International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 14
October 2016

Review of Non Parametric Regression and its application using SAS software

By
KARRAR MOHAMMED OBAID
Department of Statistics, Osmania University
India - Hyderabad - 500007
Email:karar111000@gmail.com

Abstract:

This paper aims to demonstrate the benefits of adopting a nonparametric regression approach when the standard regression model is not appropriate. It also provides an overview of circumstances where a nonparametric approach might not only be beneficial, but necessary. Many statistical methods require assumptions to be made about the format of the data to be analysed. There are situations in which even transformed data may not satisfy the assumptions, however in these cases it may be inappropriate to use traditional parametric methods of analysis. Nonparametric methods provide an alternative series of statistical methods that require no or very limited assumptions to be made about the data. There are a wide range of methods that can be used in different circumstances, but in this study we will describe the practical advantages of nonparametric regression by presenting the fundamental concepts of regression and modelling, and by discussing their desirable forms with practical example. The power of SAS in analysing data patterns and developing such models is also demonstrated appropriate and relevant portions of SAS code are included where ever it is possible.

1. Non Parametric Statistics:

There are numerous ways in which nonparametric statistical methods are defined and the existence of varying descriptions testifies to nonparametric statistical methods flexibility.

A statistical method is nonparametric if it satisfies at least one of the following criteria:

1) The method may be used on data with a nominal scale of measurement.
2) The method may be used on data with an ordinal scale of measurement.
3) The method may be used on data with an interval or ratio scale of measurement, where the distribution function of the random variable producing the date is either unspecified or specified except for an infinite of unknown parameters.

According to the criteria above, nonparametric statistics can be used on any set of data, whether it is nominal, ordinal,

interval, or ratio. In a later section of this paper, it is shown that nonparametric statistics can also be expanded, in some instances, to data including categorical variables.

2. Motivation:

The nonparametric approach for estimating a regression curve has four main purposes.

First, it provides a versatile method of exploring a general relationship between two variables. Second, it gives predictions of observations yet to be made without reference to a fixed parametric model. Third, it provides a tool for finding spurious observations by studying the influence of isolated points. Fourth, it constitutes a flexible method of substituting for missing values or interpolating between adjacent x values.

The flexibility of the method is extremely helpful in a preliminary and exploratory statistical analysis of a dataset. If no a priori model information about the regression curve is available, the nonparametric analysis could help in suggesting simple parametric formulations of the regression relationship.

The parametric model can only produce uni-modal densities; the nonparametric approach makes it possible to estimate functions of greater complexity and suggests instead a bimodal distribution.

Missing data is a problem quite often encountered in practice. Some response variables may not have been recorded since an instrument broke down or a certain entry on an inquiry form was not answered. Nonparametric smoothing bridges the gap of missing data by interpolating between adjacent data points, whereas parametric models would involve all the observations in the interpolation. An approach in spatial statistics is to interpolate points by the kriging method. This method is used by statisticians in hydrology, mining, petroleum engineering and is related to predicting values of noisy data in a nonparametric fashion [10].

3. Introduction to Non Parametric Regression:

Nonparametric regression is a form of regression analysis in which the predictor does not take a predetermined form but is constructed according to information derived from the data. Nonparametric regression requires larger sample sizes than regression based on parametric models because the data must supply the model structure as well as the model estimates.

The traditional nonlinear regression model

$$y = m(x,\theta) + \epsilon$$

Where θ is a vector of parameters to be estimated, and x is a vector of predictors; the errors ϵ

are assumed to be normally and independently distributed with mean 0 and

constant variance σ^2. The function $m(x, \theta)$, relating the average value of the response y to the predictors, is specified in advance, as it is in a linear regression model.

The general nonparametric regression model is written in a similar manner, but the function m is left unspecified.

$$Y = m(x) + \epsilon \quad (1)$$

$$= m(x_1, x_2, ..., x_p) + \epsilon$$

For the p predictors $X = (x_1, x_2, ..., x_p)'$. Moreover, the object of nonparametric regression is to estimate the regression function m(x) directly, rather than to estimate parameters. Most methods of nonparametric regression implicitly assume that m is a smooth, continuous function. As in nonlinear regression, it is standard to assume that $\epsilon_i \sim$ NID $(0, \sigma^2)$.

An important special case of the general model is nonparametric simple regression, where there is only one predictor.

$$Y = m(x) + \epsilon$$

Nonparametric simple regression is often called 'scatter plot smoothing' because an important application is to tracing a smooth curve through a scatter plot of y against x.

Because it is difficult to fit the general nonparametric regression model when there are many predictors, and because it is difficult to display the fitted model when there are more than two or three predictors, more restrictive models have been developed. One such model is the additive regression model,

$$Y = \beta_0 + m_1(x_1) + m_2(x_2) + ... + m_p(x_p) + \epsilon \quad (2)$$

Where the partial regression functions $m_j(x_j)$ are assumed to be smooth, and are to be estimated from the data. This model is much more restrictive than the general nonparametric regression model, but less restrictive than the linear regression model, which assumes that all of the partial regressions function is linear.

