AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Better Quantile Contour for Constructing Growth Charts

Dr. Marwa K. Ibrahim	Prof. Dr. Qutaiba N. Nayef Al-Qazaz
marwakhalil.202024@yahoo.com	dr.qutaiba@coadec.uobaghdad.edu.iq
Ministry of Trade, Baghdad, Iraq	Department of statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq

Article Information

Article History:
Received: February, 20, 2024
Accepted: April, 12, 2024
Available Online: 31, December, 2024

Keywords:
Quantile regression; Nadaraya-Watson (NW); B-Splines (BS); Smoothing Spline (SS); time-varying coefficient; depth; Mean Absolute Error (MAE); contour shape and bivariate growth charts.

Correspondence:
Marwa K. Ibrahim
marwakhalil.202024@yahoo.com

Abstract

In this paper we going to comparison using Mean Absolute Error (MAE), among three nonparametric methods: Nadaraya-Watson (NW), B-Splines (BS) and Smoothing Spline (SS) which use to estimate quantile functions depending on the model of time-varying coefficient under the concept of depth in the bivariate case, to prepare for construct a better contours.

We used simulation study for construct two stages: first stage is to construct the simulation models for the three nonparametric methods and then comparison among them by using MAE, where the method that gives the least MAE is the best method, and the second stage is to plot the best method which given the best contour shape.

Where it was concluded that the BS is the best method because it's given a suitable contouring shape and less MAE which equal to 0.048 which will be used to plot best bivariate growth charts.

DOI: https://doi.org/10.55562/jrucs.v56i1.33

Introduction

They regression analysis, it's a technique which has become widely used and applied in many areas of academic and applied sciences such as social sciences, medical researches, biology, meteorology, psychology, chemistry and economics. So, depending on the regression, a number of researchers have resorted to using the quantile regression models to constructing the growth charts in univariate case. So the researcher decided to develop the quantile regression from univariate to multivariate used combining the two concepts (time-varying coefficient) and (depth) by construct a new model (time-varying coefficient with depth) which depending on a proposed generalization equation to find quantile functions, then use nonparametric methods (Nadaraya-Watson (NW), B-Splines (BS) with nodes (0.5, 1, 1.5, 2 and 2.5) and Smoothing Spline (SS)) to estimate those functions and find the best method. But first we have to explain why a model was used the model of time-varying coefficient with depth?

To rank the data points in the high dimension (multidimensional space), it used depth to benefit for contours is to provide a reasonable concept of quantile contours, and it was added with that concept the concept of the time-varying coefficient which can transform the linear regression or non-linear parametric regression to linear regression since it describes the effect of variables as a constant correlation coefficient. Hence the final model used of multivariate quantile regression depending on the model of time-varying coefficient with depth.

In this paper, we applied in the bivariate case in order to obtain the contours and thus obtain the bivariate growth charts because the difficulty of generalization and programming as well as for compatibility between methods when using nonparametric methods and applied under condition dimensional for each of the models.

Theoretical and Dynamic side

First, we have to review the two stratified quantile regression models which will be to study[4]:

The marginal model $\mathcal{Q}_{\mathrm{Y}_1}(\mathrm{q})$ of $\mathrm{Y}_{\mathrm{t},1} = \mathrm{y}_1$, where:

$$\mathcal{Q}_q\big(Y_{i,j,1}\big) = \mathcal{g}_1\big(q; t_{i,j}, \boldsymbol{X}_{i,j}\big) \tag{1}$$

And

The conditional models $\mathcal{Q}_{Y_2}(q)$ of $Y_{t,2}$ given $Y_{t,1}$, where:

$$\mathcal{Q}_q\big(Y_{i,j,2}\big) = \mathcal{g}_2\big(q; t_{i,j}, \boldsymbol{X}_{i,j}, Y_{i,j,1}\big) \tag{2}$$

