Journal of Modern Applied Statistical Methods

Volume 13 | Issue 2 Article 30

11-2014

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

Recommended Citation
Edionwe, Efosa and Mbegbu, Julian L. (2014) "Local Bandwidths for Improving Performance Statistics of Model-Robust Regression 2," *Journal of Modern Applied Statistical Methods*: Vol. 13 : Iss. 2 , Article 30.
DOI: 10.22237/jmasm/1414816140

Journal of Modern Applied Statistical Methods
November 2014, Vol. 13, No. 2, 506-527.

ISSN 1538 − 9472

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

Mr. Edionwe is a Ph.D student in the Department of Mathematics. Email him trustnelson24@yahoo.com. Dr. Mbegbu is an Associate Professor in the Department of Mathematics. Email him at julian.mbegbu@yahoo.com.

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

$$y_i = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i>j} \beta_{ij} x_i x_j + \varepsilon_i \tag{1}$$

is assumed for fitting the data, where β_i, β_{ii} and β_{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

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{2}$$

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

$$\hat{\mathbf{y}}^{(OLS)} = \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathbf{OLS}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}^{(\mathbf{OLS})}\mathbf{y} \tag{3}$$

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 $\tilde{\boldsymbol{x}}_0$ is given as

$$\hat{y}_0^{(LLR)} = \tilde{x}_0' \left(\tilde{\mathbf{X}}' \mathbf{W_0} \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}' \mathbf{W_0} \mathbf{y} = \mathbf{h}_0^{(LLR)} \mathbf{y} \tag{4}$$

$$\tilde{\boldsymbol{X}} = \begin{bmatrix} \tilde{\boldsymbol{x}}_1' \\ \tilde{\boldsymbol{x}}_2' \\ \vdots \\ \tilde{\boldsymbol{x}}_n' \end{bmatrix} \tag{5}$$

where $\tilde{\mathbf{x}}_\mathbf{1}' = \left(1\, x_{i1}, \ldots, x_{ik}\right)$, the n by n diagonal matrix $\mathbf{W}_0$, known as the local weight matrix for location $\mathbf{x}_0$, is given by

$$\mathbf{W_0} = diag\left(h_{01}^{(KER)} \quad h_{02}^{(KER)} \quad \cdots \quad h_{04}^{(KER)} \right) \tag{6}$$

$h_{0i}^{(KER)}$ represents a kernel weight assigned to y_i in the estimation of y_0 at location $\mathbf{x}_0$ and is given as

$$h_{0i}^{(KER)} = \frac{K\left(\frac{x_0 - x_i}{b}\right)}{\sum_{i=1}^{n} K\left(\frac{x_0 - x_i}{b}\right)} \tag{7}$$

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

$$K\left(\frac{x_0 - x_i}{b}\right) = e^{-\left(\frac{x_0 - x_i}{b}\right)^2} \tag{8}$$

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

$$K\left(\tilde{x}_0, \tilde{x}_1\right) = \prod_{j=1}^{k} K\left(\frac{\tilde{x}_{0j} - \tilde{x}_{ij}}{b}\right) \tag{9}$$

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

$$\hat{\mathbf{y}}^{(LLR)} = \mathbf{H}^{(LLR)}\mathbf{y} \tag{10}$$

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

$$H^{(LLR)} = \begin{bmatrix} \boldsymbol{h}_1^{(LLR)\prime} \\ \boldsymbol{h}_2^{(LLR)\prime} \\ \vdots \\ \boldsymbol{h}_n^{(LLR)\prime} \end{bmatrix} \tag{11}$$

$$\mathbf{h}_i^{(LLR)\prime} = \tilde{\mathbf{x}}_{\mathbf{i}}^{\prime}\left(\tilde{\mathbf{X}}^{\prime}\mathbf{W}_{\mathbf{i}}\tilde{\mathbf{X}}^{\prime}\right)^{-1}\tilde{\mathbf{X}}^{\prime}\mathbf{W}_{\mathbf{i}} \tag{12}$$

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.

