BIU Journal of Basic and Applied Sciences 8(1): 59 – 69, 2023.

ISSN: 2563-6424

A MODIFIED PREDICTION ERROR SUM OF SQUARES CRITERION FOR BANDWIDTH SELECTION IN LOCAL QUADRATIC REGRESSION

EDIONWE, E.[1] AND OSEMWENKHA, S. O.[2]
[1]Department of Mathematical Sciences, Edwin Clark University, Kiagbodo, Delta State, Nigeria
[2]Department of Statistics, University of Benin, Benin City, Nigeria
*Corresponding Author: trustnelson24@yahoo.com

ABSTRACT

Bandwidth (smoothing parameter) is considered the most crucial parameter in the application of nonparametric regression model. For the purpose of selecting adaptive bandwidths for the Local Quadratic Regression (LQR) in the response surface settings, we present a modified Prediction Error Sum of Square (PRESS) criterion using a penalty term derived from the sum of the range of kernel weights at each of the data points. LQR is applied to a multiple response problem from the literature using the current PRESS criterion and the proposed version for optimal bandwidth selection. The proposed criterion gives comparatively better regression and optimization results than the current PRESS criterion. The Sum of Squared Error (SSE) and the Coefficient of Determination (R^2) (both of which indicate the degree of closeness of the fitted values of the response to the raw data) were used as the basis for comparing model performance and goodness-of-fit. In order to compare the version that meets the process specifications for each of the three responses simultaneously, the desirability measure (function) was applied. The results presented show that the proposed version of the PRESS criterion gives the smallest SSE (0.2127, 10.0027 and 65720 for y_1, y_2 and y_3, respectively) and the largest R^2 (99.2598, 97.2309 and 92.3804 for y_1, y_2 and y_3, respectively) across the three responses in the study. The proposed version gives a desirability measure of 69.0639% trumping that of the existing version which gives a desirability measure of 40.7450%. A desirability measure of 69.0639% indicates that the proposed version meets approximately 70% of the process specification.

KEYWORDS: *Desirability function, hat matrix, local quadratic regression, PRESS criterion, response surface study, process specification*

Introduction

Response surface methodology is a collection of mathematical and statistical tools for studying the relationship between two or more variables (Vivek *et al*., 2021). In the data modeling phase of RSM, it is assumed that the relationship between a response variable y and a k explanatory variables $x_{1,}\ x_2, \dots, x_k$, takes the form:

$$y_i = f\left(x_{i1,}\ x_{i2}, \dots, x_{ik}\right) + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n \quad (1)$$

where y_i is the output at the i^{th} data point, x_{ij}, $j = 1,2,\ldots,k$, is the value of the j^{th} explanatory variable at the i^{th} data point, f represents the true but unknown function that depicts the exact mathematical relationship between the variables, $\boldsymbol{\varepsilon}_i$ is a random error term assumed to be independent, identically distributed with mean zero and constant variance σ^2, and n is the sample size (Castillo, 2007; He *et al*., 2012; Rajewski and Dobrzynaska-Inger, 2021).

The traditional regression model for estimating f in (1) is the Ordinary Least Squares (OLS) (Myers, 1999; Myers et al., 2009). However, if the data consists of salient patterns and trends that might be overlooked by OLS, nonparametric regression models could provide better alternatives (Wan and Birch, 2011; Depalma *et al*., 2021; Ryu *et al*., 2021). Mathematically, LQR estimate $y_i^{(LQR)}$ of y_i takes the form:

$$\hat{y}_i^{(LQR)} = \boldsymbol{x}_i'(\boldsymbol{X}_q'\boldsymbol{W}_i\boldsymbol{X}_q)^{-1}\boldsymbol{X}_q'\boldsymbol{W}_i\boldsymbol{y}, \qquad (2)$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of response, $\boldsymbol{W_i}$ is an $n \times n$ diagonal matrix of weights for estimating y_i, $\boldsymbol{x}_i'$ is the i^{th} row vector of the LQR model matrix $\boldsymbol{X_q}$, $\boldsymbol{X}_q'$ is the transposed LQR model matrix whose general form is given by:

$$\boldsymbol{X_q} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} & x_{11}^2 & x_{12}^2 & \cdots & x_{1k}^2 \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} & x_{21}^2 & x_{22}^2 & \cdots & x_{2k}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} & x_{n1}^2 & x_{n2}^2 & \ldots & x_{nk}^2 \end{bmatrix},$$

