seasonal Variations. Journal of Accountancy, 64, 174.
5-Hardle, W., Chen, R. (1995). Nonparametric Time Series Analysis, a selective Review with examples. In Proceedings of the 50th Session of the ISI CONFERENCE,1,375--394
6-Hardle,W.,Lutkepohl ,H and Chen, R. (1997). A Review of Nonparametric Time Series Analysis. International Statistical Review, 65, 49-72.
7-Heiler,S.(1999). A Survey on Nonparametric Time Series Analysis . Discussion Paper, No. 99/05,Provided in Cooperation with University of Konstanz, Center of Finance and Econometrics (CoFE).
8-Hart, J.D. (1996).Some Automated Methods of Smoothing Time-dependent Data. Journal of Nonparametric Statistics, 6, 115-142.
9- Jones, H.L. (1943). Fitting of Polynomial Trends to Seasonal Data by the Method of Least Squares. Journal Amer. Statist. Assoc., 38, 453
10- MAacaulay, R.R. (1931). The smoothing of time series. Natinal Bureau of Economic Research, New York.
11- Stone, C.J. (1977). Consistent Nonparametric Regression (with discussion). Annals of Statistics, 5, 595{620.
12-Vilar-Fernández JM, Cao R. (2007). Nonparametric forecasting in time series: a comparative study.

Communication sin Statistics: Simulation and Computation **36**: 311–334