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

$$\hat{y}_0^{(MRR2)} = \breve{\mathbf{x}}_{\mathbf{0}}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} + \boldsymbol{\lambda}\tilde{\mathbf{x}}_{\mathbf{0}}'\left(\tilde{\mathbf{X}}'\mathbf{W}_{\mathbf{r}}\tilde{\mathbf{X}}'\right)^{-1}\tilde{\mathbf{X}}'\mathbf{W}_{\mathbf{r}}\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\mathbf{y} \tag{13}$$

where

$$\mathbf{h}_0^{(MRR2)\prime} = \mathbf{x}_{\mathbf{0}}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' + \boldsymbol{\lambda}\tilde{\mathbf{x}}_{\mathbf{0}}'\left(\tilde{\mathbf{X}}'\mathbf{W}_{\mathbf{r}}\tilde{\mathbf{X}}'\right)^{-1}\tilde{\mathbf{X}}'\mathbf{W}_{\mathbf{r}}\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right),$$

$\mathbf{r}$ is the vector of residuals from the parametric fits, $\mathbf{I}$ is an n by n identity matrix, $\mathbf{W_r}$ 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 $\mathbf{X}$ and $\tilde{\mathbf{X}}$ are as defined in (2) and (5) respectively. MRR2 fits is expressed in matrix form as

$$\hat{\mathbf{y}}^{(MRR2)} = \left[\mathbf{H}^{(\mathbf{OLS})} + \boldsymbol{\lambda}\mathbf{H}^{(\mathbf{LLR})}\left(\mathbf{I} - \mathbf{H}^{(\mathbf{OLS})}\right)\right]\mathbf{y} \tag{14}$$

$$= \mathbf{H}^{(\mathbf{MRR2})}\mathbf{y} \tag{15}$$

where $\mathbf{H}^{(\mathbf{MMR2})}$ is the MRR2 "HAT" matrix defined as

$$\boldsymbol{H}^{(\boldsymbol{MRR2})} = \begin{bmatrix} \boldsymbol{h}_1^{(MRR2)\prime} \\ \boldsymbol{h}_2^{(MRR2)\prime} \\ \vdots \\ \boldsymbol{h}_n^{(MRR2)\prime} \end{bmatrix} \tag{16}$$

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

$$\hat{\lambda}^* = \frac{\langle \hat{r}, r \rangle}{\|\hat{r}\|^2} \tag{17}$$

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

$$PRESS^{**}(\lambda) = \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_{i,-1}\left(b^*, \lambda\right)\right)^2}{n - tr\left(H^{(MRR2)}\left(b^*, \lambda\right)\right) + (n-k-1)\dfrac{SSE_{\max} - SSE_{b^*}}{SSE_{\max}}} \tag{18}$$

is minimized. n is the sample size, b^* is the optimal bandwidth, k is the number explanatory variables, $SSE_{\max}$ is the maximum sum of squared errors obtained as b tends to infinity, SSE_{b^*} is the sum of squared errors for the optimal bandwidth, $tr\,(H^{(MRR2)}(b^*,\lambda))$ is the trace of the MRR2 "HAT" matrix for a given b^* and λ, and $\hat{y}_{i,-i}$ is the fit at $\boldsymbol{x}_i$ 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, 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

$$PRESS^{**}(b) = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{i,-1}(b)\right)^2}{n - tr\left(H^{(MRR2)}(b)\right) + (n-k-1)\dfrac{SSE_{\max} - SSE_b}{SSE_{\max}}} \tag{19}$$

$$PRESS^{**}(b) = \sum_{i=1}^{n}\left(y_i - \hat{y}_{i,-1}(b)\right)^2, \; i = 1, 2, \ldots n \tag{20}$$

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

$$\alpha_{opt}(x) = \begin{cases} b\left(\dfrac{f_X(x)\left[m''(x)\right]^2}{\sigma^2(x)}\right)^{\frac{1}{5}}, & \text{if } W(x) > 0, \\ \alpha^*(x), & \text{if } W(x) = 0 \end{cases} \tag{21}$$

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

$$h_{opt}^{SCH}(x)=\left(\frac{\sigma^2 A}{2pnB(x)^2}\right)^{1/(2p+1)} \tag{22}$$

