

Parametric Versus Semi and Nonparametric Regression Models

Hamdy F. F. Mahmoud^{1,2}

¹ Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

² Department of Statistics, Mathematics, and Insurance, Faculty of Commerce, Assiut University, Assiut 71515, Egypt

Correspondence: Hamdy F. F. Mahmoud, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA. E-mail: ehamdy@vt.edu

Received: December 20, 2020 Accepted: February 20, 2021 Online Published: February 23, 2020

doi:10.5539/ijsp.v10n2p90 URL: <https://doi.org/10.5539/ijsp.v10n2p90>

Abstract

There are three common types of regression models: parametric, semiparametric and nonparametric regression. The model should be used to fit the real data depends on how much information is available about the form of the relationship between the response variable and explanatory variables, and the random error distribution that is assumed. Researchers need to be familiar with each modeling approach requirements. In this paper, differences between these models, common estimation methods, robust estimation, and applications are introduced. For parametric models, there are many known methods of estimation, such as least squares and maximum likelihood methods which are extensively studied but they require strong assumptions. On the other hand, nonparametric regression models are free of assumptions regarding the form of the response-explanatory variables relationships but estimation methods, such as kernel and spline smoothing are computationally expensive and smoothing parameters need to be obtained. For kernel smoothing there two common estimators: local constant and local linear smoothing methods. In terms of bias, especially at the boundaries of the data range, local linear is better than local constant estimator. Robust estimation methods for linear models are well studied, however the robust estimation methods in nonparametric regression methods are limited. A robust estimation method for the semiparametric and nonparametric regression models is introduced.

Keywords: kernel regression, nonparametric, parametric, robust estimation, semiparametric, single index model, spline smoothing

1. Introduction

The aim of this paper is to answer many questions that researchers may have when they fit a real data, such as what are the differences between parametric, semi, and nonparametric models? which one should be used to model a real data set? what estimation method should be used? and which modeling approach is better? These questions and others are addressed by examples in this paper and the R code for plots and analyses presented are available in the Appendix.

Assume that a researcher collected data about a response variable, y , and k explanatory variables, (x_1, x_2, \dots, x_k) . The model that describes the relationship between y and \mathbf{X} can be written as

$$y = f(\mathbf{X}, \boldsymbol{\beta}) + \epsilon, \quad (1)$$

where $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters, $\mathbf{X} = (x_1, x_2, \dots, x_k)$ is a $n \times k$ matrix of regressors value, ϵ is an error term, $f(\cdot, \cdot)$ is the function that describes the relationship between y and \mathbf{X} .

The model choice from parametric, semiparametric or nonparametric regression model depends on the prior knowledge about the functional form of the relationship, $f(\cdot, \cdot)$, and the random error distribution. If the form is known and it is correct, the parametric choice can model the data set well. However, if a wrong functional form is chosen, this will result in larger bias as compared to the other competitive regression models (Fan and Yao, 2003). The most common functional form is the linear model, as a type of parametric regression, which is frequently used to describe the relationship between a dependent variable and explanatory variables. Parametric linear models require the estimation of a finite number of parameters, $\boldsymbol{\beta}$.

Parametric models are easy to work with, estimate, and interpret. So, why are semiparametric and nonparametric regression important? Over the last decade, increasing attention has been devoted to these regression models as new techniques for estimation and prediction in different areas, such as epidemiology, agriculture, economics, ..., etc. Nonparametric regression analysis relaxes the assumption of linearity or even knowledge about the functional form priori in regression analysis and enables one to explore the data more flexibly. However, in case of high dimensions, the variance of the

estimates rapidly increases, this is called curse of dimensionality, due to the sparseness of data. To solve this problem, semiparametric models have been proposed, such as single index model (SIM). The three modeling approaches are compared in terms of fitting and prediction (Rajaratnan and Parmar 2011; Dhekale et al. 2017). Mahmoud et al. (2016) showed that if the link function in generalized linear models is misspecified, semiparametric models are better than parametric models. In Section 2, the three types of models are introduced. In Section 3, which model should be used is addressed. In Section 4, common estimation methods for semiparametric and nonparametric models are displayed. In Section 5, the multiple case is presented. Robust estimation method for the semi and nonparametric regression is introduced in Section 6. Section 7 includes discussion and conclusion.

2. Parametric, Semi and Nonparametric Regression Models

To differentiate between the three types of regression models, without loss of generality, assume that we have a response variable, y , and two explanatory variables, x_1 and x_2 . The regression model that describes the relationship between the response variable and two the explanatory variables can be written as:

$$y = f_1(x_1, \beta_1) + f_2(x_2, \beta_2) + \epsilon, \quad (2)$$

where β_1 and β_2 are the model parameters that need to be estimated, and $f_1(\cdot, \cdot)$ and $f_2(\cdot, \cdot)$ are the functions describe the relationships between y and the two explanatory variables, x_1 and x_2 .

2.1 Parametric Models

Parametric models are models in which the vector of parameters, β in model (1), is a vector in finite p -dimensional space (the number of parameters, p , may be less or greater than the number of explanatory variables, k). Our interest in this case is estimating the vector of parameters. In the parametric models case, the researcher assumes the form of the model and its assumptions. Applying on model (2), $f_1(\cdot, \cdot)$ and $f_2(\cdot, \cdot)$ need to be known functions and ϵ distribution is known priori for inference purposes. For example, the researcher may assume they are linear and the error term follows the normal distribution. In general, $f_1(\cdot, \cdot)$ and $f_2(\cdot, \cdot)$ can be assumed linear or nonlinear in terms of the β 's. For model validity, after fitting the data using the assumed model, the researcher needs to check whether the assumptions are met by using residuals.

2.1.1 Linear Models

Linear models are linear in parameters (i.e., linear in β 's). For example, the polynomial regression that is used to model curvature in a data set, by using higher-ordered values of the predictors, is a linear model in β 's. Hence, the final regression model is a linear combination of higher-ordered predictors. There are many examples of linear models, such as

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ (Multiple linear regression model),
- $y = \beta_0 + \beta_{10} x_1 + \beta_{11} x_1^2 + \beta_{20} x_2 + \beta_{21} x_2^2 + \epsilon$ (Polynomial regression model of the second order),
- and $\log(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ (Poisson regression in case the response variable, y , is count data).

These models are linear in parameters and the number of parameters is greater than the number of the explanatory variables with one ($p = k + 1$). For this setting, there are many methods for parameters estimation, such as the ordinary least square method and maximum likelihood method. The main interest of the researcher, in this case, is estimating the vector of parameters. Once he estimated the parameters, everything is straightforward afterwards, such as inference and prediction.