Variations on the additive regression model include semi parametric models, in which some of the predictors enter linearly, for example,

$$Y = \beta_0 + \beta_1(x_1) + m_2(x_2) + ... + m_p(x_p) + \epsilon \quad (3)$$

Particularly useful when some of the predictors are factors, and models in which some predictors enter into interactions, which appear as higher dimensional terms in the model, for example,

$$Y = \beta_0 + m_{12}(x_1, x_2) + m_3(x_3) + ... + m_p(x_p) + \epsilon \quad (4)$$

All of these models extend straightforwardly to generalize nonparametric regression, much as linear models extend to generalize linear models. The random and link components are as in generalized linear models, but the linear predictor of the GLM.

$$\eta = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + ... + \beta_p(x_p) \quad (5)$$

is replaced, for example, by an unspecified smooth function of the predictors

$$\eta = m(x_1, x_2, ..., x_p)$$

For the most general case or by a sum of smooth partial regression functions in the generalized additive model.

$$\eta = \beta_0 + m_1(x_1) + m_2(x_2) + ... + m_p(x_p) \quad (6)$$

4. Nonparametric multiplicative regression:

Nonparametric multiplicative regression (NPMR) is a form of nonparametric regression based on multiplicative kernel estimation. Like other regression methods, the goal is to estimate a response (dependent variable) based on one or more predictors (independent variables). NPMR can be a good choice for a regression method if the following are true:

1) The shape of the response surface is unknown.
2) The predictors are likely to interact in producing the response; in other words, the shape of the response to one predictor is likely to depend on other predictors.
3) The response is either a quantitative or binary (0/1) variable.

This is a smoothing technique that can be cross-validated and applied in a predictive way.

NPMR behaves like an organism:

NPMR has been useful for modelling the response of an organism to its environment. Organism response to environment tends to be nonlinear and have complex interactions among predictors. NPMR allows you to model automatically the complex interactions among predictors in much the same way that organisms integrate the numerous factors affecting their performance. [6]

A key biological feature of an NPMR model is that failure of an organism to tolerate any single dimension of the predictor space results in overall failure of the organism. For example, assume that a plant needs a certain range of moisture in a particular temperature range. If either temperature or moisture falls outside the tolerance of the organism, then the organism dies. If it is too hot, then no amount of moisture can compensate to result in survival of the plant. Mathematically this works with NPMR because the product of the weights for the target point is zero or near zero if any of the weights for individual predictors (moisture or temperature) are zero or near zero. Note further that in this simple example, the second condition listed above is probably true: the response of the plant to moisture probably depends on temperature and vice versa.

Optimizing the selection of predictors and their smoothing parameters in a

multiplicative model is computationally intensive. With a large pool of predictors, the computer must search through a huge number of potential models in search for the best model. The best model has the best fit, subject to over fitting constraints or penalties. [7]

5. Estimation Methods:

There are several approaches to estimating nonparametric regression models, but commonly used methods are local polynomial regression and smoothing splines.

5.1. Local Polynomial Regression:

Simple Regression:

Here, we are looking to fit the model

$$Y = m(x) + \epsilon \quad (7)$$

Let us focus on evaluating the regression function at a particular x value, x_0. Ultimately, we will fit the model at a representative range of values of x or simply at the n observations, x_i. We proceed to perform a pth order weighted-least-squares polynomial regression of y on x,

$$Y = b_0 + b_1(x - x_0) + b_2(x - x_0)^2 + \ldots\ldots + b_p(x - x_0)^p + \epsilon \quad (8)$$

Weighting the observations in relation to their proximity to the focal value x_0; a common weight function to use is the tri-cube function,

$$W(z) = \begin{cases} (1 - |z|3)3 & for\ |z| < 1 \\ 0 & for\ |z| \geq 1 \end{cases} \quad (9)$$

In the present context, $z = (x - x_0)/h$, where h is the half width of a window enclosing the observations for the local regression. The fitted value at x_0, that is, the estimated height of the regression curve, is simply $\hat{y}_0 = b_0$, produced conveniently by having cantered the predictor x at the focal value x_0.

It is typical to adjust h so that each local regression includes a fixed proportion s of the data; then, s is called the span of the local-regression smoother. The larger the span, the smoother the result; in contrast, the larger the order of the local regressions p, the more flexible the smooth, so the span and the order of the local regressions can be traded off against one another.

5.2 Multiple Regression:

The nonparametric multiple regression model is, $Y = f(x) + \epsilon$

$$f(x_1, x_2, x_3, \ldots\ldots x_p) + \epsilon \quad (10)$$

Extending the local-polynomial approach to multiple regression is simple conceptually, but can run into practical difficulties.