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, $h_{opt}^{SCH}(x)$ is calculated using estimates of σ^2 and $B(x)$. Hence, the quality of the final estimator $\hat{h}_{opt}^{SCH}$ 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

$$h^{**}(x)=\arg\left\{\min_h \sum_{i=1}^{l(x)}\left(Y_i'-\hat{m}_{-i}(X_i')\right)^2\right\} \tag{23}$$

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

$$\hat{m}_{-i}(X_i)=\frac{1}{(n-1)h}\sum_{j\neq i}K\left(\frac{x_i-x_j}{h}\right)\frac{M_{2i}(x_i)-\left(\frac{x_i-x_j}{h}\right)M_{1i}(x_i)}{M_{2i}(x_i)M_{0i}(x_i)-M_{1i}^2(x_i)}Y_j \tag{24}$$

and

$$M_{ci}\left(X_i\right)=\left(\left(n-1\right)h\right)^{-1}\sum_{j\neq i}K\left(\frac{x_i-x_j}{h}\right)\left(\frac{x_i-x_j}{h}\right)^c,\ \text{for } c = 1, 2, 3 \qquad (25)$$

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

$$b_i=\frac{y_i}{T},\ \text{for } i = 1, \ldots, n. \qquad (26)$$

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

$$b_i = \left(N - \frac{y_i}{T} \right) \tag{27}$$

Taking sum of both sides of (27) gives

$$\sum_i^n b_i = \sum_1^n \left(N - \frac{y_i}{T} \right) \tag{28}$$

$$\sum_i^n b_i = \sum_1^n \left(\frac{NT - y_i}{T} \right) \tag{29}$$

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)

$$\sum_i^n b_i = \left(\frac{(NT - y_1) + (NT - y_2) + \cdots + (NT - y_n)}{T} \right) \tag{30}$$

On collecting like terms in (30)

$$\sum_i^n b_i = \frac{NTn - \sum_{i=1}^n y_i}{T} \tag{31}$$

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

$$\sum_i^n b_i = \frac{NTn - T}{T} \tag{32}$$

Therefore

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

which implies $\sum_{1}^{n}\left(\frac{NT - y_i}{T}\right) = (Nn - 1)$. 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

$$\sum_{1}^{n} b_i = \sum_{1}^{n}\left(\frac{NT - y_i}{T}\right) N / (Nn - 1) \tag{34}$$

$$\sum_{1}^{n} b_i = \sum_{1}^{n} \frac{N(NT - y_i)}{T(Nn - 1)} \tag{35}$$

Therefore

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

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

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 $PRESS^{**}(\mathbf{b})$ given by

$$PRESS^{**}(\mathbf{b}) = \frac{\sum_{i}^{n}\left(y_i - y_{i,-i}(\mathbf{b})\right)^2}{n - trace\left(H^{(\cdot)}(\mathbf{b})\right) + (n - k - 1)\dfrac{(SSE_{\max} - SSE_{\mathbf{b}})}{SSE_{\max}}} \tag{38}$$

$$= \frac{PRESS(\mathbf{b})}{n - trace\left(H^{(\cdot)}(\mathbf{b})\right) + (n-k-1)\frac{(SSE_{\max} - SSE_{\mathbf{b}})}{SSE_{\max}}} \tag{39}$$

where $H^{(\cdot)}$ is the "HAT" matrix for (LLR) or MRR2 obtained by using local bandwidths from N-squared function, $SSE_{\max}$ is the maximum sum of squares of errors over all possible bandwidths which is equivalent to $\sum_{i=1}^{n}\left(y_i - \hat{y}_i^{(OLS)}\right)^2$ for LLR or $\sum_{i=1}^{n}\left(e_i - \hat{e}_i^{(OLS)}\right)^2$ for MRR2 where $\hat{\mathbf{y}}^{(OLS)}$ and $\hat{\mathbf{e}}^{(OLS)}$ are the OLS fit of a first-order model for responses and OLS residuals respectively, SSE_{b_i} is given by $\sum_{i=1}^{n}\left(y_i - \hat{y}_i^{(LLR)}(b_i)\right)^2$ for LLR or $\sum_{i=1}^{n}\left(y_i - \hat{y}_i^{(MMR2)}(b_i)\right)^2$ for the MRR2 counterpart. (See Wan, 2007). For MRR2, the mixing parameter, λ, 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^2_{adj}), PRESS given in (20), $PRESS^* = PRESS/DF_{error}$, where $DF_{error} = DF_{total} - DF_{model}$, 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