2.1.2 Nonlinear Models

The nonlinear regression models are parametric models and $f(\cdot, \cdot)$ is known but nonlinear in parameters. Below are some examples of nonlinear models

- $y = \beta_0 + \beta_1 x_1 + \beta_2 e^{\beta_3 x_2} + \epsilon$
- $y = \frac{e^{\beta_0 + \beta_1 x_1}}{1 + e^{\beta_0 + \beta_1 x_1}} + \epsilon$
- $y = \beta_0 + (0.4 - \beta_0) e^{-\beta_1 (x_1 - 5)} + \beta_2 x_2 + \epsilon$

For nonlinear regression models case, the nonlinear least square method can be used to estimate the model parameters. There are several algorithms for the nonlinear least squares estimation, such as Newton's method and Gauss-Newton algorithm.

1. Newton's method is based on a gradient approach. It can be computationally challenging and depends on initial values (Atkinson, 1989).
2. Gauss-Newton algorithm is a modification of Newton's method that gives an approximation of the solution that Newton's method but it is not guaranteed to converge (Nocedal and Wright, 1999).

There are some nonlinear models that can be linearized in parameters using a transformation. For example,

$$y = \frac{\beta_0 x}{\beta_1 + x}$$

can be written as

$$\frac{1}{y} = \frac{1}{\beta_0} + \frac{\beta_1}{\beta_0} \frac{1}{x} \quad (3)$$

$$y^* = \theta_0 + \theta_1 x^* \quad (4)$$

The last equation is linear in parameters, so it is a linear model. If a researcher found that the relationship between the response and explanatory variable is not linear, a simple solution may be used to fix the problem before considering the nonlinear, nonparametric, or semiparametric modeling, such as

- applying a transformation to the dependent and/or independent variables, such as a log transformation, square root, or power transformation.
- adding another regressor as a function of one of the explanatory variables. For example, if a researcher consider regressing y on x , it may make sense to regress y on both x and x^2 (i.e., x -squared).

Practically, the parametric models do not work well in modeling data in many cases of applications, they have very strong assumptions (Rajaratnam and Parmar (2011)).

2.2 Nonparametric Models

The parametric and nonparametric regression models differ in that the nonparametric model form is not specified a priori but is instead determined from the data set. The term nonparametric does not mean that such models are completely lacking parameters, but that the number of parameters is flexible and is not fixed a priori. In parametric models, the vector of the parameters, β , is a vector in a finite p -dimensional space and our main interest is estimating that vector of parameters. In contrast, in nonparametric models, the set of parameters is a subset of infinite-dimensional vector spaces. The primary interest is in estimating the infinite-dimensional vector of parameters. In nonparametric regression models, the relationship between the explanatory variables and response is unknown. Applying that on model (2), $f_1(\cdot, \cdot)$ and $f_2(\cdot, \cdot)$ both are unknown functions. These functions can take any form but they are unknown to the researcher, they may be linear or nonlinear. In parametric modeling, the researcher knows exactly the model of the data. In nonparametric modeling, the data shows what the regression model should be used; the data decide the form of the functions, $f_1(\cdot, \cdot)$ and $f_2(\cdot, \cdot)$.

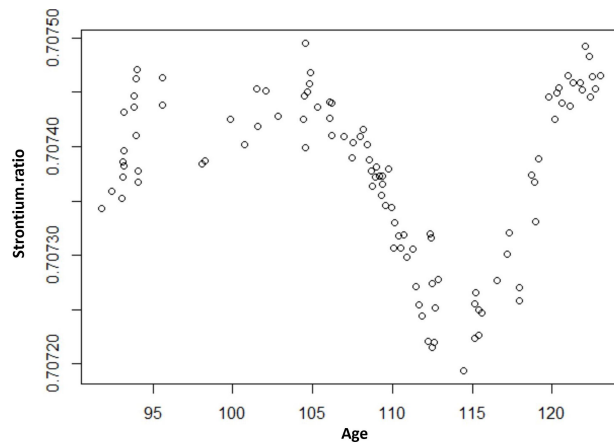


Figure 1. A scatter plot of age and strontium ratio

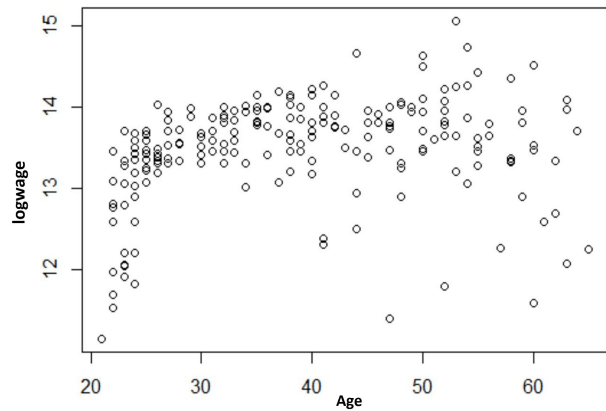


Figure 2. A scatter plot of age versus log of wage.

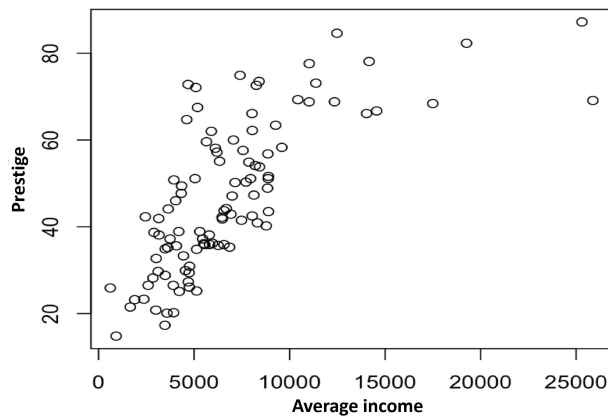


Figure 3. A scatter plot of average income and prestige

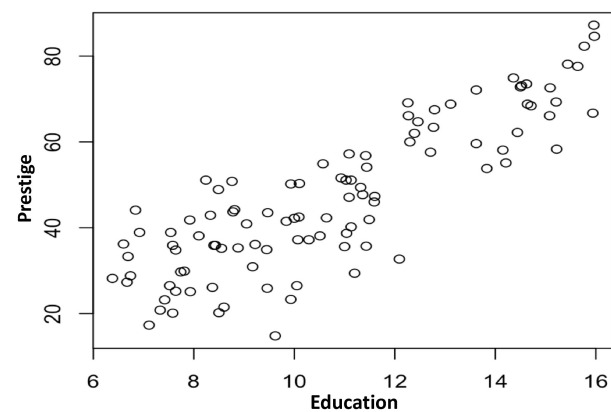


Figure 4. A scatter plot of education and prestige

2.3 Semiparametric Models

Semiparametric modeling is a hybrid of the parametric and nonparametric approaches of statistical models. It may appear at first that the semiparametric model contains the nonparametric model case, however, the semiparametric model is considered to be “smaller” than a completely nonparametric model because we are often interested only in the finite-dimensional component of β . By contrast, in nonparametric models, the primary interest is in estimating the infinite-dimensional parameter. As a result, the estimation is statistically harder in nonparametric models compared to semiparametric models.