The first step is to define a multivariate neighbourhood around a focal point $x'_0 = (x_{01}, x_{02} \ldots. x_{0k})$. The default approach in the

loess function is to employ scaled Euclidean distances.

$$D(x_i, x_0) = \sqrt{\sum_{j=1}^{k}(zij - z0j)2} \quad (11)$$

Where the z_j are the standardized predictors,

$$Z_{ij} = \frac{xij - \overline{xj}}{sj}$$

Here, x_i is the predictor vector for the ith case, x_{ij} is the value of the jth predictor for the ith case, x_j is the mean of the jth predictor, and s_j is its standard deviation.

Weights are defined using the scaled distances,

$$w_i = W\left[\frac{D(xi,x0)}{h}\right]$$

Where W (.) is a suitable weight function, such as the tricube, in which case h is the half width (i.e. radius) of the neighbourhood. As in local simple regression, h may be adjusted to define a neighbourhood including the [ns] nearest neighbours of x_0 (where the square brackets denote rounding to the nearest integer).

Perform a weighted polynomial regression of y on the x's, for example, a local linear fit takes the following form.

$$Y = b_0 + b_1(x_1 - x_{01}) + b_2(x_2 - x_{02}) + ...+ b_k(x_k - x_{0k}) + e \quad (12)$$

The fitted value at X_0 is then simply $\hat{y}_0 = b_0$.

The procedure is repeated for representative combinations of predictor values to build up a picture of the regression surface.

5.3 Idea of Smoothing:

A reasonable approximation to the regression curve m(x) will be the mean of response variables near a point x. This local averaging procedure can be defined as,

$$\hat{m}(x) = n^{-1}\sum_{i=1}^{n} W_{ni}(x)Y_i \quad (13)$$

Every smoothing method to be described is of the above form.

The amount of averaging is controlled by a smoothing parameter. The choice of smoothing parameter is related to the balances between bias and variance.

5.3 Smoothing Splines:

Smoothing splines arise as the solution to the following simple-regression problem. Find the function $\hat{m}(x)$ with two continuous derivatives that minimizes the penalized sum of squares,

$$SS^*(h) = \sum_{i=1}^{n}[yi - m(xi)]2 + h\int_{xmin}^{xmax}[m''(x)]2dx \quad (14)$$

Where h is a smoothing parameter, analogous to the neighbourhood width of the local polynomial estimator.

- The first term in the above equation is the residual sum of squares.
- The second term is a roughness penalty, which is large when the integrated second derivative of the regression function m''(x) is large, that is, when m(x) is rough (with rapidly changing slope). The endpoints of the integral enclose the data.
- At one extreme, when the smoothing constant is set to h = 0 (and if all the x-values are distinct), $\hat{m}(x)$ simply interpolates the data; this is similar to a local regression estimate with span = 1/n.
- At the other extreme, if h is very large, then $\hat{m}$ will be selected so that $\hat{m}''(x)$ is everywhere 0, which implies a globally linear least squares fit to the data (equivalent to local regression with infinitely wide neighbourhoods).

The function $\hat{m}(x)$ that minimizes the above equation is a natural cubic spline with knots at the distinct observed values of x. Although this result seems to imply that n parameters are required (when all x-values are distinct), the roughness penalty imposes additional constraints on the solution, typically reducing the equivalent number of parameters for the smoothing spline substantially, and preventing $\hat{m}(x)$ from interpolating the data. Indeed, it is common to select the smoothing parameter h indirectly by setting the equivalent number of parameters for the smoother.

Because there is an explicit objective-function to optimize, smoothing splines are more elegant mathematically than local regression. It is more difficult, however, to generalize smoothing splines to multiple regression and smoothing spline and local regression fits with the same equivalent number of parameters are usually very similar.

5.4 Selecting Smoothing parameter:

Both local polynomial regression and smoothing splines have an adjustable smoothing parameter. This parameter may be selected by visual trial and error, picking a value that balances smoothness against fidelity to the data. More formal methods of selecting smoothing parameters typically try to minimize the mean squared error of the fit, either by employing a formula approximating the mean-square error (e.g., so called plug in estimates), or by some form of cross validation.

In cross validation, the data are divided into subsets (possibly comprising the individual observations), the model is successively fit omitting each subset in turn; and then the fitted model is used to `predict' the response for the left-out subset. Trying this procedure for different values of the smoothing parameter will suggest a value that minimizes the cross validation estimate of the mean squared error. Because cross validation is very computationally intensive,

approximations and generalizations are often employed [14].