In matrix form, the vector of LQR estimates presented in (2) is expressed as:

$$\begin{bmatrix} \hat{\boldsymbol{y}}_1^{(LQR)} \\ \hat{\boldsymbol{y}}_2^{(LQR)} \\ \vdots \\ \hat{\boldsymbol{y}}_n^{(LQR)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1(\boldsymbol{X}_q'\boldsymbol{W}_1\boldsymbol{X}_q)^{-1}\boldsymbol{X}_q'\boldsymbol{W}_1 \\ \boldsymbol{x}_2(\boldsymbol{X}_q'\boldsymbol{W}_2\boldsymbol{X}_q)^{-1}\boldsymbol{X}_q'\boldsymbol{W}_2 \\ \vdots \\ \boldsymbol{x}_n(\boldsymbol{X}_q'\boldsymbol{W}_n\boldsymbol{X}_q)^{-1}\boldsymbol{X}_q'\boldsymbol{W}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_n \end{bmatrix}, \qquad (3)$$

$$= \begin{bmatrix} \boldsymbol{h}_1^{(LQR)'}(\boldsymbol{b}) \\ \boldsymbol{h}_2^{(LQR)'}(\boldsymbol{b}) \\ \vdots \\ \boldsymbol{h}_n^{(LQR)'}(\boldsymbol{b}) \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_n \end{bmatrix}, \qquad (4)$$

$$= \boldsymbol{H}^{(LQR)}(\boldsymbol{b})\boldsymbol{y}, \qquad (5)$$

where $\boldsymbol{h}_i^{(LQR)'}(\boldsymbol{b}) = \left(\boldsymbol{h}_{i1}^{(LQR)}\boldsymbol{h}_{i2}^{(LQR)} \ldots \boldsymbol{h}_{in}^{(LQR)}\right)$ is the i^{th} row vector of the $n \times n$ LQR Hat matrix, $\boldsymbol{H}^{(LQR)}(\boldsymbol{b})$.

If all the quadratic terms in the LQR model matrix $\boldsymbol{X_q}$ are deleted, LQR reduces to local linear regression (Anderson-Cook and Prewitt, 2005; Eguasa et al., 2022).

The rth entry, say w_{rr} of the weight matrix $\boldsymbol{W_o}$ for estimating y_o is obtained from the product kernel given as:

$$w_{or} = \prod_{j=1}^{k} K\left(\frac{x_{oj}-x_{rj}}{b_r}\right) \Big/ \sum_{i=1}^{n} \prod_{j=1}^{k} K\left(\frac{x_{oj}-x_{ij}}{b_i}\right), \quad i = 1,2,\ldots,n, \qquad (6)$$

where $K\left(\frac{x_{oj}-x_{rj}}{b_r}\right) = e^{-\left(\frac{x_{oj}-x_{ij}}{b_r}\right)^2}$ is the simplified Gaussian function which assigns relatively heavier weights to the observations close to x_{oj} than those far from x_{oj}, and b_i, $i = 1,2,\ldots,n$, are referred to as the local or locally adaptive bandwidths which reduce to a fixed or global bandwidth b in situations where we have $b_1 = b_2 \ldots = b_n = b$ (Edionwe and Mbegbu, 2014).

In nonparametric regression procedure, the values, x_{ij} of the explanatory variables are transformed such that $0 \leq x_{ij} \leq 1$, and consequently, $b_i, i = 1,2, \ldots, n$, are constrained to lie in the interval $0 < b_i \leq 1$ (Edionwe *et al*., 2016).

A model for selecting local bandwidths presented in Edionwe et al. (2018) is given by:

$$b_i = \frac{N(TC - y_i)}{T(Cn-1)}, \quad (7)$$

where $C \geq 0$ and $N > 0$ are data-driven tuning parameters and $T = \sum_i^n y_i$.

For small-sample studies such as RSM, the PRESS** criterion developed by Mays et al (2001) for selecting bandwidths in the application of both nonparametric and semiparametric regression models is given by:

$$PRESS^{**}(\boldsymbol{b}) = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{i,-i}^{(LQR)}(\boldsymbol{b})\right)^2}{\left[n - tr\left(\boldsymbol{H}^{(LQR)}(\boldsymbol{b})\right)\right] + \left[(n-k-1)\left(\frac{SSE_{max} - SSE(\boldsymbol{b})}{SSE_{max}}\right)\right]},$$

$$= \frac{PRESS}{DF + \left[(n-k-1)\left(\frac{SSE_{max} - SSE(\boldsymbol{b})}{SSE_{max}}\right)\right]}, \quad (8)$$

where $SSE_{max} = \sum_{i=1}^{n}\left(y_i - \hat{y}_i^{(LQR)}(\boldsymbol{b})\right)^2$ is the maximum Sum of Squared Errors obtained as b tends to infinity, $SSE(\boldsymbol{b})$ is the Sum of Squared Errors for a given vector of bandwidths, $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$, Degree of Freedom (DF)= $n - tr\{\boldsymbol{H}^{(LQR)}(\boldsymbol{b})\}$, $tr\{\boldsymbol{H}^{(LQR)}(\boldsymbol{b})\}$ is the sum of the diagonal elements of the LQR Hat matrix for a given vector of bandwidths, $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$, and $\hat{y}_{i,-i}^{(LQR)}(\boldsymbol{b})$ is the leave-one-out estimate of y_i with the i^{th} observation left out.