	Coded Variables			Response values			
	x_1	x_2	x_3	y_1	y_2	y_3	y_4
1	0.203	0.203	0.203	1.83	29.31	29.50	50.36
2	0.797	0.203	0.203	1.73	39.32	19.40	48.16
3	0.203	0.797	0.203	1.85	25.16	25.70	50.72
4	0.797	0.797	0.203	1.67	40.18	27.10	49.69
5	0.203	0.203	0.797	1.86	29.82	21.40	50.09
6	0.797	0.203	0.797	1.77	32.20	24.00	50.61
7	0.203	0.797	0.797	1.88	22.01	19.60	50.36
8	0.797	0.797	0.797	1.66	40.02	25.10	50.42
9	0	0.5	0.5	1.81	33.00	24.20	29.31
10	1	0.5	0.5	1.37	51.59	30.60	50.67
11	0.5	0	0.5	1.85	20.35	20.90	48.75
12	0.5	1	0.5	1.92	20.53	18.90	52.70
13	0.5	0.5	0	1.88	23.85	23.00	50.19
14	0.5	0.5	1	1.90	20.16	21.20	50.86
15	0.5	0.5	0.5	1.89	21.72	18.50	50.84
16	0.5	0.5	0.5	1.88	21.21	18.60	50.93
17	0.5	0.5	0.5	1.87	21.55	16.80	50.98

Table 2. Bandwidths for LLR_{VB1} for each location and response

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

Table 3. Bandwidths for $MMR2_{VB1}$ for each location and response

Location	y_1	y_2	y_3	y_4
1	0.0792	0.2568	0.3624	0.0791
2	0.0794	0.2556	0.3640	0.0793
3	0.0792	0.2573	0.3630	0.0791
4	0.0796	0.2555	0.3628	0.0792
5	0.0792	0.2567	0.3637	0.0792
6	0.0794	0.2564	0.3633	0.0791
7	0.0791	0.2577	0.3639	0.0791
8	0.0796	0.2555	0.3631	0.0791
9	0.0793	0.2563	0.3632	0.0807
10	0.0802	0.2541	0.3622	0.0791
11	0.0792	0.2579	0.3637	0.0793
12	0.0791	0.2578	0.3640	0.0790
13	0.0791	0.2574	0.3634	0.0792
14	0.0791	0.2579	0.3637	0.0791
15	0.0791	0.2577	0.3641	0.0791
16	0.0791	0.2578	0.3641	0.0791
17	0.0792	0.2577	0.3644	0.0791

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 λ = 1, fixed optimal bandwidth, *b*, and *N* as defined in equation (37) for local bandwidths in Table 2, and Table 3

