Local Bandwidths for Improving Performance Statistics of Model-Robust Regression 2

Efosa Edionwe
University of Benin, Benin City, Nigeria, trustnelson24@yahoo.com

Julian L. Mbegbu
University of Benin, Benin, Nigeria, julian.mbegbu@yahoo.com

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
DOI: 10.22237/jmasm/1414816140
Available at: http://digitalcommons.wayne.edu/jmasm/vol13/iss2/30

This Emerging Scholar is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.
Local Bandwidths for Improving Performance Statistics of Model-Robust Regression 2

Efosa Edionwe  
University of Benin  
Benin City, Nigeria

Julian I. Mbegbu  
University of Benin  
Benin City, Nigeria

Model-Robust Regression 2 (MRR2) method is a semi-parametric regression approach that combines parametric and nonparametric fits. The bandwidth controls the smoothness of the nonparametric portion. We present a methodology for deriving data-driven local bandwidth that enhances the performance of MRR2 method for fitting curves to data generated from designed experiments.

Keywords: Semi-parametric methods, Model-Robust Regression, response surface methodology, local bandwidths

Introduction

The understanding of any system or process is enhanced by the availability of fairly accurate mathematical relations connecting the explanatory variables and the dependent variables (responses) of the system. The desire to obtain such mathematical relations led to the development of response surface methodology (RSM) which is a collection of mathematical and statistical techniques employed for modeling and analysis of problems in which a response of interest is influenced by several explanatory variables (Montgomery, 1999; Wu & Hamada, 2000; Raissi & Farsani, 2009). The objective of RSM is to optimize one or more responses, which are influenced by several explanatory variables.

RSM consists of three main phases, namely the experimental design phase, the modeling phase, and the optimization phase (Del Castillo, 2007). However, the efficiency and reliability of the optimal solutions achieved at the optimization phase depends on the results obtained in the modeling phase. Better results obtained in
the modeling phase ensure better optimal solution in the optimization phase (Pickle, 2006).

The modeling phase involves the use of regression techniques to fit a curve to the data generated from the experiment. Regression techniques employed in RSM include parametric regression, nonparametric regression and semi-parametric regression (Pickle, 2006; Wan, 2007).

In Parametric regression, a low-order polynomial such as the \( k \)-factor second-order model of the form

\[
y_i = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j} \beta_{ij} x_i x_j + \epsilon_i
\]

is assumed for fitting the data, where \( \beta_i, \beta_{ii} \) and \( \beta_{ij} \) are the model parameters, \( x_i \) and \( x_j \) are the explanatory (Del Castillo, 2007). For \( n \)-sample observations, (1) can be expressed in matrix form as

\[
\mathbf{Y} = \mathbf{X}\beta + \mathbf{\epsilon}
\]

where \( \mathbf{Y} \) is an \( n \) by 1 vector of responses, \( \mathbf{X} \) is an \( n \) by \( 1 + 2k + \left( \begin{array}{c} k \\ 2 \end{array} \right) \) model matrix, \( \beta \) is a \( 1 + 2k + \left( \begin{array}{c} k \\ 2 \end{array} \right) \) by 1 vector of unknown model parameters, and \( \mathbf{\epsilon} \) is an \( n \) by 1 vector of random errors. The Ordinary Least Squares (OLS) method gives the vector of estimated responses is given as

\[
\hat{\mathbf{y}}^{(\text{OLS})} = \mathbf{X}\hat{\beta}_{\text{OLS}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}^{(\text{OLS})}\mathbf{y}
\]

A disadvantage of the parametric regression method is that if the assumed model is misspecified, the fitted curve is affected by high bias (Einsporn & Birch, 1993; Mays, 2001b; Pickle, 2006).

In nonparametric regression, the user approaches the problem without assuming a model and attempts to fit a curve to the data points by employing a weighting scheme (Uysal & Guvenir, 1999; DiNardo & Tobias, 2001). Most often, nonparametric regression (for example, the kernel regression, local linear regression) is employed when a theoretical reference curve is unavailable for a process and the data size is large (Hens, 2005; Hernández-Lobato, 2010). The local
linear regression (LLR) utilizes kernel weights for smoothing. For instance, the LLR estimate of observation $y_0$ obtained at the location $\tilde{x}_0$ is given as

$$
\hat{y}_0^{(LLR)} = \tilde{x}_0' (\tilde{X} W_0 \tilde{X})^{-1} \tilde{X} W_0 y = h_0^{(LLR)} y
$$

(4)

where $\tilde{X}' = (x_1, \ldots, x_n)$, the $n$ by $n$ diagonal matrix $W_0$, known as the local weight matrix for location $x_0$, is given by

$$
W_0 = diag \left( h_{01}^{(KER)}, \ldots, h_{04}^{(KER)} \right)
$$

(6)

$h_{0i}^{(KER)}$ represents a kernel weight assigned to $y_i$ in the estimation of $y_0$ at location $x_0$ and is given as

$$
h_{0i}^{(KER)} = \frac{K \left( \frac{x_0 - x_i}{b} \right)}{\sum_{i=1}^{n} K \left( \frac{x_0 - x_i}{b} \right)}
$$

(7)

$K$ is a univariate kernel function, $b$ is referred to as the bandwidth (Härdle, 1994). A commonly used kernel function is the simplified Gaussian kernel function given as

$$
K \left( \frac{x_0 - x_i}{b} \right) = e^{-\left( \frac{x_0 - x_i}{b} \right)^2}
$$

(8)
For the multivariate case with \( k \) explanatory variables, a common form of the Gaussian kernel function used is the product kernel given as

\[
K(\hat{x}_0, \tilde{x}_i) = \prod_{j=1}^{k} K \left( \frac{\hat{x}_{0j} - \tilde{x}_{ij}}{b} \right)
\]  

(9)

\( \hat{x}_0 = (x_{01}, x_{02}, \ldots, x_{0k}) \) is the prediction point, \( K \) is the univariate kernel function (Wan, 2007). In general, using matrix notation, LLR estimated response can be written as

\[
\hat{y}^{(LLR)} = H^{(LLR)} y
\]  

(10)

where \( H^{(LLR)} \) is the LLR “HAT” or smoother matrix defined as

\[
H^{(LLR)} = \begin{bmatrix}
    h_{1y}^{(LLR)} \\
    h_{2y}^{(LLR)} \\
    \vdots \\
    h_{ny}^{(LLR)}
\end{bmatrix}
\]  

(11)

\[
h_{iy}^{(LLR)} = \hat{x}_i' (\hat{X}' W_i \hat{X})^{-1} \hat{X}' W_i
\]  

(12)

A disadvantage of nonparametric methods is that large amounts of data are required. Moreover, the capacity of nonparametric methods to describe complex patterns makes them more prone to overfitting (Mays, 2001b; Wan, 2007; Starnes, Birch & Robinson, 2008).