In applying (7) to generate bandwidths for a given data, we search for the optimal values, $C^*$and N^*, of C and N, respectively, that give the optimal local bandwidths for minimizing the $PRESS^{**}(\boldsymbol{b})$ criterion.

The phase succeeding the data modelling phase in RSM is the optimization phase, where the setting of the explanatory variables that optimize the fitted regression model according to the process specifications (or production requirements) is sought.

In studies that involve say m responses, $m > 1$, the goal is to obtain the setting of the explanatory variables which simultaneously optimize the m fitted models with respect to their individual process specifications (Harrington, 1965; Derringer and Suich, 1980). A few criteria for carrying out multiple response optimization exist amongst which the desirability measure (function) stands out. The desirability function, with respect to the process specification of individual response, transforms the fitted model, $\hat{y}_p(\boldsymbol{x})$, into a scalar measure, $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$, $p = 1,2, \ldots, m$, after which the setting of each of the explanatory variables that maximize the geometric mean of the m transformed scalar measures is subsequently sought (Wan and Birch, 2011).

The classifications of $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$ based on the process specification of the responses is as follows:

If the response is of nominal-the-better (NTB) type where the p^{th} response acceptable value lies between an upper limit, U and a lower limit, L, $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$ is given as:

$$d_p\left(\hat{y}_p(\boldsymbol{x})\right) = \begin{cases} 0 & \hat{y}_p(\boldsymbol{x}) < L \\ \left\{\frac{\hat{y}_p(x)-L}{\emptyset-L}\right\} & L \le \hat{y}_p(\boldsymbol{x}) < \emptyset, \\ \left\{\frac{U-\hat{y}_p(x)}{U-\emptyset}\right\} & \emptyset \le \hat{y}_p(\boldsymbol{x}) \le U, \\ 0 & \hat{y}_p(\boldsymbol{x}) > U, \end{cases} \tag{9}$$

where $\emptyset$ is the target value of the p^{th} response.

If the goal is to maximize the p^{th} response, $d_p\left(\hat{y}_p(\boldsymbol{x})\right)$ is given by a one-sided transformation as:

$$d_p\left(\hat{y}_p(\boldsymbol{x})\right) = \begin{cases} 0 & \hat{y}_p(\boldsymbol{x}) < L, \\ \left\{\frac{\hat{y}_p(x)-L}{\emptyset-L}\right\} & L \le \hat{y}_p(\boldsymbol{x}) \le \emptyset, \\ 1 & \hat{y}_p(\boldsymbol{x}) > \emptyset, \end{cases} \tag{10}$$

where $\emptyset$ is interpreted as a large enough value of the p^{th} response.

If the goal is to minimize the p^{th} response,$d_p\left(\hat{y}_p(\boldsymbol{x})\right)$ is given by a one-sided transformation as:

$$d_p\left(\hat{y}_p(\boldsymbol{x})\right) = \begin{cases} 1 & \hat{y}_p(\boldsymbol{x}) < \emptyset, \\ \left\{\frac{U-\hat{y}_p(x)}{U-\emptyset}\right\} & \emptyset \le \hat{y}_p(\boldsymbol{x}) \le U, \\ 0 & \hat{y}_p(\boldsymbol{x}) > U, \end{cases} \tag{11}$$

where $\emptyset$ is a small enough value of the p^{th} response.

The overall objective of the desirability criterion is getting the values of the explanatory variables that maximize the geometric mean (D) of all the individual desirability measures given as:

$$D = maximize\left(\left(\prod_{p=1}^{m} d_p\left(\hat{y}_p(\boldsymbol{x})\right)\right)^{1/m}\right) \times 100\%, \tag{12}$$

The remainder of this paper is organized as follows: A review of the $PRESS^{**}(\boldsymbol{b})$ criterion is given in Section 2 and a modified version is presented in Section 3. Section 4 presents comparison of results from the application of the proposed criterion with that of the $PRESS^{**}(\boldsymbol{b})$ criterion. The paper concludes in Section 5.