Then, we write the models as:

$$\mathcal{Q}_q\big(Y_{i,j,1}\big) = \alpha_1\big(q; t_{i,j}\big) + \sum_{k=1}^{d} \gamma_{1k}\big(q; t_{i,j}\big)\, \boldsymbol{X}_{i,j,k} \tag{3}$$

Where

$$\widehat{\mathcal{g}}_{n1}(q) = argmin \sum_{i,j} \rho_q \left(Y_{i,j,1} - \mathcal{g}_1\big(t_{i,j}, \boldsymbol{X}_{i,j}\big)\right) \tag{4}$$

And

$$\mathcal{Q}_q\big(Y_{i,j,2}\big) = \alpha_2\big(q; t_{i,j}\big) + \beta_2\big(q; t_{i,j}\big) Y_{i,j,1} \quad + \sum_{k=1}^{d} \gamma_{2k}\big(q; t_{i,j}\big)\, \boldsymbol{X}_{i,j,k} \tag{5}$$

Where

$$\widehat{\mathcal{g}}_{n2}(q) = argmin \sum_{i,j} \rho_q \left(Y_{i,j,2} - \mathcal{g}_2\big(t_{i,j}, \boldsymbol{X}_{i,j}, Y_{i,j,1}\big)\right) \tag{6}$$

Then the researcher proposed equation for $\boldsymbol{Y} \in \mathbb{R}^p$ can be written as the general model for time-varying coefficient models as the following:

$$\mathcal{Q}_q\big(Y_{i,j,p}\big) = \sum_{L=1}^{p} \boldsymbol{\beta}\big(q; t_{i,j}\big) Y_{i,j,p-1} + \sum_{k=1}^{d} \gamma_{pk}\big(q; t_{i,j}\big)\, \boldsymbol{X}_{i,j,k} \tag{7}$$

Where

$\boldsymbol{\beta} = \big(\beta_{p1}, \beta_{p2}, \dots, \beta_{pp}\big)$, and $L = 1, \dots, p$, if $L = 1$ that means the marginal model and if $L \geq 2$ that means the conditional model, the researcher sought to build a comprehensive equation. Note that, the solution of the conditional sequence became more difficult in high dimensions.

We will used combining the two concepts (time-varying coefficient and depth) to construct a new concept (time-varying coefficient with depth) but first we must study the regression case which is the second definition of Koenker-Bassett methods when the linear regression is an extension for covariates in dimension $d \geq 1$, so in this method, pairs will be dealt with which is denoted as $(\boldsymbol{X}^{\intercal}, \boldsymbol{Y}^{\intercal})$ *where* $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{R}^d \times \mathbb{R}^p$.

Let have the random vector of covariates $\boldsymbol{X} := (\boldsymbol{X}_1, \dots, \boldsymbol{X}_d)^{\intercal}$

And, let $\boldsymbol{Y} := (Y_1, \dots, Y_p)^{\top} \in \mathbb{R}^p$ and the marginal quantiles $\mathfrak{q}_{q,j}$, $j = 1,2, \dots, p$ where are obtained by projecting $\boldsymbol{P}$ which characterized through the p marginal distribution functions on, then $\mathbf{q}_{q_1,\dots,q_p} := \left(\mathfrak{q}_{q_1,1}, \dots, \mathfrak{q}_{q_p,p}\right)^{\top}$

Where it's coordinate wise multivariate quantile. And if studying the empirical case for any observed it can be written the empirical marginal distributions $\boldsymbol{P}_n$ for n as $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$.
Note that, this projecting $\boldsymbol{P}$ be on p mutually orthogonal straight lines, where it's distinguished by canonical orthonormal basis $u_1, \dots, u_p$ direct by many origins. Where to facilitate the process to see the projections over all unit vectors where $\boldsymbol{u} \in S^{p-1}$ as previously mentioned, are direct by many origins.