Semiparametric regression methods involve fitting the data both parametrically and nonparametrically, and then combining the results to form a curve that is based on suitable theoretical form, yet still being able to adapt to aberrations or misspecifications from that form. Hence semi-parametric regression techniques are robust to model misspecifications (Starnes, 1999; Mays, 2001a; Hens, 2005).

Starnes (1999) and Pickle (2006) reported that MRR2 is the best overall semi-parametric regression procedure for fitting small-sample data in situation of small to moderate model misspecification.
IMPROVING PERFORMANCE STATISTICS OF MRR2

MRR2 technique combines a parametric fit like \( \hat{y}^{(OLS)} \) to the raw data and a nonparametric fit to the vector of residuals, \( r \), from the parametric fit. The MRR2 fit at location \( x_0 \) is given by

\[
\hat{y}_0^{(MRR2)} = \hat{x}_0' (X'X)^{-1} X'y + \hat{\lambda} \hat{x}_0' \left( \tilde{X}' W_r \tilde{X}' \right)^{-1} \tilde{X}' W_r \left( I - X(X'X)^{-1} X' \right) y
\]

(13)

where

\[
h_0^{(MRR2)'} = \hat{x}_0' (X'X)^{-1} X' + \hat{\lambda} \hat{x}_0' \left( \tilde{X}' W_r \tilde{X}' \right)^{-1} \tilde{X}' W_r \left( I - X(X'X)^{-1} X' \right)
\]

\( r \) is the vector of residuals from the parametric fits, \( I \) is an \( n \) by \( n \) identity matrix, \( W_r \) is the \( n \) by \( n \) diagonal matrix containing the kernel weights for fitting the parametric residuals and is obtained using the same procedure as in (6), the matrices \( X \) and \( \tilde{X} \) are as defined in (2) and (5) respectively. MRR2 fits is expressed in matrix form as

\[
\hat{y}^{(MRR2)} = \left[ H^{(OLS)} + \lambda H^{(LLR)} \left( I - H^{(OLS)} \right) \right] y
\]

(14)

\[
= H^{(MRR2)} y
\]

(15)

where \( H^{(MRR2)} \) is the MRR2 “HAT” matrix defined as

\[
H^{(MRR2)} = \begin{bmatrix}
h_1^{(MRR2)'} \\
h_2^{(MRR2)'} \\
\vdots \\
h_n^{(MRR2)'}
\end{bmatrix}
\]

(16)
Wan (2007) reported two expressions for the mixing parameter, $\lambda$. One is the estimated asymptotically optimal mixing parameter given by

$$\hat{\lambda}^* = \frac{\langle \hat{r}, r \rangle}{\|\hat{r}\|^2}$$

(17)

the notation $\langle \cdot \rangle$ represents the inner product and $\| \|$ is the standard $L_2$ (Euclidean) norm. A second data-driven method chooses $\hat{\lambda}^*$ such that $\text{PRESS}^{**}(\hat{\lambda})$ defined for a given optimal bandwidth, $b^*$, as

$$\text{PRESS}^{**}(\hat{\lambda}) = \sum_{i=1}^{n} \left( y_i - \hat{y}_{i,:-1}(b^*, \hat{\lambda}) \right)^2$$

$$n - \text{tr}\left( H^{(MRR2)}(b^*, \hat{\lambda}) \right) + (n - k - 1) \frac{\text{SSE}_{\max} - \text{SSE}_{b^*}}{\text{SSE}_{\max}}$$

(18)

is minimized. $n$ is the sample size, $b^*$ is the optimal bandwidth, $k$ is the number of explanatory variables, $\text{SSE}_{\max}$ is the maximum sum of squared errors obtained as $b$ tends to infinity, $\text{SSE}_{b^*}$ is the sum of squared errors for the optimal bandwidth, $\text{tr}\left( H^{(MRR2)}(b^*, \hat{\lambda}) \right)$ is the trace of the MRR2 “HAT” matrix for a given $b^*$ and $\hat{\lambda}$, and $\hat{y}_{i,:-1}$ is the fit at $x_i$ with the $i^{th}$ observation left out.

The bandwidth is an important parameter in that it determines performance of the model in terms of criteria such as variance, mean squared error (Huang & Fan, 1996). A bandwidth is said to be fixed or global if its value is constant for the full range of the data or if does not change with locations or runs in a given regression procedure otherwise it is referred to as local, variable or adaptive bandwidth. For a given location, local bandwidths are chosen according to factors involving the values of the explanatory variables, $x_i$, or of the response, $y_i$, or both (Starnes, 1999). This dependence allows different degree of smoothing for different locations in the data thereby giving the data more privilege to determine the functional form of the model fitted and to incorporate the information provided by the density of the data.

Among the categories of methods for selecting bandwidths, the most frequently employed procedures include the plug-in methods and the cross-validation methods (Fan & Gijbels, 1992; Atkeson, Moore & Schaal, 1997; Gerard & Schucany, 1999; Racine, 2008; Avery, 2010; Kohler, Schindler and Sperlich, 2011). However, all the criteria for selecting bandwidths are based on the same philosophy, and they are such that the fitted value $\hat{y}(x)$ is as close to the true value $y(x)$ as possible thereby minimizing errors associated with estimation (Härdle,
IMPROVING PERFORMANCE STATISTICS OF MRR2

1994; Galdo & Black, 2008). Researches applying MRR2 for fitting curves to RSM data employ a penalized version of prediction error sum of squares referred to as PRESS** and given by

\[
\text{PRESS}^{**}(b) = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_{i-1}(b))^2}{n - 
\text{tr}\left( H^{(MRR2)}(b) \right) + (n - k - 1) \frac{\text{SSE}_{\text{max}} - \text{SSE}_{\text{b}}}{\text{SSE}_{\text{max}}}}
\]  

(19)

where all the parameters retain their previous definitions in (18) (Pickle, 2005; Wan, 2007).