5.5 Bandwidth Selection: For Smooth densities and a normal kernel, use the bandwidth,

$$h_n = \frac{1.06\hat{\sigma}}{n1/5} \quad (15)$$

where, $\hat{\sigma} = \min\{s, Q/1.34\}$

The cross validation score is,

$$\hat{J}(h) = \int \hat{f}^2(x)\,dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-1}(X_i) \quad (16)$$

$$\hat{J}(h) = \frac{1}{hn2}\sum_i \sum_j K^*\left(\frac{Xi-Xj}{h}\right) + \frac{2}{nh}K(0) \quad (17)$$

5.6 Kernel Smoothing:

- Kernel smoothing describes the shape of the weight function $W_{ni}(x)$ by a density function K with a scale parameter that adjusts the size and the form of the weights near x. The kernel K is a continuous, bounded and symmetric real function which integrates to 1. The weight is defined by,

$$W_{hi}(x) = K_h(x-X_i) / \hat{f}_h(x) \quad (18)$$

Where $\hat{f}_h(x) = n^{-1}\sum_{i=1}^{n} K_h(x-X_i)$, and $k_h(u) = h^{-1}K(u/h)$

- The Nadaraya-Watson estimator is defined by,

$$\hat{m}_h(x) = \frac{n-1\sum_{i=1}^{n} Kh(x-Xi)Yi}{n-1\sum_{i=1}^{n} Kh(x-Xi)} \quad (19)$$

The mean squared error is $d_M(x,h) = E[\hat{m}_h(x) - m(x)]^2$. As $n\to\infty$, $h\to 0$, $nh\to\infty$. We have, under certain conditions,

$$d_M(x,h) = (nh)^{-1}\sigma^2 c_k + h^4 d^2 k[m''(x)]^2/4 \quad (20)$$

where $\sigma^2 = var(\epsilon_i)$, $ck = \int K^2(u)du$, $dk = \int u^2 K(u)du$

The bias is increasing whereas the variance is decreasing in h.

5.7 Smoothing Issues:

- Since averaging is done over neighbouring observations, an estimate of m(.) at peaks or bottoms will flatten them. This finite sample bias depends on the local curvature of m(.)
- At the boundary points, half the weights are not defined. This also creates a bias.
- When there are regions of sparse data, weights can be undefined no observations to average.

5 .8 K-Nearest Neighbour Estimates:

In k-NN, the neighbourhood is defined through those X variables which are among the k-nearest neighbours of x in Euclidean distance. The k-NN smoother is defined as,

$$\hat{m}_k(x) = n^{-1}\sum_{i=1}^{n} W_{ki}(x)Y_i \quad (21)$$

Where $\{W_{ki}(x)\}$ i=1... n is defined through the set of indexes $J_x = \{i: X_i$ is one of the k nearest observations to x$\}$, and

$$W_{ki}(x) = \begin{cases} \frac{n}{k}, if\ i \epsilon Jx \\ 0\ Otherwise \end{cases} \quad (22)$$

- The smoothing parameter k regulates the degree of smoothness of the estimated curve. It plays a role similar to the bandwidth for kernel smoothers.
- The influence of varying *k* on qualitative features of the estimated curve is similar to that observed for kernel estimation with a uniform kernel.
- When $k > n$, the *k - NN* smoother then is equal to the average of the response variables. When $k = 1$, the observations are reproduced at X_i, and for an *x* between two adjacent predictor variables a step function is obtained with a jump in the middle between the two observations.
- Let $k \to \infty$, $k/n \to 0$, $n \to \infty$. Bias and variance of the K-NM estimate with weights as in equation(22) are given by,

$$E\hat{m}_k(x) - m(x) = \frac{1}{24f(x)^3}[m``f + 2m'f')(x)](k/n)^2 \quad (23)$$

$$Var\{\hat{m}_k(x)\} = \sigma^2(x)/k$$

In addition to the uniform weights, the k-NN weights can be generally thought of as being generated by a kernel function K,

$$W_{Ri}(x) = K_R(x - Xi) / \hat{f}_R(x) \quad (24)$$

Where, $\hat{f}_R(x) = n^{-1}\sum_{i=1}^{n} K_R(x - Xi)$

And, $K_R(u) = R^{-1}K(u/R)$

R is the distance between x and its k-th nearest neighbour

5 .9 Additive Non Parametric Regression:

The additive non parametric regression model is,

$$Y = \beta_0 + m_1(x_1) + m_2(x_2) + m_3(x_3) + ... + m_k(x_k) + \epsilon \quad (25)$$

where the partial regression functions m_j are fit using a simple regression smoother, such as local polynomial regression or smoothing splines.

Introduction to Data and Variables:

In this section we present the data. This data is collected from Canadian cross section wage data consisting of a random sample, which represents Canadian high school graduate earnings. We have two variables

Different SAS Procedures for Non-Parametric Regression:

Nonparametric regression relaxes the usual assumption of linearity and enables you to uncover relationships between the independent variables and the dependent variable that might otherwise be missed.

The SAS/STAT nonparametric regression procedures include the following:

- ADAPTIVEREG Procedure - Multivariate adaptive regression splines
- GAM Procedure (Generalized Additive Models) - Fits generalized additive models
- GAMPL Procedure (Generalized Additive Model using Penalized Likelihood Estimation) - Fits generalized additive models that are based on low rank regression splines
- LOESS Procedure or LOWESS (Logically Weighted Scatter plot Smoothing) - Nonparametric method for estimating regression surfaces
- TPSPLINE Procedure (Thin Plan Spline) - Provides penalized least squares estimates

age and log wage and there are 205 observations in total.