While parametric models are being easy to understand and easy to work with, they fail to give a fair representation of what is happening in the real world. Semiparametric models allow us to have the best of both worlds: a model that is understandable and offering a fair representation of the messiness that is involved in real life. The semiparametric regression take several structures. One structure is a form of regression analysis in which a part of the predictors do not take predetermined forms and the other part takes known forms with the response. For example, in model (2), $f_1(\cdot, \cdot)$ may be known and $f_2(\cdot, \cdot)$ is unknown. In case the known function is linear, the model can be written as

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + f(\mathbf{x}_2, \beta_2) + \epsilon.$$

Another structure of the semiparametric regression modeling, which is a well-known example, is the single index model that is extensively studied and has many applications. In general, it takes the form

$$\mathbf{y} = f(\mathbf{X}\beta) + \epsilon, \quad (5)$$

where $f(\cdot)$ is an unknown function, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ is a $n \times k$ matrix of regressors values, β is a $k \times 1$ vector of parameters, and $E(\epsilon|X) = \mathbf{0}$. The term $\mathbf{X}\beta$ is called a “single index” because it reduces the p -dimension into one dimension. In this

case, the functional form of $f(\cdot)$ is unknown to the researcher. This model is semiparametric since the functional form of the linear index is specified, while $f(\cdot)$ is unspecified. In the single index model, the explanatory variables affect the response variable through the index $(\mathbf{X}\beta)$. The SIM is extensively studied and applied in many different fields including biostatistics, medicine, economics, financial econometrics and epidemiology (Ruppert et al. 2003; Mahmoud et. al 2016, 2019; Toma and Fulga 2018; Li, et al. 2017; Qin et al. 2018).

SIM is more flexible compared to parametric models and does not lack from the curse of dimensionality problem compared to nonparametric models. It assumes that the link between the response and the explanatory variables is unknown and should be estimated nonparametrically. This gives the single index model two main advantages over parametric and nonparametric models: (1) It avoids misspecifying the link function and its misleading results (Horowitz and Hardle, 1996) and (2) the reduction of dimension which is achieved by assuming the link function to be a univariate function applied to the projection of explanatory covariate vector onto some direction. To fix the identifiability problem of the SIM, the coefficient of one of the continuous explanatory variables is set to be equal to 1 (Ichimura 1993, Sherman 1994) or $\|\beta\|=1$ is used (Lin and Kulasekera 2007, Xia et al. 2004).

Model (2) can be written in the form of single index model as

$$\mathbf{y} = f(\beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2) + \epsilon \quad (6)$$

Figures 1-4 show different relationships between a response variable and an explanatory variable. For Figure 1-3, the relationship between the response and explanatory variable is not linear, so a researcher may use the nonparametric model to fit this data or try to find a polynomial regression model that can fit the data that is because there is no known model for these data sets. Figure 4 shows a linear relationship between the two variables so the linear parametric model can be used to fit the data.

3. Which Type of Modeling You Should Use to Fit Your Data?

The first step in statistical analysis is summarizing the data numerically and graphically of the response (dependent) variable and the explanatory (independent) variables. The researcher should look at scatter plots, boxplots, histograms, and numerical summaries. That is to get an initial information about the data and see whether the model assumptions are met and whether there are outliers.

Figure 5 shows that the relationship between age and the log of wage is nonlinear. It displays the fitted lines of linear, quadratic, cubic and the 4th degree of polynomial regression along with the scatterplot.

Assume that a researcher used the linear model to fit the data. So to check the linearity assumption, the researcher needs to look at the residual plots. Figure 6 shows that the linearity assumption is not satisfied because fitted values versus residuals plot reveals a nonlinearity form and also normality is not satisfied that is from the the Normal Q-Q plot. Comparing the linear model to 4th degree polynomial model, Figure 7 shows a better fitted versus residuals graph, and normality of residual is much improved. In addition, we need to look at the p-value associated with the fitted model. For the linear model, p-value = 0.0008407, which is significant, however, it is clear that the linear model is not suitable to fit this data and R-squared = 0.05357 is very small. For the 4th degree polynomial regression, p-value = 1.202e-15, R-squared = 0.315.