The Penalizing Factor of the $PRESS^{**}(\boldsymbol{b})$ vis-à-vis the Flexibility of the LQR Hat Matrix

In general, the $n \times n$ diagonal matrix weights $\boldsymbol{W_o}$ derived from (6) for estimating y_o can be expressed as:

$$\boldsymbol{W_o} = \begin{bmatrix} w_{01} & 0 & \cdots & 0 \\ 0 & w_{02} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & w_{0n} \end{bmatrix},$$

$$= \begin{bmatrix} \left(\frac{\prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{1j}}{b_1}\right)^2}}{\sum_{i=1}^{n}\prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{ij}}{b_i}\right)^2}}\right) & 0 & \cdots & 0 \\ 0 & \left(\frac{\prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{2j}}{b_2}\right)^2}}{\sum_{i=1}^{n}\prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{ij}}{b_i}\right)^2}}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\frac{\prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{nj}}{b_n}\right)^2}}{\sum_{i=1}^{n}\prod_{j=1}^{k} e^{-\left(\frac{x_{oj}-x_{ij}}{b_i}\right)^2}}\right) \end{bmatrix},$$

According to Fan and Gijbels (1996), Mays and Birch (1998), the advantage of LQR over OLS is its flexibility which is a function of the choice of bandwidths that are selected for the procedure.

In situations where $b_1, b_2, \ldots, b_n$ all tend to 1 and above in (3), the elements of $\boldsymbol{W_1}$, $\boldsymbol{W_2}, \ldots, \boldsymbol{W_n}$ which respectively form part of the row vectors $\boldsymbol{h_1^{(LQR)'}}(\boldsymbol{b}), \boldsymbol{h_2^{(LQR)'}}(\boldsymbol{b}), \ldots, \boldsymbol{h_n^{(LQR)'}}(\boldsymbol{b})$, all would return the same value, say w. That is $w_{11} = w_{22} = \cdots = w_{nn} = w$, and consequently, $\boldsymbol{W_1} = \boldsymbol{W_2} = \cdots = \boldsymbol{W_n}$. In this case, assuming that the same model matrix $\boldsymbol{X_q}$ is used for OLS procedure, we will get $\widehat{\boldsymbol{y}}_{\boldsymbol{1}}^{(\boldsymbol{LQR})} = \widehat{\boldsymbol{y}}_{\boldsymbol{1}}^{(\boldsymbol{OLS})}$, $\widehat{\boldsymbol{y}}_{\boldsymbol{2}}^{(\boldsymbol{LQR})} = \widehat{\boldsymbol{y}}_{\boldsymbol{2}}^{(\boldsymbol{OLS})}, \ldots,$ $\widehat{\boldsymbol{y}}_{\boldsymbol{n}}^{(\boldsymbol{LQR})} = \widehat{\boldsymbol{y}}_{\boldsymbol{n}}^{(\boldsymbol{OLS})}$, meaning that LQR returns the same vector of estimated responses as that of the OLS. On the other hand, the more distinct or dissimilar the elements of $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, … , $\boldsymbol{W_n}$ in the vectors $\boldsymbol{h_1^{(LQR)'}}(\boldsymbol{b}), \boldsymbol{h_2^{(LQR)'}}(\boldsymbol{b}), \ldots, \boldsymbol{h_n^{(LQR)'}}(\boldsymbol{b})$, respectively, the higher the flexible of the resulting LQR over that of the OLS.

The advantage of nonparametric regression models over their parametric counterpart is flexibility and, one of the ways to upgrade the flexibility of LQR is to ensure that the bandwidths selected allow the $1 \times n$ elements of each of the vectors $\boldsymbol{h_i^{(LQR)'}}(\boldsymbol{b}) = \left(\boldsymbol{h_{i1}^{(LQR)}}\boldsymbol{h_{i2}^{(LQR)}} \ldots \boldsymbol{h_{in}^{(LQR)}}\right)$, $i = 1,2, \ldots, n$, to be as distinct from one another as possible. Distinctiveness implies variability. Statistically, the range is directly proportional to variability. Hence, the range of the n elements of each of the rows of the Hat matrix should be as high as possible.

We highlight two shortcomings of the DF as it relates to its computation in the $PRESS^{**}(\boldsymbol{b})$ criterion:

One, the DF is not a function of the variability (or the range) of the elements of the row vectors of the LQR Hat matrix and so does not in any way enhance the flexibility of the LQR.