Logically Weighted Scatter plot Smoothing(LOESS) Procedure:

The LOESS procedure implements a nonparametric method for estimating regression surfaces. PROC LOESS allows great flexibility because no assumptions about the parametric form of the regression surface are needed.

The following are highlights of the LOESS procedure's features:

- Supports the use of multidimensional data
- Supports multiple dependent variables
- Supports both direct and interpolated fitting that uses k trees
- Performs statistical inference
- Performs automatic smoothing parameter selection
- Performs iterative reweighting to provide robust fitting when there are outliers in the data
- Scores external data sets
- Performs BY group processing, which enables you to obtain separate analyses on grouped observations
- Performs weighted estimation

- Creates a SAS data set that contains the predicted values and other requested statistics
- Creates a SAS data set that corresponds to any output table
- Automatically creates graphs by using ODS Graphics

- **Below are the results for non parametric regression model from the above code.**

The SAS System

The LOESS Procedure

Independent Variable Scaling	
Scaling applied: None	
Statistic	age
Minimum Value	21.00000
Maximum Value	65.00000

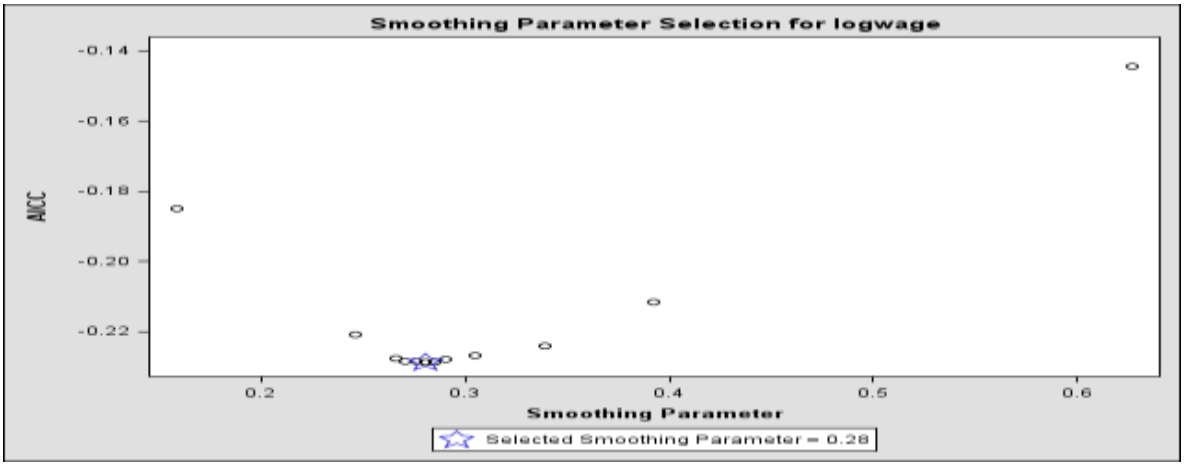
The SAS System

The LOESS Procedure

Dependent Variable: log wage

Model Summary				
Smoothing Parameter	Local Points	Residual SS	GCV	AICC
0.39268	80	57.29854	0.00144	-0.21157
0.62683	128	62.35093	0.00154	-0.14443
0.24634	50	55.00880	0.00142	-0.22098

Model Summary				
Smoothing Parameter	Local Points	Residual SS	GCV	AICC
0.15854	32	54.58793	0.00147	-0.18508
0.30488	62	55.59025	0.00141	-0.22705
0.33902	69	56.15054	0.00142	-0.22408
0.28049	57	55.12079	0.00141	-0.22888
0.26585	54	54.98814	0.00141	-0.22762
0.29024	59	55.36521	0.00141	-0.22795
0.27073	55	55.07917	0.00141	-0.22857
0.28537	58	55.20755	0.00141	-0.22873
0.27561	56	55.07917	0.00141	-0.22857
0.28049	57	55.12079	0.00141	-0.22888