	METHOD	*b*	*N*	DF_{error}	S^2	*R*	R^2_{adj}	*PRESS*	*PRESS**	*PRESS***
	OLS	-	-	14.0000	1.65E-03	0.9211	0.9090	0.0582	0.0042	0.0042
	LLR_{FB}	0.146		12.1381	1.04E-03	0.9570	0.9433	0.0682	0.0056	0.0026
y_1	$MRR2_{FB}$	0.17		12.2680	1.03E-03	0.9568	0.9436	0.0472	0.0039	0.0025
	LLR_{VB1}		1.362	12.0000	**1.00E-03**	**0.9579**	**0.9439**	0.0216	0.0018	**0.0008**
	$MRR2_{VB1}$		1.348	12.0000	**1.00E-03**	**0.9579**	**0.9439**	0.0405	0.0034	0.0021
	OLS	-	-	12.0000	7.5417	0.9341	0.9122	**234.1166**	**19.5097**	19.5097
	LLR_{FB}	0.436		11.2120	21.8508	0.8217	0.7456	785.7855	70.0873	36.4222
y_2	$MRR2_{FB}$	0.277		8.9400	4.8253	0.9686	0.9438	319.3332	35.7214	19.6311
	LLR_{VB1}		7.441	11.2260	21.9206	0.8209	0.7448	785.9495	70.0115	36.4328
	$MRR2_{VB1}$		4.366	8.6923	**4.6819**	**0.9704**	**0.9455**	305.1765	35.1090	**18.5803**
	OLS	-	-	9.0000	4.5641	0.8408	0.7170	182.4468	**20.2719**	20.2719
	LLR_{FB}	0.537		8.3730	9.7990	0.6821	0.3925	287.0564	34.2849	17.0554
y_3	$MRR2_{FB}$	0.542		6.5960	2.9031	0.9258	0.8200	177.6750	26.9357	13.1264
	LLR_{VB1}		9.121	8.3672	9.7791	0.6829	0.3937	286.6772	34.2622	17.0261
	$MRR2_{VB1}$		6.179	3.9265	**1.3817**	**0.9790**	**0.9143**	**173.9599**	44.3046	**11.4358**
	OLS	-	-	14.0000	14.2182	0.5407	0.4751	684.7407	48.9101	48.9101
	LLR_{FB}	0.12		12.0310	1.0197	0.9717	0.9624	454.5871	37.7832	17.1484
y_4	$MRR2_{FB}$	0.119		12.0290	1.0158	0.9718	0.9625	486.8458	40.4725	18.6472
	LLR_{VB1}		1.361	12.0000	**1.0116**	**0.9720**	**0.9627**	**407.8131**	**33.9844**	**15.3990**
	$MRR2_{VB1}$		1.347	12.0000	1.0116	**0.9720**	**0.9627**	451.5303	37.6275	17.3105

Table 5. LLR_{FB} versus LLR_{VB1} for values of b and b_i indicated in Table 2

	METHOD	DF_{error}	S^2	R	R^2_{adj}	*PRESS*	$PRESS^*$	$PRESS^{**}$
y_1	LLR_{FB}	12.1380	1.04E-03	0.9570	0.9433	0.0682	0.0056	0.0026
	LLR_{VB1}	12.0000	**1.00E-03**	**0.9579**	**0.9439**	0.0216	0.0018	**0.0008**
y_2	LLR_{FB}	11.2120	**21.8508**	**0.8217**	**0.7456**	**785.7855**	70.0873	**36.4222**
	LLR_{VB1}	11.2260	21.9206	0.8209	0.7448	785.9495	**70.0115**	36.4328
y_3	LLR_{FB}	8.3730	9.7990	0.6821	0.3925	287.0564	34.2849	17.0554
	LLR_{VB1}	8.3672	9.7791	**0.6829**	**0.3937**	**286.6772**	**34.2622**	**17.0261**
y_4	LLR_{FB}	12.0310	1.0197	0.9717	0.9624	454.5871	37.7832	17.1484
	LLR_{VB1}	12.0000	**1.0116**	**0.9720**	**0.9627**	**407.8131**	**33.9844**	**15.3990**

Table 6. $MMR2_{FB}$ versus $MMR2_{VB1}$ for values of b and b_i indicated in Table 3.

	METHOD	DF_{error}	S^2	R	R^2_{adj}	*PRESS*	$PRESS^*$	$PRESS^{**}$
y_1	$MRR2_{FB}$	12.2680	1.03E-03	0.9568	0.9436	0.0472	0.0039	0.0025
	$MRR2_{VB1}$	12.0000	**1.00E-03**	**0.9579**	**0.9439**	0.0405	0.0034	**0.0021**
y_2	$MRR2_{FB}$	8.9400	4.8253	0.9686	0.9438	319.3332	35.7214	19.6311
	$MRR2_{VB1}$	8.6923	**4.6819**	**0.9704**	**0.9455**	**305.1765**	**35.1090**	**18.5803**
y_3	$MRR2_{FB}$	6.5960	2.9031	0.9258	0.8200	177.6750	**26.9357**	13.1264
	$MRR2_{VB1}$	3.9265	**1.3817**	**0.9790**	**0.9143**	**173.9599**	44.3046	**11.4358**
y_4	$MRR2_{FB}$	12.0290	1.0158	0.9718	0.9625	486.8458	40.4725	18.6472
	$MRR2_{VB1}$	12.0000	**1.0116**	**0.9720**	**0.9627**	**451.5303**	**37.6275**	**17.3105**

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.