The remainder of this paper is organized as follows: A review of methods for deriving local bandwidth is presented in the next section. A new methodology for deriving data-driven local bandwidths follows. After that, results of application of MRR2 and LLR methods utilizing the new data-driven local bandwidths to a multiresponse problem are presented. Finally, a discussion on the comparison of results from OLS, LLR and MRR2 (both fixed optimal bandwidth and local optimal bandwidths) is presented.

A Review of Methods for Deriving Local Bandwidths

For kernel density estimation, Fan and Gijbels (1992), using Average Mean Integrated Squared Error, gave an expression for optimal variable bandwidths as

\[
\alpha_{\text{opt}}(x) = \begin{cases} 
 b \left( \frac{f_X(x) [m''(x)]^2}{\sigma^2(x)} \right)^{\frac{1}{2}}, & \text{if } W(x) > 0, \\
\alpha^*(x), & \text{if } W(x) = 0
\end{cases}
\]  

(21)

where \( b \) is any arbitrarily positive constant, \( \alpha^*(x) \) can take any value greater than zero, \( W(x) \) is a diagonal matrix of weights, \( m''(x) \) is the second-derivative of the unknown function, \( f_X(.) \) and \( \sigma^2(.) \) are the marginal density of \( X \) and the conditional variance of \( Y \) given \( X \) respectively. The limitation in the use of this variable bandwidth is that it requires estimates of \( f_X(.) \), \( m''(x) \), and \( \sigma^2(.) \) respectively. Hence, the efficiency of (21) depends on how close these estimates are to the true values.
Schucany (1995) proposed a variable bandwidth selector for both the kernel and local linear regression. An expression for the optimal bandwidth is given by

$$h^{\text{SCH}}_{\text{opt}}(x) = \left( \frac{\sigma^2 A}{2pnB(x)^2} \right)^{1/(2p+1)}$$

(22)

where $p$ is the degree of the polynomial, $n$ is the number of observations, $A$ is a constant which depends on the kernel, $B(x)$ is an approximation for the bias. Again, $h^{\text{SCH}}_{\text{opt}}(x)$ is calculated using estimates of $\sigma^2$ and $B(x)$. Hence, the quality of the final estimator $\hat{h}^{\text{SCH}}_{\text{opt}}$ depends on the choice of a “pilot bandwidth” from which an estimate of the $B(x)$ is obtained. Moreover, (22) is developed for cases where the levels of a single explanatory variable are equally-spaced.

Few of the plug-in methods for obtaining variable bandwidths are used in practice due to computational difficulty. Plugs methods seem logically inconsistent since they require higher order smoothness of the unknown function (Bickels & Li, 2007; Galdo & Black., 2008; Avery, 2010).

A Local cross-validation variable bandwidth which reflects the impact of the responses and suitable for a single explanatory is considered in Zheng (2010) and is given as

$$h^{**}(x) = \arg \left\{ \min_h \sum_{i=1}^{l(x)} \left( Y_i - \hat{m}_{l-i}(X'_i) \right)^2 \right\}$$

(23)

where $l(x)$ denotes the number of covariate values falling in a certain defined interval $I_{x-[x-h'(x),x+h'(x)]}$ and $(X'_i,Y'_i), i=1,\ldots,l(x)$ denotes the number observations falling in the interval, $h'(x)$ is a sequence of a version of optimal Bayesian bandwidths, and $\hat{m}_{l-i}(X'_i)$ is given as

$$\hat{m}_{l-i}(X'_i) = \frac{1}{(n-1)h} \sum_{j \neq i} K \left( \frac{x_j - x_i}{h} \right) \frac{M_{zi}(x_i) - \left( \frac{x_j - x_i}{h} \right) M_{li}(x_i)}{M_{zi}(x_i)M_{0i}(x_i) - M_{li}(x_i)^2} Y_j$$

(24)
and

\[ M_{ci}(X_i) = \left( (n-1)h \right)^{-1} \sum_{j \neq i} K \left( \frac{x_i - x_j}{h} \right)^c, \quad \text{for } c = 1, 2, 3 \]  

(25)

The method in (24) works for a single explanatory variable. Besides, the choice of \( l(x) \) is dependent on \( h^*(x) \) which, according to the author, requires estimates of some prior parameters.

A methodology for the derivation of a function for generating local bandwidth is presented in the section that follows. The local bandwidth generated by the function can be applied to data with more than one explanatory variable. Furthermore, typical of cross validation procedures, no estimates of parameters is required for the utilization of the proposed function.

**Methodology**

**Derivation of a Function for Generating Local Bandwidth**

A new methodology used to derive a function for generating data-driven local bandwidth is presented. In deriving the function, the basic objectives to achieve are:

- to allow the values of the bandwidths to be a function of the observations we intend to fit;
- to assume that a real number N, which also acts as a tuning parameter is the sum of all the bandwidths that minimize PRESS**. The simplified kernel function, which is a decreasing function, is utilized in the paper. Therefore, the function generating the local bandwidth is modeled in a manner that locations with relatively smaller observations are assigned smaller bandwidths (corresponding to heavier weights via the kernel function), and vice versa. For convenience, this function is referred to as N-squared function and its derivation is as follows:

Given that \( T = \sum_{i=1}^{n} y_i \) , \( (T \) is the sum of all the observations), \( n \) is the number of observations or locations, or sample size, \( b_i, i = 1, \ldots, n \), is the bandwidth for the \( i^{th} \) location and \( N \) is the sum of the bandwidths that minimize PRESS**.