Optimal Smoothing Criterion	
AICC	Smoothing Parameter
-0.22888	0.28049

The SAS System

The LOESS Procedure

Selected Smoothing Parameter: 0.28

Dependent Variable: log wage

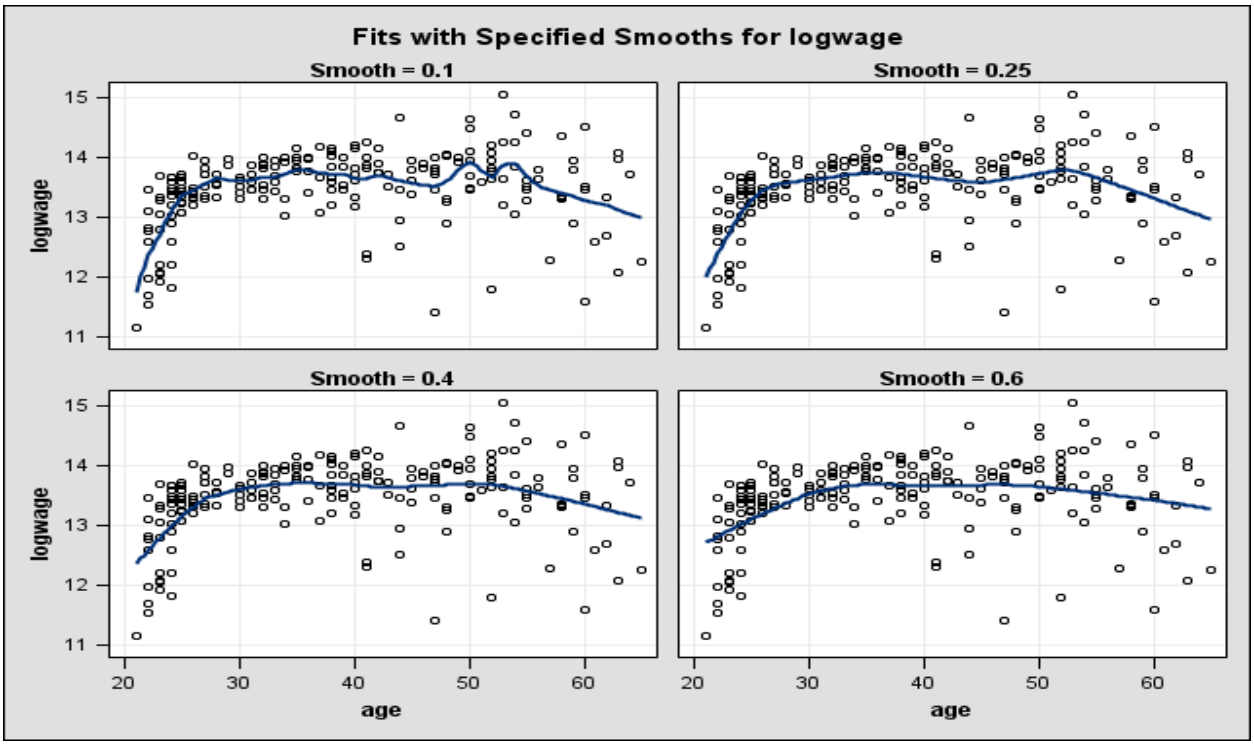
Output Statistics				
Obs	age	Log wage	Predicted log wage	Residual
1	21.00000	11.15600	12.06531	-0.90931
2	22.00000	12.81300	12.39935	0.41365
3	22.00000	13.09600	12.39935	0.69665
4	22.00000	11.69500	12.39935	-0.70435
5	22.00000	11.53300	12.39935	-0.86635
6	22.00000	12.76600	12.39935	0.36665
7	22.00000	12.58800	12.39935	0.18865
8	22.00000	11.98300	12.39935	-0.41635
9	22.00000	13.45900	12.39935	1.05965
10	23.00000	12.20600	12.72431	-0.51831

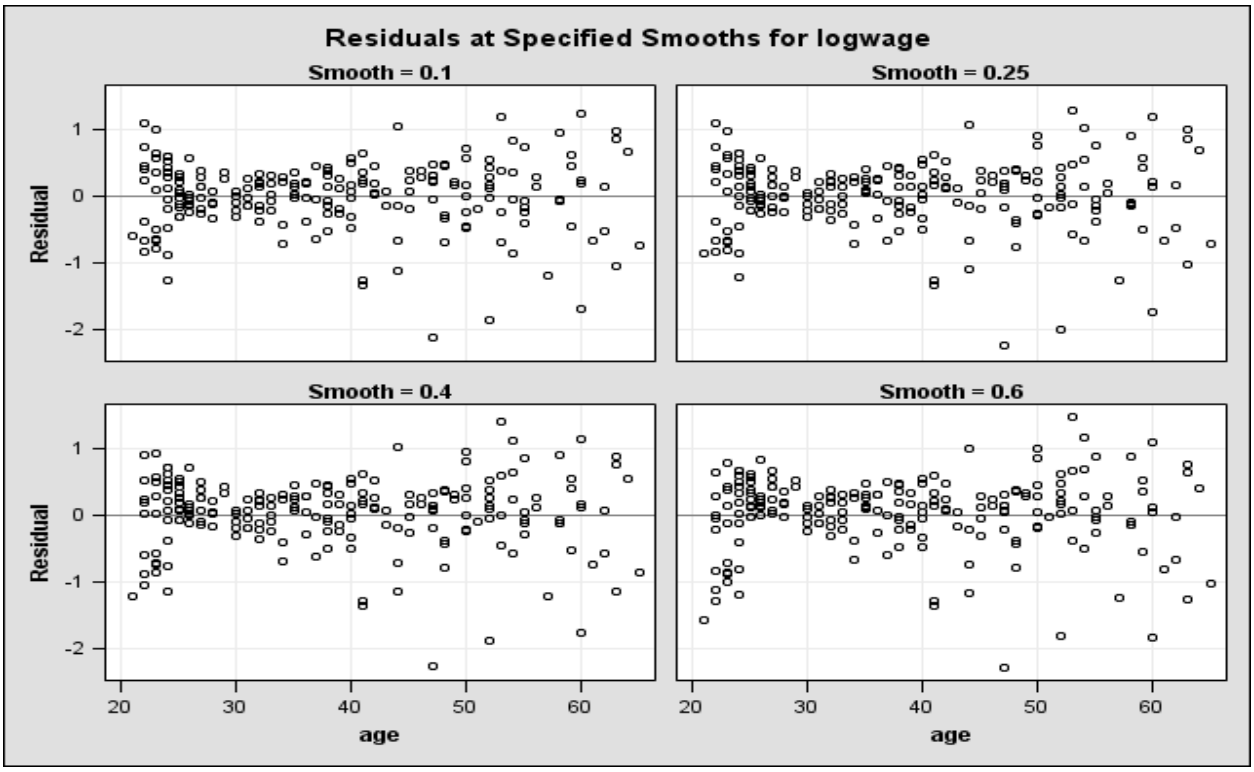
200	62.00000	12.69500	13.18314	-0.48814
201	63.00000	13.96800	13.11261	0.85539
202	63.00000	12.07800	13.11261	-1.03461
203	63.00000	14.09200	13.11261	0.97939
204	64.00000	13.71000	13.04207	0.66793
205	65.00000	12.25500	12.97153	-0.71653