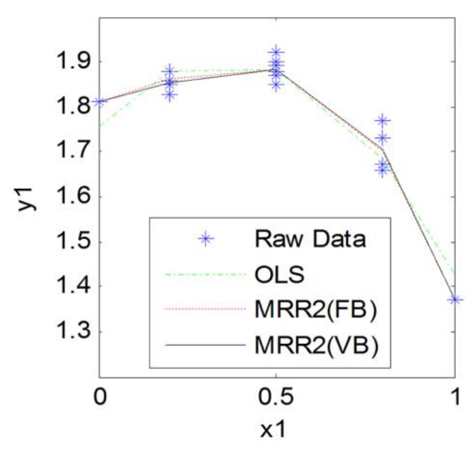

Figure 1. Plot of $\hat{y}_1$ vs. x_1

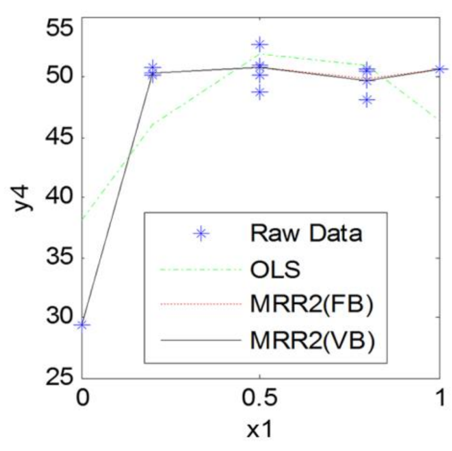

Figure 2. Plot of $\hat{y}_4$ vs. x_1

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).