First, it is required that for each location, the bandwidth be a function of the size of observation at that location, and one of the ways to achieve this is to express the bandwidth, \( b_i \), as ratio of the \( i^{th} \) observation to the sum of the data, \( T \)

\[ b_i = \frac{y_i}{T}, \quad \text{for } i = 1, \ldots, n. \]  

(26)
Because the simplified Kernel function is a decreasing function, hence to ensure heavier kernel weights are assigned to smaller observations and vice versa, smaller observations need to have smaller bandwidths and to achieve this, (26) is expressed as

\[ b_i = \left( N - \frac{y_i}{T} \right) \]  

(27)

Taking sum of both sides of (27) gives

\[ \sum_i^n b_i = \sum_i^n \left( N - \frac{y_i}{T} \right) \]  

(28)

\[ \sum_i^n b_i = \sum_i^n \left( \frac{NT - y_i}{T} \right) \]  

(29)

Next, proceed to determine a factor that multiplies the right-hand side of (29) to ensure the bandwidths sum to a value \( N \).

By expanding the right side of (29)

\[ \sum_i^n b_i = \left( \frac{(NT - y_1)}{T} + \frac{(NT - y_2)}{T} + \cdots + \frac{(NT - y_n)}{T} \right) \]  

(30)

On collecting like terms in (30)

\[ \sum_i^n b_i = \frac{NTn - \sum_{i=1}^n y_i}{T} \]  

(31)

But the sum of the data, \( \sum_{i=1}^n y_i \), is equal to \( T \) as previously defined hence (31) reduces to

\[ \sum_i^n b_i = \frac{NTn - T}{T} \]  

(32)
Therefore

$$\sum_i^n b_i = (Nn - 1) \tag{33}$$

which implies

$$\sum_i^n \left( \frac{NT - y_i}{T} \right) = (Nn - 1). \tag{34}$$

Hence, to ensure the bandwidths sum to a value $N$, we need to multiply the right hand side of (29) by a factor $N/(Nn - 1)$, giving

$$\sum_i^n b_i = \sum_i^n \left( \frac{NT - y_i}{T} \right) N/I(Nn - 1) \tag{35}$$

Therefore

$$b_i = \frac{N^2 T - Ny_i}{T(Nn - 1)}, \quad i = 1, 2, \ldots, n, \tag{36}$$

$$b_i = \frac{N^2 \sum_{i=1}^n y_i - Ny_i}{(Nn - 1) \sum_{i=1}^n y_i}, \quad i = 1, 2, \ldots, n. \tag{37}$$

Equation (37) gives the $N$-squared function for data-driven variable bandwidths. The optimal local bandwidth, $b$ is a vector whose elements are the bandwidths $b_i$, (for smoothing $i^{th}$ location of the observation), $i = 1, 2, \ldots, n$, obtained at the value of $N$ in (37) where $PRESS^\ast (b)$ given by

$$PRESS^\ast (b) = \frac{\sum_i^n \left( y_i - y_{i-i} (b) \right)^2}{n - \text{trace} \left( H^{(i)} (b) \right) + (n-k-1) \left( \frac{SSE_{\max} - SSE_b}{SSE_{\max}} \right)} \tag{38}$$
where $H^{(.)}$ is the “HAT” matrix for (LLR) or MRR2 obtained by using local bandwidths from N-squared function, $SSE_{\text{max}}$ is the maximum sum of squares of errors over all possible bandwidths which is equivalent to $\sum_{i=1}^{n} \left( y_i - \hat{y}_{i}^{(\text{OLS})} \right)^2$ for LLR or $\sum_{i=1}^{n} \left( e_i - \hat{e}_{i}^{(\text{OLS})} \right)^2$ for MRR2 where $\hat{y}_{i}^{(\text{OLS})}$ and $\hat{e}_{i}^{(\text{OLS})}$ are the OLS fit of a first-order model for responses and OLS residuals respectively, $SSE_{b_i}$ is given by $\sum_{i=1}^{n} \left( y_i - \hat{y}_{i}^{(\text{LLR})} (b_i) \right)^2$ for LLR or $\sum_{i=1}^{n} \left( y_i - \hat{y}_{i}^{(\text{MRR2})} (b_i) \right)^2$ for the MRR2 counterpart. (See Wan, 2007). For MRR2, the mixing parameter, $\lambda$, is obtained using equations (17) or (18).

LLR and MRR2 methods are applied using local bandwidth derived from N-squared function to the Minced Fish Quality problem Wan (2007) and its performance is compared with results from parametric, (OLS), LLR, (fixed bandwidth), and MRR2, (fixed bandwidth), approaches. The comparison is based on some performance criteria including, estimate of the variance, ($S^2$), the coefficient of determination, ($R^2$), adjusted coefficient of determination, ($R_{\text{adj}}^2$), PRESS given in (20), $PRESS^* = PRESS / DF_{\text{error}}$, where $DF_{\text{error}} = DF_{\text{total}} - DF_{\text{model}}$, and $PRESS^{**}$.

**Application of Local Bandwidths from N-Squared Function**

The data for the Minced Fish Quality problem presented in Wan (2007) is from food science and is used here to compare the performance of the modeling techniques discussed herein. The problem involves three independent variables $x_1$, $x_2$, $x_3$ which represent washing temperatures, washing time, washing ratio of water volume to sample weight respectively, and four response variables $y_1$, $y_2$, $y_3$, $y_4$, representing springiness, thiobarbituric acid number, (TBA), cooking loss, and whiteness index respectively. The observed data were collected through a Central Composite Design, (CCD), and is presented in Table 1.

According to Wan (2007), the final fitted second-order models for OLS for responses $y_1$ and $y_4$ include three terms: intercept, $x_1$ and $x_1^2$. The OLS model for response $y_2$ includes five terms: intercept, $x_1$, $x_2$, $x_1^2$, and $x_{12}$. The OLS model for
response $y_3$ has eight terms: intercept, $x_1$, $x_2$, $x_3$, $x_1^2$, $x_{12}$, $x_{13}$, and $x_3^2$. Therefore, the model spaces for the OLS part of the MRR2 are the same for the final fitted OLS models for each of the responses while the LLR and the nonparametric part of MRR2 utilize a first-order version of the OLS models since LLR and the nonparametric part of MRR2 are based local linear smoothing. Thus, the model matrix for the LLR and the nonparametric part of MRR2 for response $y_3$ consists of four terms: intercept, $x_1$, $x_2$, and $x_3$, the model matrices for response $y_1$ and $y_4$ both consist of the intercept and $x_1$, and that for response $y_2$ consists of the intercept, $x_1$ and $x_2$.