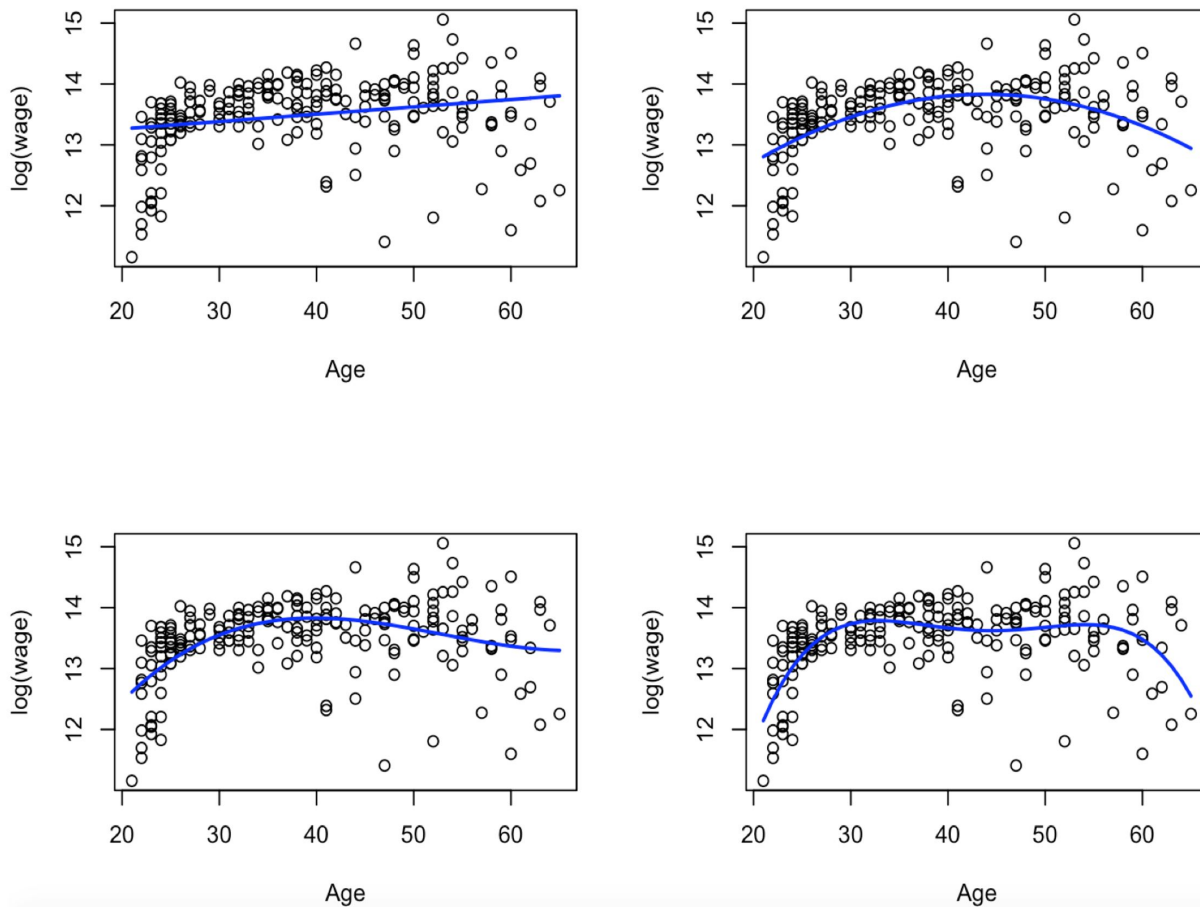


Figure 5. A scatterplot of age and log(wage) along with different polynomial regression lines: the linear model $y = \beta_0 + \beta_1 x$ (top left), quadratic $y = \beta_0 + \beta_1 x + \beta_2 x^2$ (top right), cubic $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ (bottom left), and quartic $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$ (bottom right)

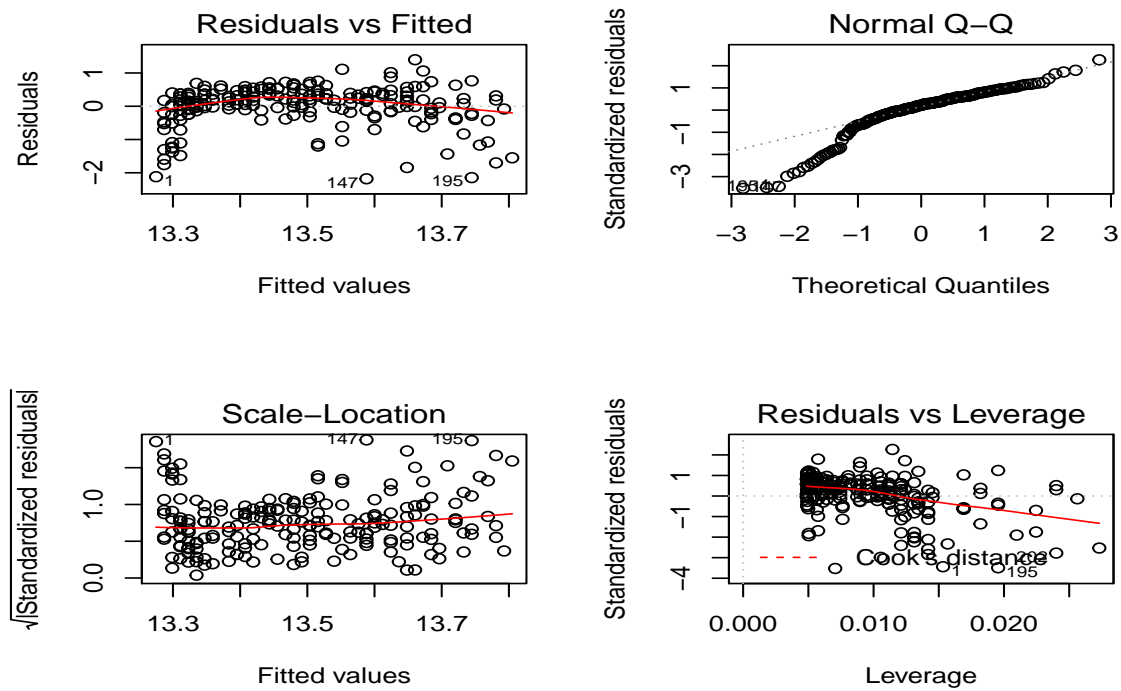
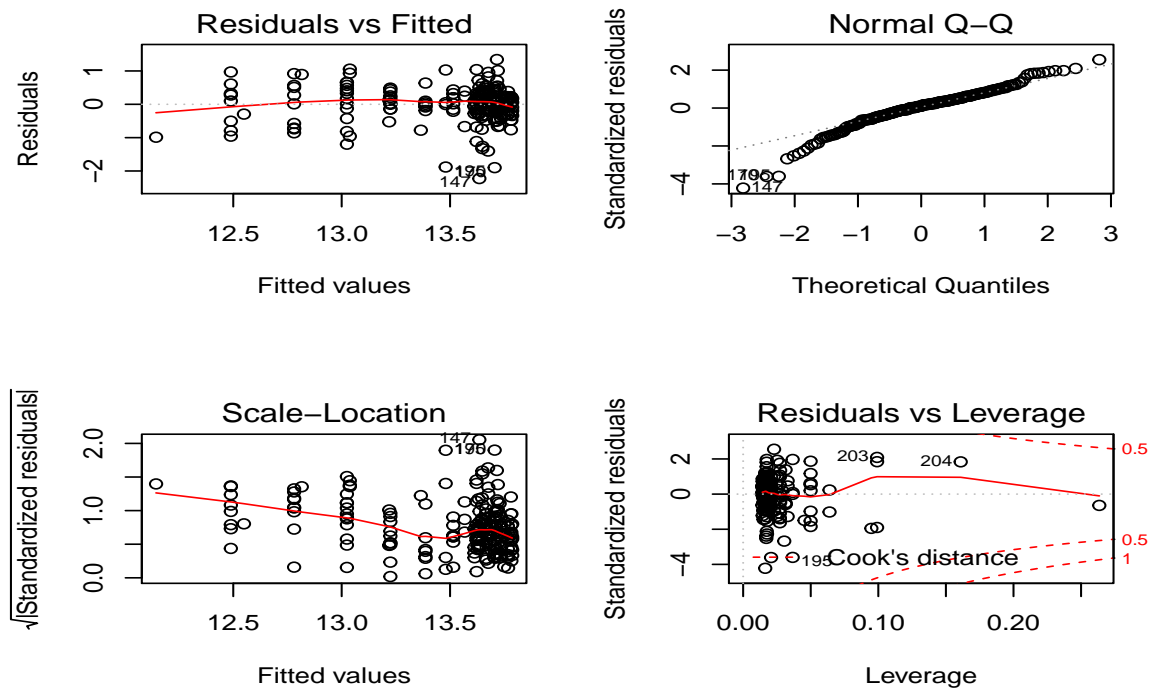
4. Estimation Methods of Semi/Nonparametric Regression Models