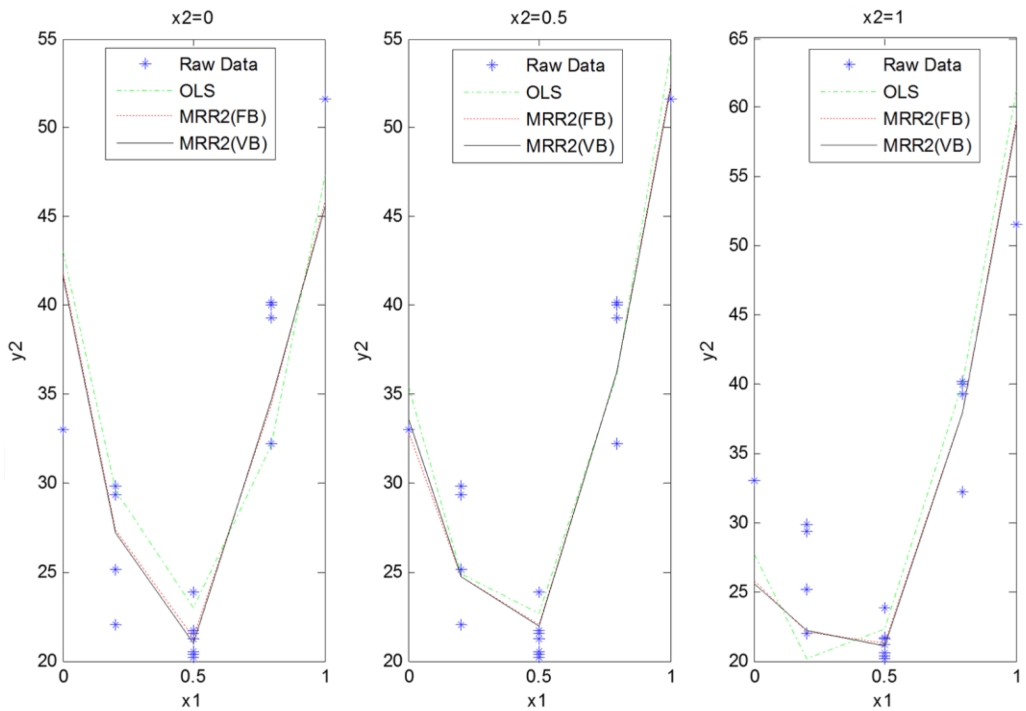

Figure 3. Plots of $\hat{y}_2$ versus x_1

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

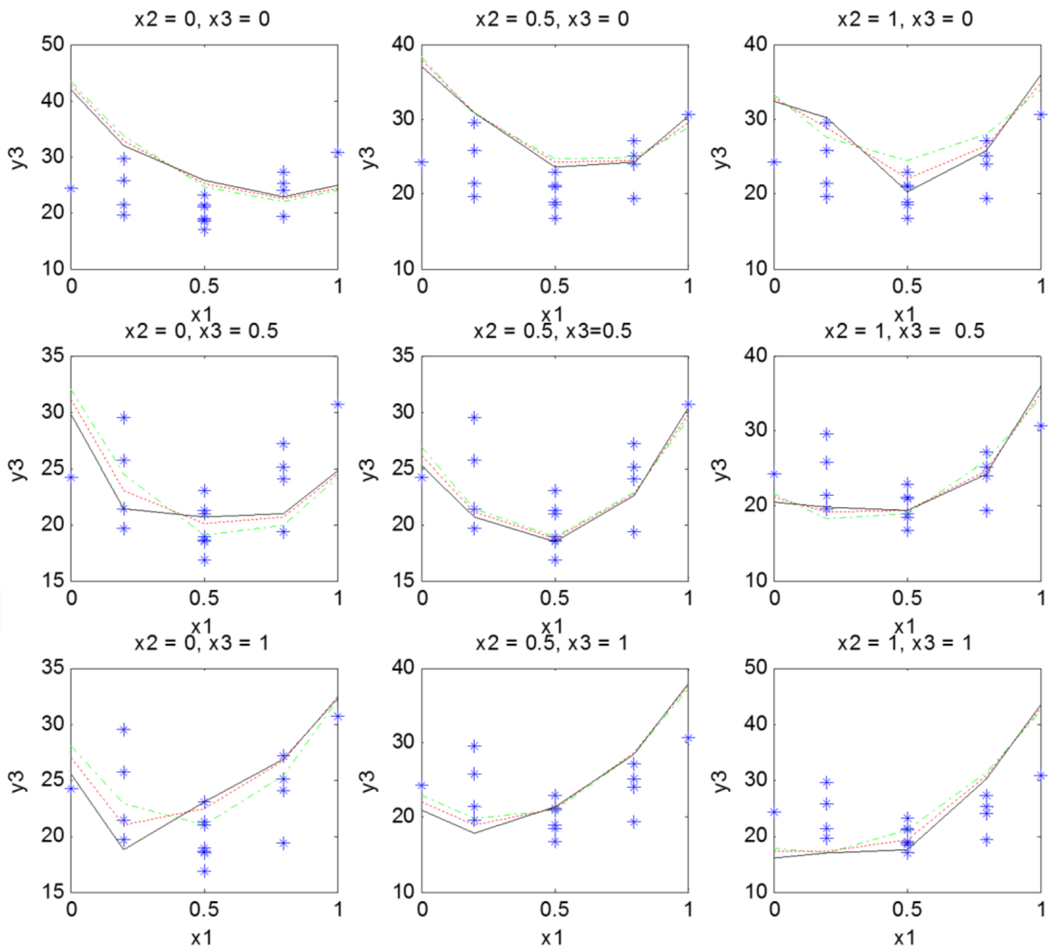

Figure 4. Plots of $\hat{y}_3$ versus x_1

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

Atkeson, C. G., Moore, A. W., & Schaal, S. (1997). Locally weighted learning. *Journal of Artificial Intelligence Review, 11*, 11–73.

Avery, M. (2010). *Literature review for local polynomial regression*. Unpublished manuscript, Department of Statistics, North Carolina State University, Raleigh, NC.

Bickel, P. J., & Li, B. (2007). Local polynomial regression on unknown manifolds. *IMS Lecture Notes Monograph Series, 54*, 177–186.