Table 1. A CCD with three factors and four responses on minced fish quality

<table>
<thead>
<tr>
<th>Coded Variables</th>
<th>Response values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
</tr>
<tr>
<td>1</td>
<td>0.203</td>
</tr>
<tr>
<td>2</td>
<td>0.797</td>
</tr>
<tr>
<td>3</td>
<td>0.203</td>
</tr>
<tr>
<td>4</td>
<td>0.797</td>
</tr>
<tr>
<td>5</td>
<td>0.203</td>
</tr>
<tr>
<td>6</td>
<td>0.797</td>
</tr>
<tr>
<td>7</td>
<td>0.203</td>
</tr>
<tr>
<td>8</td>
<td>0.797</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>0.5</td>
</tr>
<tr>
<td>12</td>
<td>0.5</td>
</tr>
<tr>
<td>13</td>
<td>0.5</td>
</tr>
<tr>
<td>14</td>
<td>0.5</td>
</tr>
<tr>
<td>15</td>
<td>0.5</td>
</tr>
<tr>
<td>16</td>
<td>0.5</td>
</tr>
<tr>
<td>17</td>
<td>0.5</td>
</tr>
</tbody>
</table>
### Table 2. Bandwidths for $LLR_{VB1}$ for each location and response

<table>
<thead>
<tr>
<th>Location</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0801</td>
<td>0.4377</td>
<td>0.5355</td>
<td>0.0800</td>
</tr>
<tr>
<td>2</td>
<td>0.0803</td>
<td>0.4365</td>
<td>0.5370</td>
<td>0.0801</td>
</tr>
<tr>
<td>3</td>
<td>0.0800</td>
<td>0.4382</td>
<td>0.5361</td>
<td>0.0799</td>
</tr>
<tr>
<td>4</td>
<td>0.0804</td>
<td>0.4364</td>
<td>0.5358</td>
<td>0.0800</td>
</tr>
<tr>
<td>5</td>
<td>0.0800</td>
<td>0.4376</td>
<td>0.5367</td>
<td>0.0800</td>
</tr>
<tr>
<td>6</td>
<td>0.0802</td>
<td>0.4373</td>
<td>0.5363</td>
<td>0.0800</td>
</tr>
<tr>
<td>7</td>
<td>0.0800</td>
<td>0.4385</td>
<td>0.5370</td>
<td>0.0800</td>
</tr>
<tr>
<td>8</td>
<td>0.0804</td>
<td>0.4364</td>
<td>0.5361</td>
<td>0.0800</td>
</tr>
<tr>
<td>9</td>
<td>0.0801</td>
<td>0.4372</td>
<td>0.5363</td>
<td>0.0815</td>
</tr>
<tr>
<td>10</td>
<td>0.0810</td>
<td>0.4350</td>
<td>0.5353</td>
<td>0.0799</td>
</tr>
<tr>
<td>11</td>
<td>0.0800</td>
<td>0.4387</td>
<td>0.5368</td>
<td>0.0801</td>
</tr>
<tr>
<td>12</td>
<td>0.0799</td>
<td>0.4387</td>
<td>0.5371</td>
<td>0.0798</td>
</tr>
<tr>
<td>13</td>
<td>0.0800</td>
<td>0.4383</td>
<td>0.5365</td>
<td>0.0800</td>
</tr>
<tr>
<td>14</td>
<td>0.0799</td>
<td>0.4388</td>
<td>0.5367</td>
<td>0.0799</td>
</tr>
<tr>
<td>15</td>
<td>0.0799</td>
<td>0.4386</td>
<td>0.5372</td>
<td>0.0799</td>
</tr>
<tr>
<td>16</td>
<td>0.0800</td>
<td>0.4386</td>
<td>0.5371</td>
<td>0.0799</td>
</tr>
<tr>
<td>17</td>
<td>0.0800</td>
<td>0.4386</td>
<td>0.5374</td>
<td>0.0799</td>
</tr>
</tbody>
</table>

### Table 3. Bandwidths for $MMR_{VB1}$ for each location and response

<table>
<thead>
<tr>
<th>Location</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0792</td>
<td>0.2568</td>
<td>0.3624</td>
<td>0.0791</td>
</tr>
<tr>
<td>2</td>
<td>0.0794</td>
<td>0.2556</td>
<td>0.3640</td>
<td>0.0793</td>
</tr>
<tr>
<td>3</td>
<td>0.0792</td>
<td>0.2573</td>
<td>0.3630</td>
<td>0.0791</td>
</tr>
<tr>
<td>4</td>
<td>0.0796</td>
<td>0.2555</td>
<td>0.3628</td>
<td>0.0792</td>
</tr>
<tr>
<td>5</td>
<td>0.0792</td>
<td>0.2567</td>
<td>0.3637</td>
<td>0.0792</td>
</tr>
<tr>
<td>6</td>
<td>0.0794</td>
<td>0.2564</td>
<td>0.3633</td>
<td>0.0791</td>
</tr>
<tr>
<td>7</td>
<td>0.0791</td>
<td>0.2577</td>
<td>0.3639</td>
<td>0.0791</td>
</tr>
<tr>
<td>8</td>
<td>0.0796</td>
<td>0.2555</td>
<td>0.3631</td>
<td>0.0791</td>
</tr>
<tr>
<td>9</td>
<td>0.0793</td>
<td>0.2563</td>
<td>0.3632</td>
<td>0.0807</td>
</tr>
<tr>
<td>10</td>
<td>0.0802</td>
<td>0.2541</td>
<td>0.3622</td>
<td>0.0791</td>
</tr>
<tr>
<td>11</td>
<td>0.0792</td>
<td>0.2579</td>
<td>0.3637</td>
<td>0.0793</td>
</tr>
<tr>
<td>12</td>
<td>0.0791</td>
<td>0.2578</td>
<td>0.3640</td>
<td>0.0790</td>
</tr>
<tr>
<td>13</td>
<td>0.0791</td>
<td>0.2574</td>
<td>0.3634</td>
<td>0.0792</td>
</tr>
<tr>
<td>14</td>
<td>0.0791</td>
<td>0.2579</td>
<td>0.3637</td>
<td>0.0791</td>
</tr>
<tr>
<td>15</td>
<td>0.0791</td>
<td>0.2577</td>
<td>0.3641</td>
<td>0.0791</td>
</tr>
<tr>
<td>16</td>
<td>0.0791</td>
<td>0.2578</td>
<td>0.3641</td>
<td>0.0791</td>
</tr>
<tr>
<td>17</td>
<td>0.0792</td>
<td>0.2577</td>
<td>0.3644</td>
<td>0.0791</td>
</tr>
</tbody>
</table>
IMPROVING PERFORMANCE STATISTICS OF MRR2