Two, $DF = n - tr\{\boldsymbol{H}^{(LQR)}(\boldsymbol{b})\}$, where $tr\{\boldsymbol{H}^{(LQR)}(\boldsymbol{b})\}$ is the sum of the diagonal

elements of the LQR Hat matrix, makes use of only n of the n times $n = n^2$ elements of the Hat matrix, neglecting the remaining $n(n-1) = n^2 - n$ elements. This negligence has negative consequences on the regression procedure since the unused $n^2 - n$ elements contain important information about the data under study.
Thus, there is a need for the inclusion of an appropriate statistic in the penalizing factor of the $PRESS^{**}(\boldsymbol{b})$ in order to address these shortcomings.

Methodology for the Modification of the Penalizing Factor in the $\boldsymbol{PRESS^{**}(b)}$ Criterion

In the modification of the $PRESS^{**}(\boldsymbol{b})$ criterion, we seek to achieve the following objectives:

i. to provide for the inclusion of a penalizing factor that utilizes every bit of information which the entire n^2 elements of the LQR Hat matrix can provide.
ii. to provide for the inclusion of a penalizing factor that allows the flexibility of LQR to be over and above that of the OLS.

If the proposed modified $PRESS^{**}(\boldsymbol{b})$ criterion is designated by $PRESS^{sr}(\boldsymbol{b})$, then $PRESS^{sr}(\boldsymbol{b})$ may be given by:

$$PRESS^{sr}(\boldsymbol{b}) = \frac{PRESS}{\left[(n-k-1)\left(\frac{SSE_{max}-SSE(\boldsymbol{b})}{SSE_{max}}\right)\right]Q}, \qquad (9)$$

where Q is a function of a statistical measure that encapsulates objectives (i) and (ii) above.
From Section 2, it is shown that the increase in the flexibility of the LQR derives from the increase in variability (or increase in the range) of the n elements in each of the rows of LQR Hat matrix. Hence, we compute the range of the n elements in each vector $\boldsymbol{h}_i^{(LQR)}(\boldsymbol{b}) = \left(\boldsymbol{h}_{i1}^{(LQR)}\boldsymbol{h}_{i2}^{(LQR)} \dots \boldsymbol{h}_{in}^{(LQR)}\right)$, $i = 1,2, \dots, n$, of the LQR Hat matrix.

This gives a $n \times 1$ vector of range, say $\boldsymbol{R} = \begin{pmatrix} range\left(\boldsymbol{h}_1^{(LQR)}(\boldsymbol{b})\right) \\ range\left(\boldsymbol{h}_2^{(LQR)}(\boldsymbol{b})\right) \\ \vdots \\ range\left(\boldsymbol{h}_n^{(LQR)}(\boldsymbol{b})\right) \end{pmatrix}$, for a given vector of bandwidths, $\boldsymbol{b}$.

Next, we get the sum of the range in $\boldsymbol{R}$. That is $\sum \boldsymbol{R} = \sum_{i=1}^{n} range(\boldsymbol{h}_i^{(LQR)}(\boldsymbol{b}))$.
Finally, in order to ensure that the proposed modified $PRESS^{**}(\boldsymbol{b})$ selects bandwidths according to objective (ii) above, we will have Q given by:

$Q = \sum_{i=1}^{n} range\left(\boldsymbol{h}_i^{(LQR)}(\boldsymbol{b})\right) - \sum_{i=1}^{n} range\left(\boldsymbol{h}_i^{(OLS)}\right)$.

where $range\left(\boldsymbol{h}_i^{(OLS)}\right)$, $i = 1,2, \dots, n$, is the range of the elements in the i^{th} row of the n by n OLS Hat matrix, $\boldsymbol{H}^{(OLS)}$.
Therefore, the proposed bandwidths selection criterion reduces to comes out as:

$$PRESS^{sr}(\boldsymbol{b}) = \frac{PRESS}{\left[(n-k-1)\left(\frac{SSE_{max}-SSE(\boldsymbol{b})}{SSE_{max}}\right)\right]\left[\sum_{i=1}^{n} range\left(\boldsymbol{h}_i^{(LQR)}(\boldsymbol{b})\right)-\sum_{i=1}^{n} range\left(\boldsymbol{h}_i^{(LQR)}(\boldsymbol{b})\right)\right]}, \quad (10)$$

Clearly, the statistic, $\sum \boldsymbol{R} = \sum_{i=1}^{n} range(\boldsymbol{h}_i^{(LQR)}(\boldsymbol{b}))$ is computed from the entire n^2 elements of the LQR Hat matrix. Further, $PRESS^{sr}(\boldsymbol{b})$ is minimized at a given vector of bandwidths, $\boldsymbol{b}$ at which the difference $\sum_{i=1}^{n} range\left(\boldsymbol{h}_i^{(LQR)}(\boldsymbol{b})\right) - \sum_{i=1}^{n} range\left(\boldsymbol{h}_i^{(LQR)}(\boldsymbol{b})\right)$ (that is the flexibility between the LQR and OLS) is as large as possible.