For each direction $\boldsymbol{u}$ it can be got a good quantile $\mathfrak{q}_{qu}$ of order $q \in (0,1)$ from the univariate distribution of $\boldsymbol{u}^{\top}\boldsymbol{Y}$.
The directional q −quantile hyperplane $\boldsymbol{\Pi}_{qu}$, where $(a, \boldsymbol{b}^{\top}, \boldsymbol{\beta}^{\top})^{\top} \in \mathbb{R}^{d+p}$ become as the equation:

$$\boldsymbol{u}^{\top}\boldsymbol{y} = \boldsymbol{b}_{qu}^{\top}\boldsymbol{\Gamma}_{\boldsymbol{u}}^{\top}\boldsymbol{y} + \boldsymbol{\beta}_{q\boldsymbol{u}}^{\top}\boldsymbol{x} + a_{q\boldsymbol{u}} \tag{8}$$

Then the estimates of $\mathfrak{q}$ or written as $\left(a_{q\boldsymbol{u}}, \boldsymbol{b}_{q\boldsymbol{u}}^{\top}, \boldsymbol{\beta}_{q\boldsymbol{u}}^{\top}\right)^{\top}$ can be obtained by minimizing the quantile regression of the objective function $\mathfrak{q}$ as:

$$\left(a_{q\boldsymbol{u}}, \boldsymbol{b}_{q\boldsymbol{u}}^{\top}, \boldsymbol{\beta}_{q\boldsymbol{u}}^{\top}\right)^{\top} = \underset{(a,\boldsymbol{b}^{\top},\boldsymbol{\beta}^{\top})^{\top} \in \mathbb{R}^{d+p}}{argmin} E\left[\rho_q(\boldsymbol{Y}_{\boldsymbol{u}} - \boldsymbol{b}^{\top}\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} - \boldsymbol{\beta}^{\top}\boldsymbol{X} - a)\right] \tag{9}$$

And for the empirical distribution the minimizing become as:

$$\left(a_{q\boldsymbol{u}}^{(n)}, \boldsymbol{b}_{q\boldsymbol{u}}^{(n)^{\top}}, \boldsymbol{\beta}_{q\boldsymbol{u}}^{(n)^{\top}}\right)^{\top} = \underset{(a,\boldsymbol{b}^{\top},\boldsymbol{\beta}^{\top})^{\top} \in \mathbb{R}^{d+p}}{argmin} \sum_{i=1}^{n} \rho_q\left(\boldsymbol{Y}_{\boldsymbol{i},\boldsymbol{u}} - \boldsymbol{b}^{\top}\boldsymbol{Y}_{\boldsymbol{i},\boldsymbol{u}}^{\perp} - \boldsymbol{\beta}^{\top}\boldsymbol{X} - a\right) \tag{10}$$

The empirical distribution for the directional q −quantile hyperplane $\boldsymbol{\Pi}_{qu}$, where $(a, \boldsymbol{b}^{\top}, \boldsymbol{\beta}^{\top})^{\top} \in \mathbb{R}^{d+p}$ become as the equation:

$$\boldsymbol{u}^{\top}\boldsymbol{y} = \boldsymbol{b}_{q\boldsymbol{u}}^{(n)\top}\boldsymbol{\Gamma}_{\boldsymbol{u}}^{\top}\boldsymbol{y} + \boldsymbol{\beta}_{q\boldsymbol{u}}^{(n)^{\top}}\boldsymbol{x} + a_{q\boldsymbol{u}}^{(n)} \tag{11}$$

Estimation by using quantile technique

We estimate the two quantile function by used three of nonparametric methods: kernel regression (Nadaraya-Watson), B-splines and Smoothing splines depending on the concept of time-varying coefficient with directional Koenker-Bassett quantile regression. Now, it will study the estimation of the directional q −quantile hyperplane $\boldsymbol{\Pi}_{qu}$, under the nonparametric case with the distribution $\boldsymbol{P}$ of $\boldsymbol{Y}$ *conditional on* $\boldsymbol{X} = x$, *where* $\boldsymbol{Y} := (Y_1, \dots, Y_p)^{\top} \in \mathbb{R}^p$ and $\boldsymbol{X} := (\boldsymbol{X}_1, \dots, \boldsymbol{X}_d)^{\top}$, which is possible to apply this concept only for $d + p \leq 3$ where $d = 1, p = 2$, the general model is