Estimating the unknown function nonparametrically means that the data itself is used to estimate the function, $f(\cdot, \cdot)$, that describes the relationship between the explanatory variables and the response. There are two commonly used approaches to estimate the nonparametric regression term:

1. Kernel Regression: estimates the conditional expectation of y at a given value x using a weighted filter to the data.
2. Spline Smoothing: minimize the sum of squared residuals plus a term which penalizes the roughness of the fit.

4.1 Kernel Regression

Kernel Regression smoothing is one of the most popular methods for nonparametric kernel regression estimation that was proposed by Nadaraya (1965) and Watson (1964) and it is known as the Nadaraya-Watson estimator (also known as the local constant estimator), though the local polynomial estimator has emerged as a popular alternative. For more details in kernel smoothing, see Wand and Jones (1995). The local polynomial regression is an approach that combines the simplicity of the linear least squares regression with the flexibility of nonparametric regression by fitting simple models to localized subsets of the data to build up a function that describes the unknown function of the regression model, point by point. At each point in the data range, a p -degree polynomial is fitted to a subset of the data, with explanatory variable values near the point whose response is being estimated. The polynomial is fitted using weighted least squares by giving more weight to points near the point whose response is being estimated and less weight to the points further away. The subsets of data used for each point is determined by an input to the procedure called the "bandwidth" that determines how much of the data is used to fit each local polynomial. If the local polynomial that is used is of the first degree, so it is locally linear (a straight line is fitted) and if a zero-degree polynomial is used so it is a weighted moving average or a local constant.

Figure 6. Residuals plots of the linear model, $y = \beta_0 + \beta_1 x$.Figure 7. Residuals plots of the 4th degree polynomial model, $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$

Assume that we have the following model:

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \dots, n. \quad (7)$$

One can show that the p th-order Taylor series expansion for the nonparametric function, $f(x_i)$, assuming that the $(p+1)$ th

derivative of the conditional mean at the point x exists, can be written as

$$y_i = f(x) + (x_i - x) \frac{\partial f(x)}{\partial x} + (x_i - x)^2 \frac{\partial^2 f(x)}{\partial x^2} \frac{1}{2!} + \dots + (x_i - x)^p \frac{\partial^p f(x)}{\partial x^p} \frac{1}{p!} + \epsilon_i. \quad (8)$$

This equation can be estimated by minimizing the following quantity:

$$\sum_{i=1}^n [y_i - \sum_{j=0}^p \frac{\partial^j f(x)}{j!} (x_i - x)^j]^2. \quad (9)$$

This quantity is a function of the derivative of the unknown function which is also unknown. A solution for this problem is to set $\beta_j = \frac{\partial^j f(x)}{j!}$ and, in this case, the above quantity will be a linear regression problem and can be rewritten as

$$\sum_{i=1}^n [y_i - \sum_{j=0}^p \beta_j (x_i - x)^j]^2. \quad (10)$$

Using the weighted least square method, the estimation of the nonparametric function is the estimate that minimizes the following quantity

$$\sum_{i=1}^n K_h(x_i - x) \{y_i - [\beta_0 + \beta_1(x_i - x) + \dots + \frac{1}{j!} \beta_p (x_i - x)^p]\}^2, \quad (11)$$

When $p = 0$, equation (8) will be

$$y_i = f(x) + (x_i - x) \beta_1(x) + \epsilon_i, \quad (12)$$

and the estimate of the nonparametric function at the value x is the weighted average of y_1, y_2, \dots, y_n . This estimator is called Nadaraya-Watson estimator or local constant and takes the form:

$$\hat{f}_h(x) = \frac{\sum_{i=1}^n K(\frac{x-x_i}{h}) y_i}{\sum_{i=1}^n K(\frac{x-x_i}{h})} \quad (13)$$

where K is a Kernel function (weight function) with a bandwidth h . The K function assigns weights to the values around x and these weights decline as the values get further from the target value (x). When $p = 1$, equation (8) will be

$$y_i = f(x) + (x_i - x) \beta_1(x) + (x_i - x)^2 \beta_2(x) + \epsilon_i \quad (14)$$

or

$$y_i = \beta_0 + (x_i - x) \beta_1(x) + (x_i - x)^2 \beta_2(x) + \epsilon_i \quad (15)$$

and the solution of this regression equation parameters $(\beta_0, \beta_1, \beta_2)$ is called the local polynomial estimator which is affected by the smoothing parameter h that is called "bandwidth". Loader (1999) explained the cross validation (CV), $CV(h) = \sum_{i=1}^n [y_i - \hat{f}_{h,-i}(x_i)]^2$, and the generalized cross-validation (GCV) methods that can be used to select the bandwidth. In case the Gaussian function is used, the optimal h that minimizes the cross validation quantity takes the form

$$h = \left(\frac{4\hat{\sigma}^5}{3n} \right)^{\frac{1}{5}} \approx 1.06 \hat{\sigma} n^{-1/5},$$

where $\hat{\sigma}$ is the standard deviation of the samples and n is the sample size.

Some of the common kernel functions are listed below

- Epanechnikov: $K(x_i) = \frac{3}{4}(1 - d^2)$, $d^2 < 1$, 0 otherwise,
- Minimum var: $K(x_i) = \frac{3}{8}(1 - 5d^2)$, $d^2 < 1$, 0 otherwise,
- Gaussian density: $K(x_i) = \exp(-\frac{d^2}{h})$
- Tricube function: $K(x_i) = (1 - |z|^3)^3$ for $|z| < 1$ and 0 otherwise.