Application

LQR that utilizes the proposed $PRESS^{sr}(\mathbf{b})$ criterion for bandwidths selection (herein designated LQR^* for ease of reference) is applied to two multiple response problems from RSM literature and two sets of simulated data and results compared with those from OLS and LQR that utilizes the $PRESS^{**}(\boldsymbol{b})$ criterion.

The performance statistics for comparison include SSE and Coefficient of Determination, (R^2), which respectively indicate the nearness of the fitted responses to the observed values and a measure of variability in the data that is explained or captured by each regression model.

For each model, the values of the sum of the range, $\sum_{i=1}^{n} range\left(\boldsymbol{h}_i^{(.)}\right)$ of the elements in each row of the Hat matrix is presented under the column labelled SRR in Table 4. For the comparison of optimization results, the values of desirability measures in (4*) in the Appendix were used. The best value for each performance statistics (goodness-of-fit and optimization solution) are shown in bold.

The Chemical Process Data

This problem originates from Montgomery (2005) were it was analyzed using OLS. It involves three response variables, namely the y_1 (yield), y_2 (viscosity), and y_3 (molecular weight). Two inputs (factors) were found to influence these responses: reaction time (x_1) and temperature (x_2). A full second-order polynomial were found to be adequate for each of the response variables.

The process specifications for each response are as follows:
Maximize y_1 with lower limit $L = 78.5$, and target value, $\emptyset = 80$;
y_2 is to take a value in the range of $L = 62$ and $U = 68$ with target value, $\emptyset$ =65;
Minimize y_3 with upper limit $U = 3300$ and target value, $\emptyset = 3100$.

The data, collected via a Central Composite Design (CCD), is presented in Table 1. The optimal tuning parameters for the nonparametric models based on $PRESS^{sr}(\boldsymbol{b})$ and $PRESS^{**}(\boldsymbol{b})$ for LQR^* and LQR, respectively, are given in Table 2, and the corresponding locally adaptive bandwidths are shown in Table 3. The goodness-of-fit and optimization results for each of the regression models are presented in Tables 4 and 5, respectively.

Table 1: The chemical process data

I	x_1	x_2	y_1	y_2	y_3
1	0.1464	0.1464	76.5	62	2940
2	0.8536	0.1464	78.0	66	3680
3	0.1464	0.8536	77.0	60	3470
4	0.8536	0.8536	79.5	59	3890
5	0.0000	0.5000	75.6	71	3020
6	1.0000	0.5000	78.4	68	3360
7	0.5000	0.0000	77.0	57	3150
8	0.5000	1.0000	78.5	58	3630
9	0.5000	0.5000	79.9	72	3480
10	0.5000	0.5000	80.3	69	3200
11	0.5000	0.5000	80.0	68	3410
12	0.5000	0.5000	79.7	70	3290
13	0.5000	0.5000	79.8	71	3500

Table 2: Optimal tuning parameters for the chemical process data

Response	Models	N^*	C^*
y_1	LQR	6.3536	0.0797
	LQR*	3.3938	0.0846
y_2	LQR	5.3234	0.0228
	LQR*	3.6232	0.0192
y_3	LQR	5.9081	0.0884
	LQR*	2.4242	0.0884

Table 3: Bandwidths obtained from optimal tuning parameters in Table 2

	y_1		y_2		y_3	
i	**LQR**	**LQR***	**LQR**	**LQR***	**LQR**	**LQR***
1	0.8298	0.3270	0.3787	0.2591	0.8558	0.3512
2	0.5710	0.2770	0.4143	0.2818	0.1901	0.0780
3	0.7435	0.3103	0.3609	0.2477	0.3790	0.1555
4	0.3122	0.2270	0.3520	0.2420	0.0012	0.0005
5	0.9851	0.3570	0.4587	0.3101	0.7838	0.3216
6	0.5020	0.2636	0.4321	0.2931	0.4780	0.1961
7	0.7435	0.3103	0.3343	0.2307	0.6669	0.2736
8	0.4848	0.2603	0.3432	0.2364	0.2351	0.0965
9	0.2432	0.2136	0.4676	0.3158	0.3700	0.1518
10	0.1742	0.2003	0.4410	0.2988	0.6219	0.2552
11	0.2260	0.2103	0.4321	0.2931	0.4330	0.1777
12	0.2777	0.2203	0.4498	0.3045	0.5410	0.2270
13	0.2605	0.2170	0.4587	0.3101	0.3521	0.1445