$$\mathcal{Q}_q(\boldsymbol{Y}_{i,j,p}) = q_p \tag{12}$$

The directional q −quantile hyperplane $\boldsymbol{\Pi}_{qu}$, under location case where $(a, \boldsymbol{b}^{\top}, \boldsymbol{\beta}^{\top})^{\top} \in \mathbb{R}^{d+p}$ *is:*

$$\mathcal{Q}_q(\boldsymbol{Y}_{i,j,p}) = q_p = \boldsymbol{u}^{\top}\boldsymbol{y} = \boldsymbol{b}_{q\boldsymbol{u}}^{\top}\boldsymbol{\Gamma}_{\boldsymbol{u}}^{\top}\boldsymbol{y} + a_{q\boldsymbol{u}} \tag{13}$$

The directional q −quantile hyperplane $\boldsymbol{\Pi}_{qu}$, under the nonparametric case where $(a, \boldsymbol{b}^{\top}, \boldsymbol{\beta}^{\top})^{\top} \in \mathbb{R}^{d+p}$ is:

$$\mathcal{Q}_q(\boldsymbol{Y}_p) = q_p = \boldsymbol{u}^{\top}\boldsymbol{y} = \boldsymbol{b}_{q\boldsymbol{u}}^{\top}\boldsymbol{\Gamma}_{\boldsymbol{u}}^{\top}\boldsymbol{y} + \boldsymbol{\beta}_{q\boldsymbol{u}}^{\top}\boldsymbol{x} + a_{q\boldsymbol{u}} \tag{14}$$

Then, the estimates of $\mathfrak{q}$ or written as $\left(a_{q\boldsymbol{u}}, \mathbf{b}_{q\boldsymbol{u}}^{\top}, \boldsymbol{\beta}_{q\boldsymbol{u}}^{\top}\right)^{\top}$ can be obtained by minimizing the quantile regression of the objective function $\mathfrak{q}$ as:

$$\left(a_{q\boldsymbol{u}}, \boldsymbol{b}_{q\boldsymbol{u}}^{\top}, \boldsymbol{\beta}_{q\boldsymbol{u}}^{\top}\right)^{\top} = \underset{(a,\boldsymbol{b}^{\top},\boldsymbol{\beta}^{\top})^{\top} \in \mathbb{R}^{d+p}}{argmin} E\left[\rho_q(\boldsymbol{Y}_{\boldsymbol{u}} - \boldsymbol{b}^{\top}\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} - \boldsymbol{\beta}^{\top}\boldsymbol{X} - a)\right] \tag{15}$$

Now, we will write the above equations under the concept of time-varying coefficient with directional Koenker-Bassett (depth) quantile regression

- First method: it will to using Nadaraya-Watson method [5], depending on the concept of directional Koenker-Bassett quantile regression we get the following equation [3]:

$$\hat{q}_{(NW-dir)}(q) = \underset{(a,\boldsymbol{b}^{\intercal},\boldsymbol{\beta}^{\intercal})^{\intercal}\in\mathbb{R}^{d+p}}{argmin} \sum \rho_q\left(\boldsymbol{Y} - (\boldsymbol{b}^{\intercal}\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} + \boldsymbol{\beta}^{\intercal}X + a)\right) K\left(\frac{t-T}{h}\right) \tag{16}$$

- Second method (proposal 1): it will to suggest B-spline in a linear space $\mathbb{G}_{\ell}$ of spline function on $\boldsymbol{T}$ with a fixed degree and with a Knot sequence depending on the concept of time-varying coefficient with directional Koenker-Bassett quantile regression as:

$$\hat{q}_{(BS-dir)M}(q) = argmin \sum_{i=1}^{n} \boldsymbol{W}_i \sum \left(\rho_q\left(\boldsymbol{Y} - (\boldsymbol{b}^{\intercal}\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} + \boldsymbol{\beta}^{\intercal}X + a)\right)\right)^2 \tag{17}$$