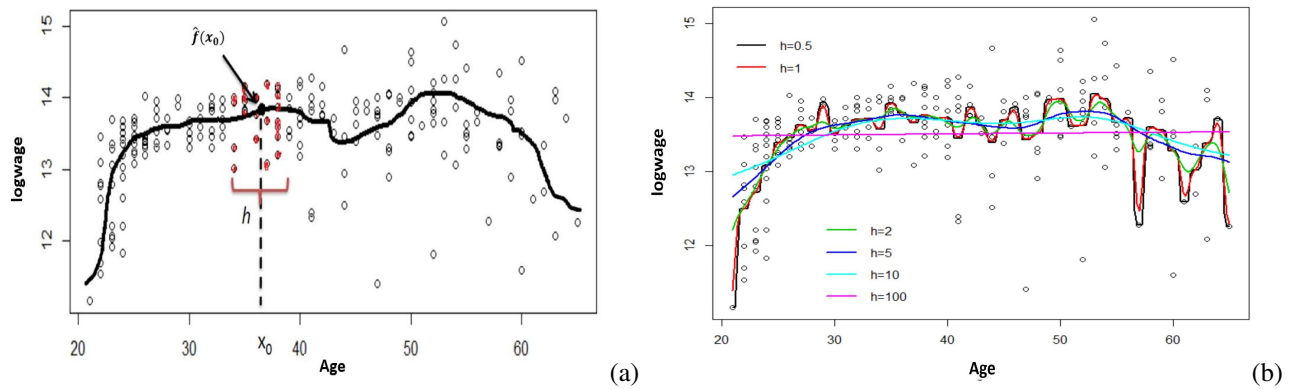


Figure 8. The scatter plot and kernel smoothing function using the optimal value of h (a), and the smoothing function at different values of the bandwidth, h . The Kernel smoothing *Ksmooth* function in **R** Package is used

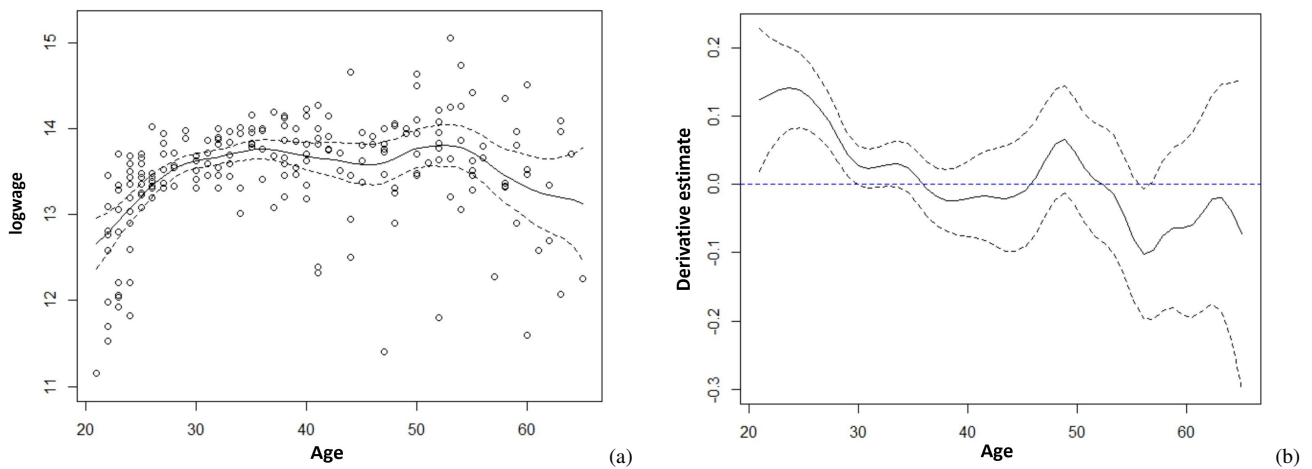


Figure 9. The unknown function estimate along with the 95% confidence interval (a), and the derivative of the smoothed function along with the 95% confidence interval (b). The *npreg* **R** function is used

where $K(x_i)$ is the kernel function that assigns a weight for x_i ($i = 1, 2, \dots, n$) based on its distance (d) between x_i and the target value x .

Bandwidth has an impact on the estimation of the nonparametric function. Figure 8(a) shows the idea of bandwidth and Figure 8(b) shows the effect of the bandwidth on the smoothness of the estimated function. The derivative of the smoothed function can be used to see whether the curvature is significant. Figure 9 shows the smoothed estimated function and its derivative along with 95% confidence interval using the optimal value of bandwidth. The figure shows that the function increases and at some point it gets constant in slope.

The behavior of the Nadaraya-Watson regression estimator ($p = 0$) for points in the boundary region is unsatisfactory compared to the local polynomial estimator ($p = 1$). This phenomenon is referred to as “boundary effects”. Kheireddine et al. (2015) showed under some conditions: $E(y^2) \leq \infty$, $E(x^2) \leq \infty$, and $f(x)$ is twice continuously differentiable in a neighborhood of x , that the bias of the Nadaraya-Watson is

$$\text{Bias}(\hat{f}(x)) = \frac{h^2 \mu_2(K)}{2} \left[\frac{\partial^2}{\partial x^2} f(x) + 2 \frac{\frac{\partial}{\partial x} f(x) \frac{\partial}{\partial x} f_y(x)}{f_y(x)} \right] + O\{h^2\}, \quad (16)$$

and of the local linear estimator is

$$\text{Bias}(\hat{f}(x)) = \frac{h^2 \mu_2(K)}{2} \frac{\partial^2}{\partial x^2} f(x) + O\{h^2\}, \quad (17)$$

where K is an integrable smoothing nonnegative kernel function, $f(x)$ is a nonparametric function, and $f_y(x)$ is the data density function.

By comparing these two formulas of bias, one can conclude that the bias decreases with the bandwidth h for both the local constant and the local polynomial estimators. The smaller the bandwidth value, the smaller the bias. In addition, the local

constant estimator is affected by $\frac{\partial}{\partial x}f(x)$, $\frac{\partial}{\partial x}f_y(x)$, and $f_y(x)$ which are not present in the bias formula of the local linear estimator.