Results

The values of $SSE_{max}$ for $LLR$ for $y_1$, $y_2$, $y_3$, and $y_4$ are 0.1638, 942.9793, 234.8291, and 352.1950 respectively, and those for $MRR2$ are 0.0231, 90.9033, 41.1338, and 198.8048, respectively. The optimal Local bandwidths for each response generated for a given value of $N$ in N-squared function are given in Table 2, and Table 3 for $LLR_{VB1}$ and $MMR2_{VB1}$ respectively.

Tables 4–6 present the results of numerical values of performance statistics for comparing OLS, $LLR$ for both fixed bandwidth and local bandwidth generated via N-squared function, and MRR2 for both fixed bandwidth and local bandwidth generated via N-squared function. For convenience, $LLR$ and MRR2 for fixed bandwidth reported in Wan (2007) are referred to as $LLR_{FB}$ and $MMR2_{FB}$ respectively while $LLR$ and MRR2 for N-squared variable bandwidths function are designated $LLR_{VB1}$ and $MMR2_{VB1}$ respectively. The case where the values of the mixing parameters for all responses are all 1 is considered for comparison sake. This will enable one attribute the performance of the models solely to the type of bandwidth used rather than to values of the mixing parameters. Best values for each performance statistics and for each response are shown in bold print.

Table 4 compares the performance statistics of fitted responses from the three regression methods discussed here. $LLR_{FB}$ produces best results exclusively in zero cell and joint best result in zero cell. $MMR2_{FB}$ produces best result exclusively in zero cell and joint best results in zero cell. OLS produces best results exclusively in three cells and joint best result in zero cell. $LLR_{VB1}$ produces best results exclusively in six cells and joint best results in six points. $MMR2_{VB1}$ produces best results exclusively in nine cells and joint best results in six cells. $MMR2_{VB1}$ produces the smallest $S^2$, highest $R^2$ and $R^2_{adj}$ exclusively across two of the responses and joint best results for these statistics in the remaining two responses. For $DF_{error}$, $MMR2_{VB1}$ produces either the best or competitive results across all responses. In addition, $MMR2_{VB1}$ produces competitive results in several cells where it fails to produce the best results. Table 5 compares the performance statistics of fitted responses from the two versions of local linear regression, $LLR_{FB}$ and $LLR_{VB1}$.

$LLR_{FB}$ produces best results in just five cells in a total of twenty-four cells and $LLR_{VB1}$ produces best results in nineteen cells which is equivalent to 79.17% of the total cells for comparison. Table 6 compares the performance statistics of fitted responses from the two versions of model-robust regression, $MMR2_{FB}$ and $MMR2_{VB1}$. $MMR2_{FB}$ produces best results exclusively in just one cell and joint best results in zero cell. $MMR2_{VB1}$ produces best results exclusively in twenty-three cells and best results in zero cell which is equivalent to 95.83% of the total cells for
comparison. Figures 1 through 4 present the plots of each response against $x_1$ for various values of $x_2$ and $x_3$ were applicable. The $MMR2_{FB}$ and $MMR2_{VB1}$ overlap in virtually all the plots for $y_1$ and $y_4$ reflecting the closeness of several performance statistics for the two approaches. However, in $y_2$ and $y_3$ plots, the $MMR2_{VB1}$ plots are seen passing through the mean values of the responses for instance plot in the top right in Figure 4.

The results in Tables 4 and 6 clearly show that $MMR2_{VB1}$ is the overall best regression technique outperforming the $MMR2_{FB}$ that produces best results in sixteen cells out of twenty-four cells in results presented in wan (2007) where it is compared with OLS and $LLR_{FB}$. However, in situations where the user has no prior knowledge of the true underlying model $LLR_{VB1}$ will certainly come in handy as results in Table 4 reveal.

Table 4. Results of comparison of performance statistics of OLS, fixed bandwidth LLR and MRR2, and Variable bandwidth LLR and MRR2 all for $\lambda = 1$, fixed optimal bandwidth, $b$, and $N$ as defined in equation (37) for local bandwidths in Table 2, and Table 3