Table 4: Goodness-of-fit of the regression models for the chemical process data

Response	Model	$PRESS^{**}$	$PRESS^{sr}$	DF	SRR	SSE	R^2
y_1	**OLS**	-	-	7.0000	6.0003	0.4962	98.2735
	LQR	0.2526	-	6.5802	7.1504	0.4468	98.4456
	LQR*	-	0.1628	4.0748	**8.9457**	**0.2127**	**99.2598**
y_2	**OLS**	-	-	7.0000	6.0003	36.2242	89.9720
	LQR	13.0953	-	4.9192	8.2690	12.4398	96.5563
	LQR*	-	7.0752	4.0270	**8.9784**	**10.0027**	**97.2309**
y_3	**OLS**	-	-	7.0000	6.0003	207870	75.8990
	LQR	82840	-	5.0093	8.4253	77067	91.0648
	LQR*	-	43674	4.0006	**8.9999**	**65720**	**92.3804**

From Table 4, we observe that LQR^* gives better SSE and R^2 across the three responses, signifying a model of better fit than the OLS and LQR. Further, LQR^* has largest SRR across the three responses as well.

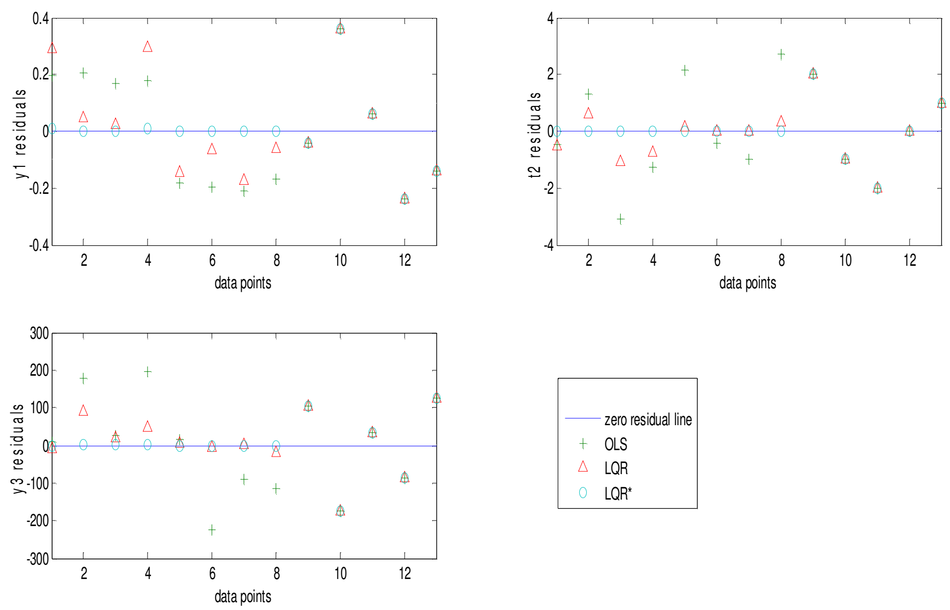

Fig. 1: Graphical comparison of plots of residuals of response estimates

From Figure 1, the plots of residuals of $\hat{y}_1$ (Top Left), $\hat{y}_2$ (Top Right), and $\hat{y}_3$ (Bottom Left) show that those from the LQR* are seen to lie relatively closest to the zero residual lines, indicative of relatively better fit of the LQR* to the given data.

Table 5: Optimization results based on the desirability measure for the chemical process data

Models	x_1	x_2	Max$(\hat{y}_1)$	$\emptyset(\hat{y}_2)$	min $(\hat{y}_3)$	d$(\hat{y}_1)$	d$(\hat{y}_2)$	d$(\hat{y}_3)$	D(%)
OLS	0.4449	0.2226	78.7616	66.4827	3229.9	0.1744	0.5058	0.3504	31.3800
LQR	0.4892	0.2093	78.7993	66.1764	3188.5	0.1996	0.6079	0.5576	40.7450
LQR*	0.4451	0.2152	**78.9944**	**65.0015**	**3085.8**	0.3296	0.9995	1.0000	**69.0639**

The results presented in Table 5 show that LQR^* found a better setting of the explanatory variables that simultaneously optimizes the three responses with a desirability measure of 69.0639%, indicating a product that meets approximately 70% of the production requirements as compared with the 41% for the LQR. The enhanced flexibility of LQR enables its exploration of the solution space for better optimal results.