4.2 Spline Smoothing

Spline is a piecewise polynomial. The polynomials of order j are pieced together at a sequence of knots ($\theta_1 < \theta_2 < \dots < \theta_C$) such that spline and its first $j-1$ derivative of the spline are continuous at those knots. In spline estimate, the unknown function is approximated by a power series of degree p , $f(x) \approx \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots + \beta_p x^p + \sum_{c=1}^C \beta_{1c} (x - \theta_c)_+^p$, where $(x - \theta_c)_+ = x - \theta_c$, $x > \theta_c$ and 0 otherwise.

In spline estimation, the vector of parameters, β , is obtained by minimizing the following quantity:

$$\sum_{i=1}^n [y_i - \hat{f}(x_i)]^2 + \lambda \int [\hat{f}''(x)]^2 dx, \quad (18)$$

where λ is a tuning parameter controls smoothness and $\hat{f}''(\cdot)$ is the second derivative of the nonparametric function. There are many methods for choosing λ , such as CV and GCV (Wahba 1977). For more details in spline smoothing, see Wang (2011).

How many knots need to be used? and where those knots should be located? A possible solution is to consider fitting a spline with knots at every data point, so it could fit perfectly and estimate the parameters by minimizing the usual sum of squares plus a roughness penalty. The smoothing parameter λ has the same effect on smoothing such as h in the kernel smoothing approach. When λ is very big, the smoothing line would be similar to linear regression line and when it is small, the smoothing line would be wiggly, more fluctuated. Figure 10(b) shows the effect of the smoothing parameter on the smoothness of the estimated function and Figure 10(a) shows the smoothed function at the optimal value of λ . Figure 11(a) shows the Spline smoothed estimated function using the optimal λ along with 95% confidence interval. Spline smoothing of the unknown function and its derivative can be used to see whether the curvature is significant. Figure 11(b) shows the smoothed derivative function along with 95% confidence interval using the optimal value of λ .

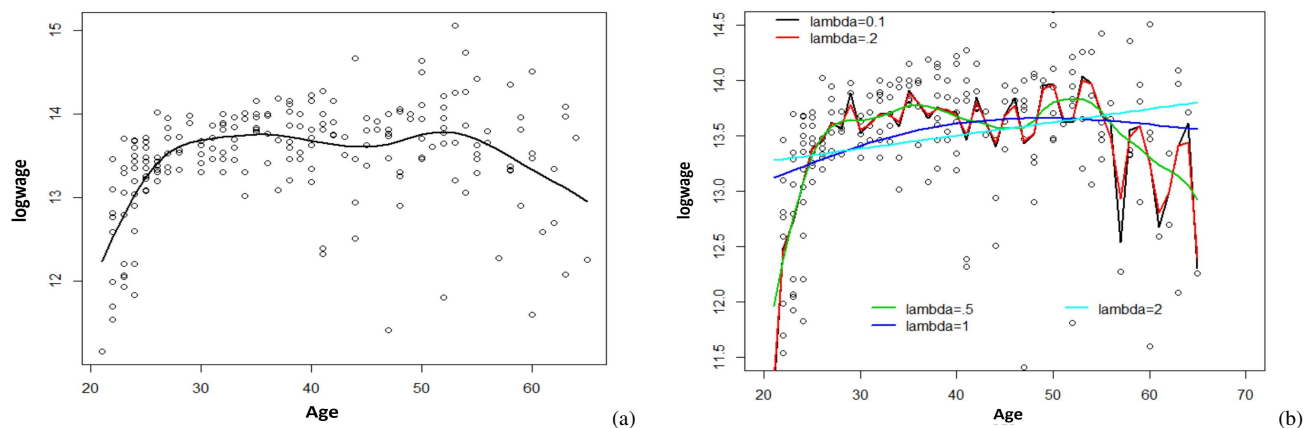


Figure 10. The estimate of the unknown function using the optimal λ that is determined by GCV criteria (a), and the estimates at different values of the penalty, λ

There are many R functions that can be used to estimate the semi/nonparametric regression models by the kernel and smoothing splines techniques. Table 1 displays the R packages and functions for estimating the nonparametric function.

5. Multiple Case

The kernel and spline smoothing approaches can be extended to any number of explanatory variables. Assume that a researcher wants to study the prestige as the response variable with education and income as the explanatory variables. Figure 1 shows that the relationship between education and prestige is linear and the relationship between income and prestige is nonlinear (unknown). So the following semiparametric model can be assumed

$$y = \beta_0 + \beta_1 \text{education} + f_2(\text{income}, \beta_2) + \epsilon, \quad (19)$$

where $f(\cdot, \cdot)$ is an unknown function. Smoothing splines or kernel regression can be used to fit this nonparametric term. Figure 12 shows the estimated two functions: a linear function of prestige and education and the nonparametric function

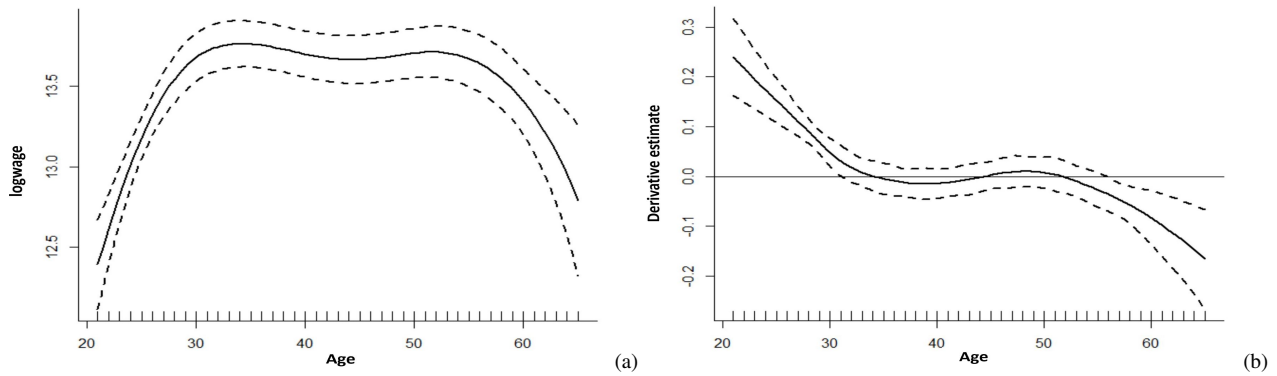


Figure 11. Estimate of the unknown function (a), and its derivative with 95% confidence interval (b) using the *spm* function in **R** software at the optimal smoothing parameter, λ , determined by REML criteria. Number of knots used is equal to 10 and $\lambda = 25.65$

Table 1. Different **R** packages and functions for estimating the nonparametric function and its derivative

	Package	Function	Smooth parameter
Local Polynomial	locfit	locfit	GCV
Kernel smoothing	KernSmooth	locpoly	—
	lokern	glkerns	plug-in
	lpridge	lpridge	—
	np	npreg	—
	stats	ksmooth	—
Spline	pspline	smooth.Pspline	CV/GCV
	sfsmisc	D1D2	GCV
	SemiPar	spm	RE(ML)

of prestige and income. Also the two relationships can be assumed unknown and the following model can be assumed

$$y = f_1(\text{education}, \beta_1) + f_2(\text{income}, \beta_2) + \epsilon, \quad (20)$$

where $f_1(\cdot, \cdot)$ and $f_2(\cdot, \cdot)$ are nonparametric functions. Figure 10 show the estimated nonparametric functions assuming both relationships or functions are unknown.