- Third method (proposal 2): it will to suggest smoothing spline depending on the concept of time-varying coefficient with directional Koenker-Bassett quantile regression [2] as:

$$\hat{q}_{(SS-dir)M}(q) = argmin \sum_{i,j} \rho_q\left(\boldsymbol{Y} - (\boldsymbol{b}^{\intercal}\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} + \boldsymbol{\beta}^{\intercal}X + a)\right)^2 + \lambda \|m''\|_2^2 \tag{18}$$

Simulation study

We used simulation study in two stages:

- First stage which represented by a comparison nonparametric methods (Nadaraya-Watson (NW), B-Splines (BS) with nodes (0.5, 1, 1.5, 2 and 2.5) and Smoothing Spline (SS)), which used to estimation the quantile functions for the models under the concept (time-varying coefficient with depth), then we chose the methods which gave least MAE in the different sample sizes (10, 30, 50), so it was excluded (NW) see the table below [4]:

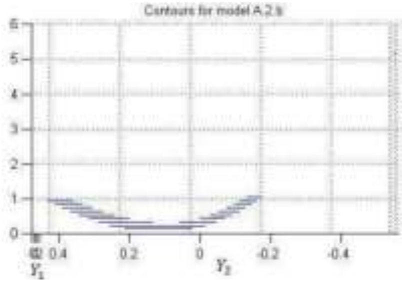
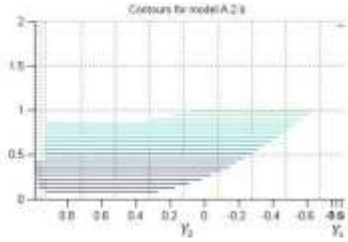
Table 1. Represent the MAE for first stage

MAE	NW (T & directional)		BS (T & directional)		SS (T & directional)		Min
n \ Q	$Q_q(Y_1)$	$Q_q(Y_2)$	$Q_q(Y_1)$	$Q_q(Y_2)$	$Q_q(Y_1)$	$Q_q(Y_2)$	
10	13.5007416	4.77627179	194.541118	8.08919566	5.59156263	2.96445351	2.96E+00
30	15.7056731	20.7030898	260.283264	197.869744	3.84958012	5.94649665	3.85E+00
50	3.12250646	21.0582218	0.28432887	820.911835	0.83553168	8.72855374	2.84E -01

- Second stage which represented to find the best contour shape from the best methods chosen from the first stage which gaves the best quantiles and with the least MAE).

It was noted that the best method is B-splines (BS) for the time-varying coefficient with directional Koenker-Bassett which achieved the lowest MAE, as shown in the table and figures below:

Table 2: Represent the MAE for first and second stages with degree of quantiles

Model	First stage	Second stage	Figure	
A.2.b SS	**4.84E-01**	**0.87**		**Given acceptable contours and good MAE.**
A.2.b BS	**0.28432887**	**0.048**		**Given good contours and good MAE because it's less MAE.**

Practical application

We will be used real data to plotting standard growth charts for children under the age of five years from Iraq Multiple Indicator Cluster Survey (MICS-4) 2011 for all governorates of Iraq and applied on the model that gave best method chosen in the Simulation study [1, 6].