CONCLUSION

In this paper, we proposed a modification of the PRESS** criterion to suit the selection of bandwidths for LQR in the response surface settings. The modified PRESS** criterion, designated $PRESS^{sr}$ criterion, involves the replacement of the DF term in the denominator of the PRESS** criterion with the difference in the sum of the range of the rows of the Hat matrix of the LQR and that of the OLS models. This was done in order to improve on the flexibility of the LQR model which happens to be the most appealing feature of nonparametric regression models in general.

From the data analyzed, LQR utilizing $PRESS^{sr}$ (designated LQR^*) gives the best SSE and R^2 in the three out of the three responses in the problem taken from literature. Much more significantly, LQR^*gives superior optimization results, trumping those from OLS and the LQR. Specifically, better optimization result (Table 5) translates to better use of scarce resources (raw material, time, e.tc).

REFERENCES

Anderson-Cook, C. M. and Prewitt, K. (2005). Some guidelines for using nonparametric models for modeling data from response surface designs, *Journal of Modern Applied Statistical Models,* 4: 106-119.

Castillo, D. E. (2007). *Process optimization: A statistical method.* New York: Springer International Series in Operations Research and Management Science.

DePalma, K., Smith, B. and McDonald, A. G. (2021). Synergistic effects of processing parameters on the biochemical and physical properties of tofu made from yellow field pea (*Pisum sativum*), as determined by response surface methodology. *Food Sci. Nutr.* 9: 1132-1142.

Derringer, G. and Suich, R. (1980). Simultaneous optimization of several response variables, *Journal of Quality Technology,* 12(4): 214 – 219.

Edionwe, E. and Mbegbu, J. I. (2014). Local bandwidths for improving the performance statistics of model robust regression 2, *Journal of Modern Applied Statistical Methods,* 13(2): 506-527.

Edionwe, E., Mbegbu, J. I. and Chinwe, R. (2016). A new function for generating local bandwidths for semiparametric MRR2 model in response surface methodology, *Journal of Quality Technology,* 48(4): 388 –404.

Edionwe, E., Mbegbu, J. I., Ekhosuehi, N., Obiora-Ilouno, H. O. (2018).

An improved robust regression model for response surface methodology. *Croatian Operational Research Review,* 9: 317 – 330.

Eguasa, O., Edionwe E. and Mbegbu J. I. (2022). Local Linear Regression and the Problem of Dimensionality: A Remedial Strategy via a New Locally Adaptive Bandwidths Selector, *Journal of Applied Statistics*, DOI: 10.1080/02664763.2022.2026895

Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its applications.* Chapman and Hall, London.

Harrington, E. C. (1965). The desirability function, *Industrial Quality Control*, 21(10): 494 – 498.

He, Z., Zhu, P. F. and Park, S. H. (2012). A robust desirability function for multi-response surface optimization, *European Journal of Operational Research,* 221: 241-247.

Mays, J. E., Birch, J. B. and Starnes, B. A. (2001). Model-robust regression: combining parametric, nonparametric, and semi-parametric models, *Journal of Nonparametric Statistics,* 13: 245-277.

Mays, J. E. and Birch, J. B. (1998). Smoothing considerations in nonparametric and semiparametric regression, *Technical Report Number 98-2,* Department of Statistics, Virginia Polytechnic Institute and State University.

Montgomery, D. C. (2005). *Design and analysis of experiments.* Sixth ed. Wiley, New York.

Myers, R. H. (1999). Response surface methodology - current status and future directions. *Journal of Quality Technology,* 31: 30-44.

Myers, R., Montgomery, D. C. and Anderson-Cook, C. M. (2009). *Response surface methodology: process and product optimization using designed experiments.* 3rd Edition, Wiley, ISBN: 978-0-470-17446-3.

Rajewski, J. and Dobrzynska-Inger, A. (2021). Application of Response Surface Methodology (RSM) for the Optimization of Chromium (III) Synergistic Extraction by Supported Liquid Membrane, 11, 854, 1-17

Ryu, D. H., Cho, J. Y., Sadiq, N. B., Kim, J. C., Lee, B., Hamayun, M., Lee, T. S., Kim, H. S., Park, S. H. and Nho, C. W. (2021). Optimization of antioxidant, anti-diabetic, and anti-inflammatory activities and ganoderic acid content of differentially dried *Ganoderma lucidum* using response surface methodology. *Food Chem.,* 335: 127645.

Vivek, K., Mishra, S. and Pradhan, R. C. (2021), *Optimization of Spray Drying Conditions for Developing Nondairy Based Probiotic Sohiong Fruit Powder. J. Fruit Sci*., 21(1): 193-204.

Wan, W. and Birch, J. B. (2011). A semi-parametric technique for multi-response optimization, *Journal of Quality and Reliability Engineering International,* 27: 47-59.