<table>
<thead>
<tr>
<th>METHOD</th>
<th>$b$</th>
<th>$N$</th>
<th>$DF_{error}$</th>
<th>$S^2$</th>
<th>$R$</th>
<th>$R^2_{adj}$</th>
<th>PRESS</th>
<th>PRESS*</th>
<th>PRESS**</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>OLS</td>
<td>-</td>
<td>14.0000</td>
<td>1.65E-03</td>
<td>0.9211</td>
<td>0.9090</td>
<td>0.0582</td>
<td>0.0042</td>
<td>0.0042</td>
</tr>
<tr>
<td></td>
<td>LLRF</td>
<td>0.146</td>
<td>12.1381</td>
<td>1.04E-03</td>
<td>0.9570</td>
<td>0.9433</td>
<td>0.0682</td>
<td>0.0056</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>MRR2FB</td>
<td>0.17</td>
<td>12.2680</td>
<td>1.03E-03</td>
<td>0.9568</td>
<td>0.9436</td>
<td>0.0472</td>
<td>0.0039</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>LLRF</td>
<td>1.362</td>
<td>12.0000</td>
<td>1.00E-03</td>
<td>0.9579</td>
<td>0.9439</td>
<td>0.0216</td>
<td>0.0018</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>MRR2VB</td>
<td>1.348</td>
<td>12.0000</td>
<td>1.00E-03</td>
<td>0.9579</td>
<td>0.9439</td>
<td>0.0405</td>
<td>0.0034</td>
<td>0.0021</td>
</tr>
<tr>
<td>$y_2$</td>
<td>OLS</td>
<td>-</td>
<td>12.0000</td>
<td>7.5417</td>
<td>0.9341</td>
<td>0.9122</td>
<td>234.1166</td>
<td>19.5097</td>
<td>19.5097</td>
</tr>
<tr>
<td></td>
<td>LLRF</td>
<td>0.436</td>
<td>11.2120</td>
<td>21.8508</td>
<td>0.8217</td>
<td>0.7456</td>
<td>785.7855</td>
<td>70.0873</td>
<td>36.4222</td>
</tr>
<tr>
<td></td>
<td>MRR2FB</td>
<td>0.277</td>
<td>8.9400</td>
<td>4.8253</td>
<td>0.9686</td>
<td>0.9438</td>
<td>319.3332</td>
<td>35.7214</td>
<td>19.6311</td>
</tr>
<tr>
<td></td>
<td>LLRF</td>
<td>7.441</td>
<td>11.2260</td>
<td>21.9206</td>
<td>0.8209</td>
<td>0.7448</td>
<td>785.9495</td>
<td>70.0115</td>
<td>36.4328</td>
</tr>
<tr>
<td></td>
<td>MRR2VB</td>
<td>4.366</td>
<td>8.6923</td>
<td>4.6819</td>
<td>0.9704</td>
<td>0.9455</td>
<td>305.1765</td>
<td>35.1090</td>
<td>18.5803</td>
</tr>
<tr>
<td>$y_3$</td>
<td>OLS</td>
<td>-</td>
<td>9.0000</td>
<td>4.5641</td>
<td>0.8408</td>
<td>0.7170</td>
<td>182.4468</td>
<td>20.2719</td>
<td>20.2719</td>
</tr>
<tr>
<td></td>
<td>LLRF</td>
<td>0.537</td>
<td>8.3730</td>
<td>9.7990</td>
<td>0.6821</td>
<td>0.3925</td>
<td>287.0564</td>
<td>34.2849</td>
<td>17.0554</td>
</tr>
<tr>
<td></td>
<td>MRR2FB</td>
<td>0.542</td>
<td>6.5960</td>
<td>2.9031</td>
<td>0.9258</td>
<td>0.8200</td>
<td>177.6750</td>
<td>26.9357</td>
<td>13.1264</td>
</tr>
<tr>
<td></td>
<td>LLRF</td>
<td>9.121</td>
<td>8.3672</td>
<td>9.7791</td>
<td>0.6829</td>
<td>0.3937</td>
<td>286.6772</td>
<td>34.2622</td>
<td>17.0261</td>
</tr>
<tr>
<td></td>
<td>MRR2VB</td>
<td>6.179</td>
<td>9.3265</td>
<td>1.3817</td>
<td>0.9790</td>
<td>0.9143</td>
<td>173.9599</td>
<td>44.3046</td>
<td>11.4358</td>
</tr>
<tr>
<td>$y_4$</td>
<td>OLS</td>
<td>-</td>
<td>14.0000</td>
<td>14.2182</td>
<td>0.5407</td>
<td>0.4751</td>
<td>684.7407</td>
<td>48.9101</td>
<td>48.9101</td>
</tr>
<tr>
<td></td>
<td>LLRF</td>
<td>0.12</td>
<td>12.0310</td>
<td>1.0197</td>
<td>0.9717</td>
<td>0.9624</td>
<td>454.5871</td>
<td>37.7832</td>
<td>17.1484</td>
</tr>
<tr>
<td></td>
<td>MRR2FB</td>
<td>0.119</td>
<td>12.0290</td>
<td>1.0158</td>
<td>0.9718</td>
<td>0.9625</td>
<td>486.8458</td>
<td>40.4725</td>
<td>18.6472</td>
</tr>
<tr>
<td></td>
<td>LLRF</td>
<td>1.361</td>
<td>12.0000</td>
<td>1.0116</td>
<td>0.9720</td>
<td>0.9627</td>
<td>407.8131</td>
<td>33.9844</td>
<td>15.3990</td>
</tr>
<tr>
<td></td>
<td>MRR2VB</td>
<td>1.347</td>
<td>12.0000</td>
<td>1.0116</td>
<td>0.9720</td>
<td>0.9627</td>
<td>451.5303</td>
<td>37.6275</td>
<td>17.3105</td>
</tr>
</tbody>
</table>
**Table 5.** \(LLR_{FB} \) versus \(LLR_{VB1} \) for values of \(b \) and \(b_i \) indicated in Table 2

<table>
<thead>
<tr>
<th>METHOD</th>
<th>(DF_{error} )</th>
<th>(S^2 )</th>
<th>(R )</th>
<th>(R^2_{adj} )</th>
<th>PRESS</th>
<th>PRESS'</th>
<th>PRESS''</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_1^1 )</td>
<td>(LLR_{FB} )</td>
<td>12.1380</td>
<td>1.04E-03</td>
<td>0.9570</td>
<td>0.9433</td>
<td>0.0682</td>
<td>0.0056</td>
</tr>
<tr>
<td></td>
<td>(LLR_{VB1} )</td>
<td>12.0000</td>
<td>1.00E-03</td>
<td>0.9579</td>
<td>0.9439</td>
<td>0.0216</td>
<td>0.0018</td>
</tr>
<tr>
<td>(y_2^1 )</td>
<td>(LLR_{FB} )</td>
<td>11.2120</td>
<td>21.8508</td>
<td>0.8217</td>
<td>0.7456</td>
<td>785.7855</td>
<td>70.0873</td>
</tr>
<tr>
<td></td>
<td>(LLR_{VB1} )</td>
<td>11.2200</td>
<td>21.9206</td>
<td>0.8209</td>
<td>0.7448</td>
<td>785.9495</td>
<td>70.0115</td>
</tr>
<tr>
<td>(y_3^1 )</td>
<td>(LLR_{FB} )</td>
<td>8.3730</td>
<td>9.7990</td>
<td>0.6821</td>
<td>0.3925</td>
<td>287.0564</td>
<td>34.2849</td>
</tr>
<tr>
<td></td>
<td>(LLR_{VB1} )</td>
<td>8.3672</td>
<td>9.7791</td>
<td>0.6829</td>
<td>0.3937</td>
<td>286.6772</td>
<td>34.2622</td>
</tr>
<tr>
<td>(y_4^1 )</td>
<td>(LLR_{FB} )</td>
<td>12.0310</td>
<td>1.0197</td>
<td>0.9717</td>
<td>0.9624</td>
<td>454.5871</td>
<td>37.7832</td>
</tr>
<tr>
<td></td>
<td>(LLR_{VB1} )</td>
<td>12.0000</td>
<td>1.0116</td>
<td>0.9720</td>
<td>0.9627</td>
<td>407.8131</td>
<td>33.9844</td>
</tr>
</tbody>
</table>