For example: class (0-5 month) was taken as an applied example, see the illustrate figure below:

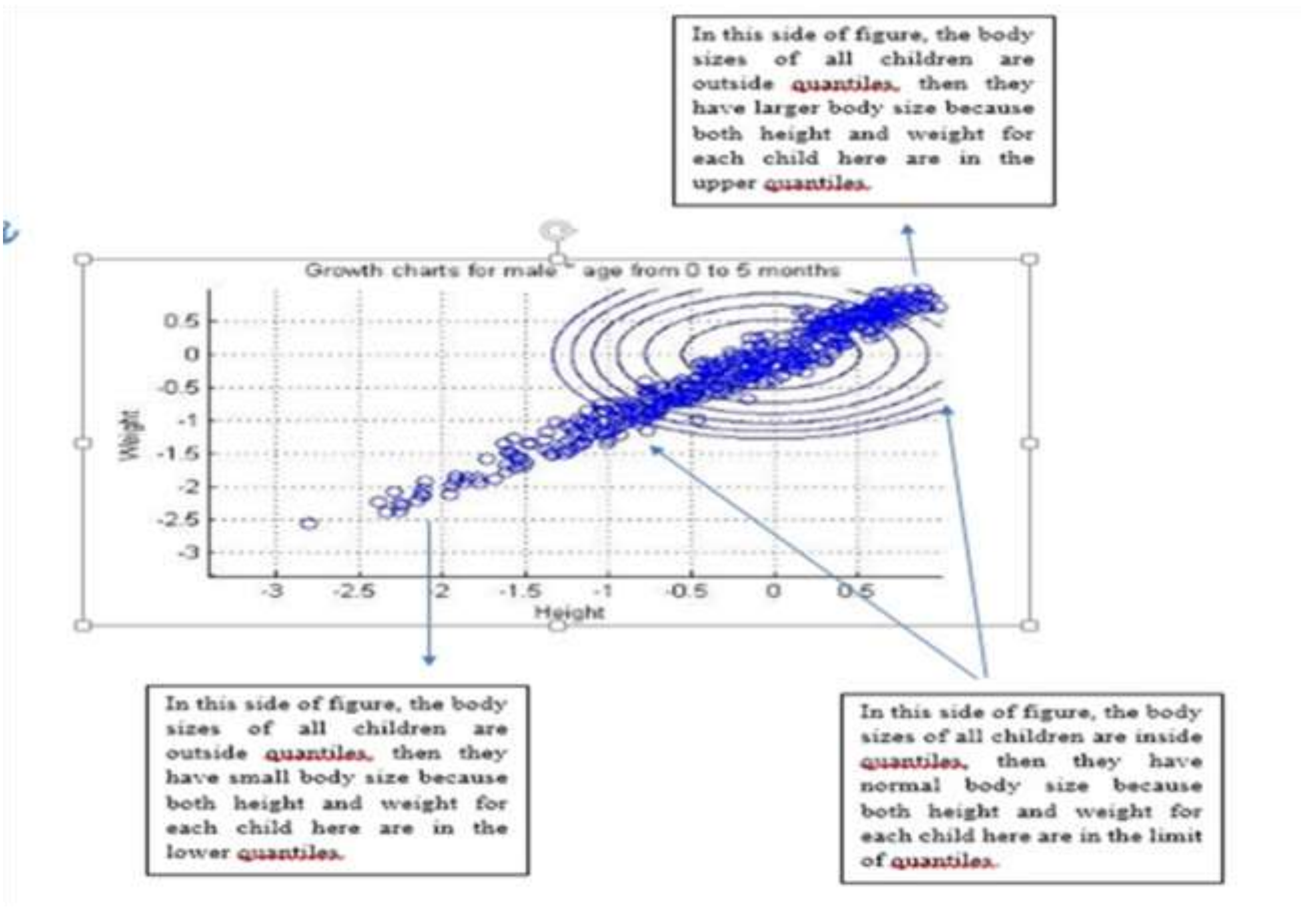


Figure 1: Represent the Analysis bivariate growth charts

Conclusion

in this paper, we discussed the bivariate case by used combining the two concepts (time-varying coefficient) and (depth) by construct a new model (time-varying coefficient with depth) which depending on a proposed generalization equation to find quantile functions, then use nonparametric methods (Nadaraya-Watson (NW), B-Splines (BS) with nodes (0.5, 1, 1.5, 2 and 2.5) and Smoothing Spline (SS)) to estimate those functions and find the best method. As it was found that model B gave less MAE = 0.048, note that we using the nonparametric method (B-spline). Then applied real data by using the best model, see the figure 1.