The SAS System

The LOESS Procedure

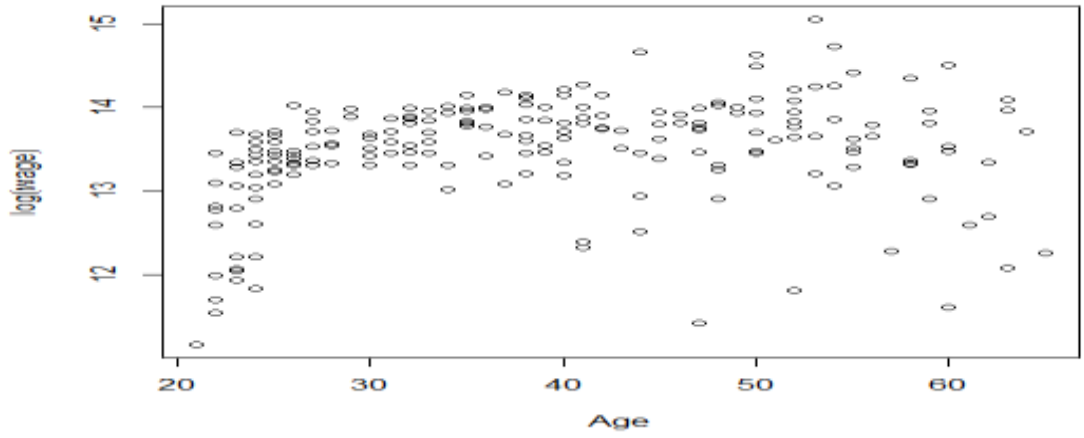
Dependent Variable: log wage



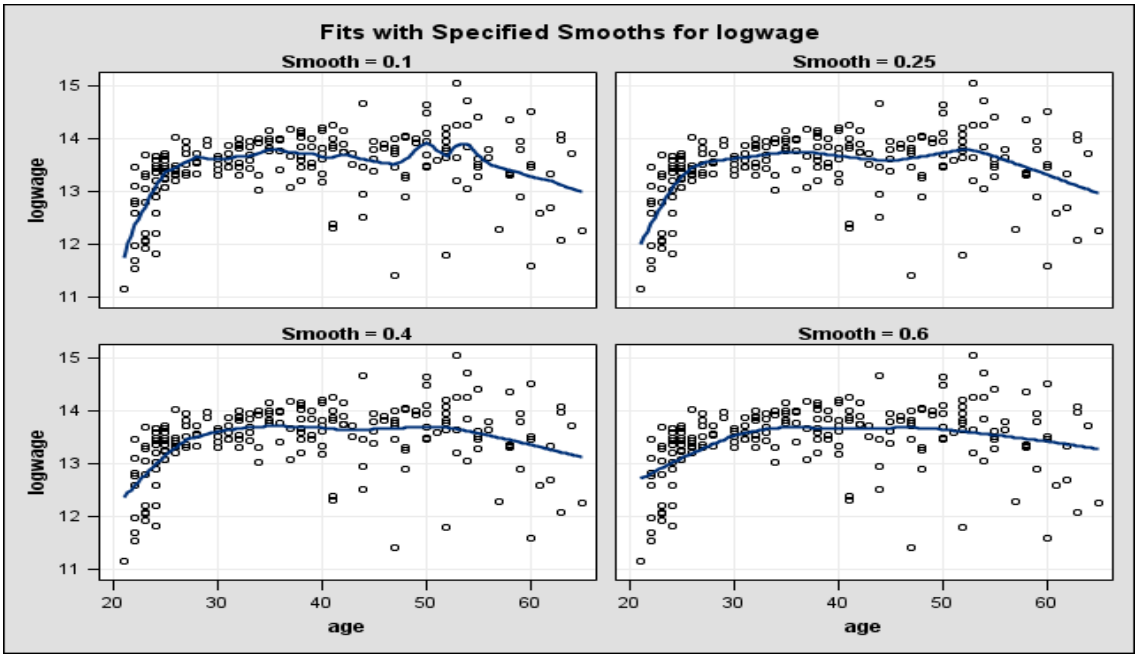


6. Conclusions:

Below is the scatter plot between the original variables, Age and Log Wage. From the below plot we can observe that our data is not satisfying linearity condition and most of the observations were scattered randomly. Hence we cannot directly go ahead with linear regression to fit an equation with minimum error in order to study the relationship between those two variables Age and Log Wage. For this reason, there is a need for non parametric regression analysis to study the relationship between Age and Log Wage.



From the results we can observe that, non parametric regression model with smoothing parameters 0.1 and 0.25 are resulting better fit with minimum residuals sum of squares, in order to study the relationship between Age and log wage.



7. References:

[1] "A Distribution-Free Theory of Nonparametric Regression" – Gyorfi Laszlo, Michael Kohler, Adam Krzyzak and Harro Walk (2002).

[2] "Understanding Regression Analysis" – Allen, Mikael Patrick (1997).

[3] "Nonparametric Regression and Generalized Linear Models" - Green P J and B W Silverman (1994)

[4] "Nonparametric Statistical Methods" - Hollander, Myles, and Douglas A. Wolfe (1973).

[5] "SAS Online Documentation", SAS/STAT.

[6] "Nonparametric habitat models with automatic interactions" – McCune B, (2006).

[7] "Spatial and temporal dynamics of potato tuber worm in the Columbia Basin of the Pacific Northwest" – DeBano SJ, Hamm PB, Jensen A, Rondon SI, Landolt PJ, (2010).

[8] "The Elements of Statistical Learning" - Hastie T, Tibsharani R, Friedman J, (2001).

[9] "Applied Nonparametric Regression" – Wolfgang Hardle (1994).

[10] "An Introduction to Numerical Computations" – Yakowitz S and Szidarovsky F (1986).

[11] "Non Parametric Regression in R" – John Fox and Sanford Weisberg (2010).

[12] "An R Companion to Applied Regression" – Fox J and Weisberg (2011).

[13] "Multidimensional Smoothing with smoothing splines" – Gu C (2000).