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

DiNardo, J., & Tobias, J. L. (2001). Nonparametric density and regression estimation. *The Journal of Economic Perspectives, 15*(4), 11–28.

Einsporn, R., & Birch, J. B. (1993). *Model robust regression: Using non-parametric regression to improve parametric regression analysis*. (Technical Report, 93–5). Blacksburg, VA: Department of Statistics, Virginia Polytechnic Institute & State University.

Fan J., & Gibjels I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics, 20*(4), 2008–2036.

Galdo, J., Smith, J. & Black, D.I. (2008). Bandwidth selection and the estimation of treatment effects with unbalanced data. *Annals of Economics and Statistics, 91-92*, 189-216.

Gerard, P. D. & Schucany, W. R. (1999). Local bandwidth selection for kernel estimation of population densities with line transect sampling. *Biometrics, 55*(3), 769-773.

Härdle, W. (1994). *Applied nonparametric regression* (Econometric Society Monographs #19). Cambridge, England: Cambridge University Press.

Hens, N. (2005). *Non- and semiparametric techniques for handling missing data.* (Doctoral dissertation). Diepenbeek, Belgium: Transnationale Universiteit Limburgs.

Hernández–Lobato, J. M. (2010). *Balancing flexibility and robustness in machine learning: Semiparametric methods and sparse linear models.* (Doctoral dissertation). Madrid, Spain: University of Autonoma Department of Computer Science.

Huang, L. S., & Fan J. (1996). Optimal bandwidth selection for local linear regression. *FSU Technical Report, N0. M.912.* Tallahassee, FL: Florida State University.

Kohler, M., Schindler, A. & Sperlich, S. (2011). A review and comparison for bandwidth selection methods for kernel regression. *Courant Research Centre: Poverty, Equity and Growth – Discussion Papers 95*.

Mays, J. E. (2001a). Model-robust confidence intervals and capturing nonconformity points. *Technical Report Number 01-1*, Dept. of Mathematical Sciences, Virginia Commonwealth University.

Mays, J. E. (2001b). Small-sample model-robust confidence intervals in regression. In *Proceedings from Annual Meeting of the American Statistical Association*. Retrieved from http://www.amstat.org/sections/SRMS/Proceedings/y2001/Proceed/00530.pdf

Montgomery, D. C. (1999). Experimental design for product and process design and development. *The Statistician, 48*(2), 159–177.

Pickle, S. M. (2006). *Semiparametric techniques for response surface methodology* (Doctoral dissertation). Virginia Polytechnic Institute and State University: Blacksburg, VA.

Racine, J. S. (2008). Nonparametric econometrics: A primer. *Journal of Foundations and Trends in Econometrics, 3*(1), 1–88.

Raissi, S., & Farsani, R. E. (2009). Statistical process optimization through multiresponse surface methodology. *Journal of World Academy of Science, Engineering and Technology, 3*(3), 250–255.

Schucany, W. (1995). Adaptive bandwidth choice for kernel regression. *Journal of American Statistical Association, 90*(430), 535–540.

Starnes, B. A. (1999). *Asymptotic results for model-robust regression* (Doctoral dissertation). Virginia Polytechnic Institute and State University: Blacksburg, VA.

Starnes, B. A., Birch, J. B., & Robinson, T. (2008, May). *A semiparametric approach to dual modeling with unreplicated data.* Paper presented at the 2008 International Conference on Interdisciplinary Mathematical and Statistical Techniques, Memphis, TN.

Uysal, I., & Guvenir, A. (1999). An overview of regression techniques for knowledge discovery. *The Knowledge Engineering Review, 14*(4), 319–340.

Wan, W. (2007). *Semiparametric techniques for multiresponse optimization* (Doctoral dissertation). Virginia Polytechnic Institute and State University: Blacksburg, VA.

Wu, C. F. J., & Hamada, M. (2000). *Experiments: planning, analysis and parameter design optimization.* New York: John Wiley and Sons.

Zheng, Q., Kulasekera, K. B. & Gallagher, C. (2010). Local adaptive smoothing in kernel regression estimation. *Statistics & Probability Letters, 80*(7-8), 540-547.