References

[1] Bocek P. and Siman M., "Directional Quantile Regression in Octave (and Matlab)". Kybernetika, Vol. 52, No. 1, pp. 28-51, 2016.

[2] Cox D. D., "Multivariate Smoothing Spline". Journal on Numerical Analysis, Vol. 21, No. 4, pp.789-813, 1984

[3] Härdle, W., Müller, M., "Multivariate and Semi parametric Kernel Regression". Working Paper, SFB 373 Discussion Paper, No. 26, 1997.

[4] Ibrahim, M. K., Al-Qazaz Q. N., "Comparison between the estimated of nonparametric methods by using the methodology of Quantile regression models". Periodicals of Engineering and Natural Sciences ISSN 2303-4521, Vol. 7, No. 2, 2019.

[5] Ley Ch. and Verdebout T., "Modern Directional Statistics". ISBN 9780367573010, Published by Chapman and Hall/CRC, 190 Pages, June 29, 2020.

[6] Liu X. and Zuo Y., "CompPD: A Matlab Package for Computing Projection Depth". Journal of Statistical Software, Vol. 65, Issue 2, May 2015.

[7] R., Koenker, V., Chernozhukov, X., He and L., Peng, "Handbook of Quantile Regression", Taylor & Francis Group, LLC, 2018.

.

Journal of AL-Rafidain University College for Sciences (2024); Issue 56 EISSN (2790-2293) PISSN (1681-6870)

مجلة كلية الرافدين الجامعة للعلوم (2024)؛ العدد 56؛ 373 -379

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

افضل تقسيمات كونتورية لبناء مخططات النمو

د. مروه خليل ابراهيم	أ. د. قتيبة نبيل نايف القزاز
marwakhalil.202024@yahoo.com	dr.qutaiba@coadec.uobaghdad.edu.iq
وزارة التجارة، بغداد، العراق	قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد، بغداد، العراق

المستخلص

تناول البحث اجراء مقارنة باستخدام متوسط الخطأ المطلق (MAE) بين ثلاث طرق لامعلمية وهي Nadaraya-Watson (NW)، B-Splines (BS) و Spline (SS) Smoothing والتي تستخدم لتقدير الدوال التقسيمية اعتمادًا على نموذج معامل متغير بمرور الوقت في ظل مفهوم العمق في حالة ثنائية المتغير، للتحضير لبناء أفضل شكل كونتوري.

استخدمنا دراسة المحاكاة لبناء مرحلتين: المرحلة الأولى هي بناء نماذج المحاكاة للطرق الثلاثة اللامعلمية ومن ثم المقارنة فيما بينها باستخدام متوسط الخطأ المطلق MAE، حيث أن الطريقة التي تعطي أقل MAE هي الطريقة الأفضل، والمرحلة الثانية هي رسم أفضل طريقة تعطي أفضل شكل كونتوري.

حيث تم التوصل إلى أن طريقة BS هي الطريقة الأفضل لأنها تعطي شكل كونتور مناسب وبأقل MAE والذي بلغت قيمته 0.048 والتي سيتم استخدامها لرسم أفضل مخططات النمو ثنائية المتغير.

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث:20/2/2024

تاريخ قبول البحث:12/4/2024

تاريخ رفع البحث على الموقع: 31/12/2024

الكلمات المفتاحية:

الانحدار التقسيمي؛ Nadaraya-Watson (NW)؛ B-Splines (BS) Smoothing Spline (SS)؛ معامل متغير بمرور الوقت؛ العمق؛ متوسط الخطأ المطلق، شكل كونتوري ومخططات النمو ثنائية المتغير.

للمراسلة:

د. مروه خليل ابراهيم

marwakhalil.202024@yahoo.com

DOI: https://doi.org/10.55562/jrucs.v56i1.33