**Table 6.** \(MMR2_{FB} \) versus \(MMR2_{VB1} \) for values of \(b \) and \(b_i \) indicated in Table 3.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>(DF_{error} )</th>
<th>(S^2 )</th>
<th>(R )</th>
<th>(R^2_{adj} )</th>
<th>PRESS</th>
<th>PRESS'</th>
<th>PRESS''</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_1^1 )</td>
<td>(MMR2_{FB} )</td>
<td>12.2680</td>
<td>1.03E-03</td>
<td>0.9568</td>
<td>0.9436</td>
<td>0.0472</td>
<td>0.0039</td>
</tr>
<tr>
<td></td>
<td>(MMR2_{VB1} )</td>
<td>12.0000</td>
<td>1.00E-03</td>
<td>0.9579</td>
<td>0.9439</td>
<td>0.0405</td>
<td>0.0034</td>
</tr>
<tr>
<td>(y_2^1 )</td>
<td>(MMR2_{FB} )</td>
<td>8.9400</td>
<td>4.8253</td>
<td>0.9686</td>
<td>0.9438</td>
<td>319.3332</td>
<td>35.7214</td>
</tr>
<tr>
<td></td>
<td>(MMR2_{VB1} )</td>
<td>8.6923</td>
<td>4.6819</td>
<td>0.9704</td>
<td>0.9455</td>
<td>305.1765</td>
<td>35.1090</td>
</tr>
<tr>
<td>(y_3^1 )</td>
<td>(MMR2_{FB} )</td>
<td>6.5960</td>
<td>2.9031</td>
<td>0.9258</td>
<td>0.8200</td>
<td>177.6750</td>
<td>26.9357</td>
</tr>
<tr>
<td></td>
<td>(MMR2_{VB1} )</td>
<td>3.9265</td>
<td>1.3817</td>
<td>0.9790</td>
<td>0.9143</td>
<td>173.9599</td>
<td>44.3046</td>
</tr>
<tr>
<td>(y_4^1 )</td>
<td>(MMR2_{FB} )</td>
<td>12.0290</td>
<td>1.0158</td>
<td>0.9718</td>
<td>0.9625</td>
<td>486.8458</td>
<td>40.4725</td>
</tr>
<tr>
<td></td>
<td>(MMR2_{VB1} )</td>
<td>12.0000</td>
<td>1.0116</td>
<td>0.9720</td>
<td>0.9627</td>
<td>451.5303</td>
<td>37.6275</td>
</tr>
</tbody>
</table>

Figures 1 and 2 compare the plots of \(\hat{y}_1 \) versus \(x_1 \) and \(\hat{y}_4 \) versus \(x_1 \) respectively, using OLS, MRR2 via fixed bandwidth \(MMR2_{FB} \), and MRR2 via local bandwidths \(MMR2_{VB1} \) from N-squared function.

![Figure 1. Plot of \(\hat{y}_1 \) vs. \(x_1 \)](image1)

![Figure 2. Plot of \(\hat{y}_4 \) vs. \(x_1 \)](image2)
Figure 3 compares the plots of $\hat{y}_1$ versus $x_1$ for OLS, $MRR2_{(FB)}$, and $MRR2_{(VB)}$, when $x_2 = 0$ (left), $x_2 = 0.5$ (center), and $x_2 = 1$ (right).

Figure 3. Plots of $\hat{y}_2$ versus $x_1$

Figure 4 compares the plots of $\hat{y}_2$ versus $x_1$ for OLS, $MRR2_{(FB)}$, and $MRR2_{(VB)}$, for all respective values of $x_2$ and $x_3$ specified.
Conclusion

One of the shortcomings of parametric regression is that the user has to specify a model that perfectly fits the data under consideration and failure to achieve this leads to highly biased estimates. Nonparametric regression is usually employed when the user is unable to specify a model for the data. However, in studies that require small-sample data such as RSM, nonparametric tends to produce fitted values that are highly variable. Semi-parametric regression such as MRR2 technique which combine parametric regression with a nonparametric technique are employed in scenarios where there is partial knowledge of the underlying model.
for small-sample data. Both the nonparametric and semi-parametric methods require a parameter referred to as smoothing parameter or bandwidth which determines the smoothness of the estimates.

Regression methods for fitting data suitable for RSM were reviewed. Also reviewed are methods for selecting local bandwidth. A new methodology for deriving a function was presented. The function, herein referred to as N-squared function, was employed for generating data-driven local bandwidths and MRR2 technique utilizing local bandwidth derived from the N-squared function was applied to the multi-response problem of minced fish quality and the results of performance statistics of fitted responses was compared with the results for performance statistics for MRR2 utilizing fixed bandwidth reported in Wan (2007). The comparisons presented in Tables 4−6 show the superiority of fits from local bandwidths derived from N-squared function over fits obtained using fixed bandwidth. Indeed, these results are confirmation of statements made Wan (2007), Mays (2001a), and several other researchers regarding improvement that MRR2 and other semi-parametric methods stand to gain if performed using suitable local bandwidths.

References