[14] "Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties" – Wood SN (2000).

[15] "Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models" – Wood SN (2004).

[16] "Generalized Additive Models" – Wood SN (2006).

[17] "Introduction to Non Parametric Estimation" – Tsybakov A (2009).

[18] "Non Parametric Statistics" – Wasserman L (2006).

[19] "Applied Regression Analysis and Generalized models" – Fox J (2008).

[20] "An Introduction to PROC LOESS for Non Parametric Regression" – Robert A Cohen.

[21] "Wavelet Methods in Statistics with R" – Nason G (2008).

[22] "Non Parametric Regression" – Ryan Tibshirani (2015).

[23] "A Distribution Free Theory of Non Parametric Regression" – Gyorfi L, Kohler M, Walk H (2002).

[24] "Component Selection and smoothing in multivariate non parametric regression" – Lin Y, Zhang HH (2006).

[25] "The Elements of Statistical Learning" – Hastie T, Tibshirani R and Friedman J (2009).

[26] "Non Parametric Regression Techniques" – Geoff Bohling (2005).

[27] "Non Parametric Inference with R" – Chad Schafer and Larry Wasserman.

[28] "A Distribution Free Theory of Nonparametric Regression" – Adam k and Michael Kohler.

[29] "An Introduction to Nonparametric Regression" – Ning Li (2004).

[30] "Semi parametric Regression for the Applied Econometrician" – Yatchew A (2003).

[31] "Kernel Regression in Empirical Microeconomics" – Blundell and Duncan (1998).

[32] "Implementing Nonparametric and Semi Parametric Estimators" – Ichimura and Todd (2007)

[33] "Nonparametric Econometrics" – Pagan A and Ullah A (1999).

[34] "Smoothing and Nonparametric Regression" – German Rodriguez (2001).

[35] "A Nonparametric Framework" – Truong YK (1993).

[36] "Nonparametric Probability Density Estimation" – Tapia D and Thompson J (1978).

[37] "Consistent Nonparametric Regression" – Stone CJ (1977).

[38] "Smoothing Methods in Statistics" – Simonoff J (1996).

[39] "Adaptive Piecewise Polynomial Estimation via Trend Filtering" – Tibshirani RJ (2009).

[40] "Nonparametric Regression and Generalized Linear Models" - Green P and Silverman B.