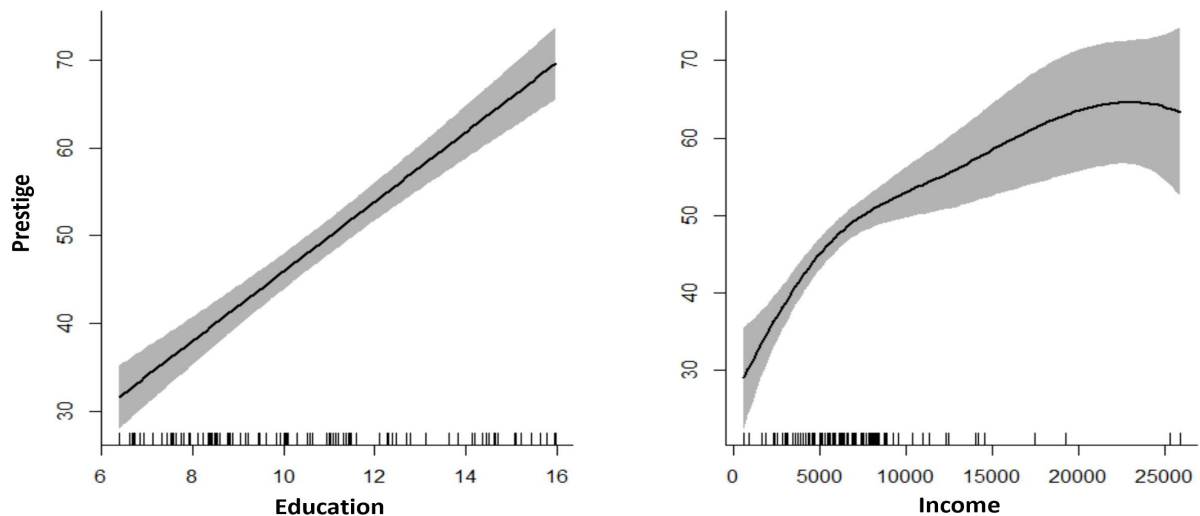


Figure 12. Estimated relationship between education, income and prestige using smoothing spline for the model $y = \beta_0 + \beta_1 \text{education} + f_2(\text{income}, \beta_2) + \epsilon$

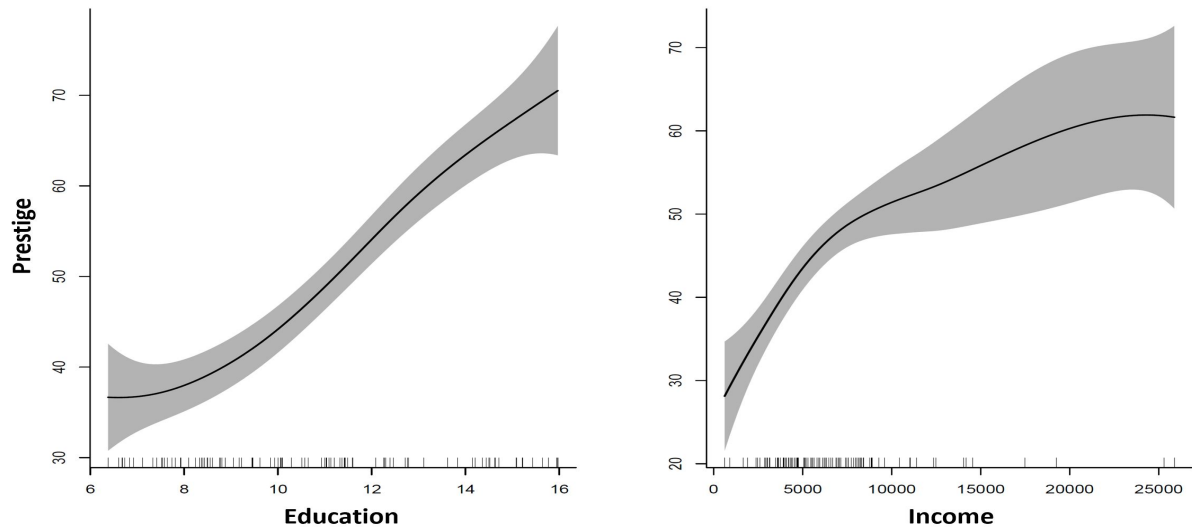


Figure 13. Estimated relationship between education, income and prestige using smoothing spline for the model $y = f_1(\text{education}, \beta_1) + f_2(\text{income}, \beta_2) + \epsilon$

As another example for the multiple case, a real data from Wooldridge (2019) is considered. This data has quantitative and categorical variables measured for 526 observations. The data is available in **R** in “np” **R** package. The response variable is $\log(\text{wage})$ and the explanatory variables are: educ (years of education), exper (the number of years of experience), and tenure (the number of years of current employment), and female (Female or Male). The scatterplots in Figure 14 show the relationships of the explanatory variables with the response. A researcher can assume the relationships between the response and quantitative variables are unknown and estimate them nonparametrically by fitting the following model

$$y = \beta_0 + \beta_1 \text{female} + f_1(\text{educ}, \beta_1) + f_2(\text{exper}, \beta_2) + f_3(\text{tenure}, \beta_3) + \epsilon, \quad (21)$$

where the categorical variable (female) is considered as a factor and the the form of the other relationships, $f_1(\cdot, \cdot)$, $f_2(\cdot, \cdot)$, and $f_3(\cdot, \cdot)$, are unknown. The output is displayed below and the smoothed functions are displayed on Figure 15.

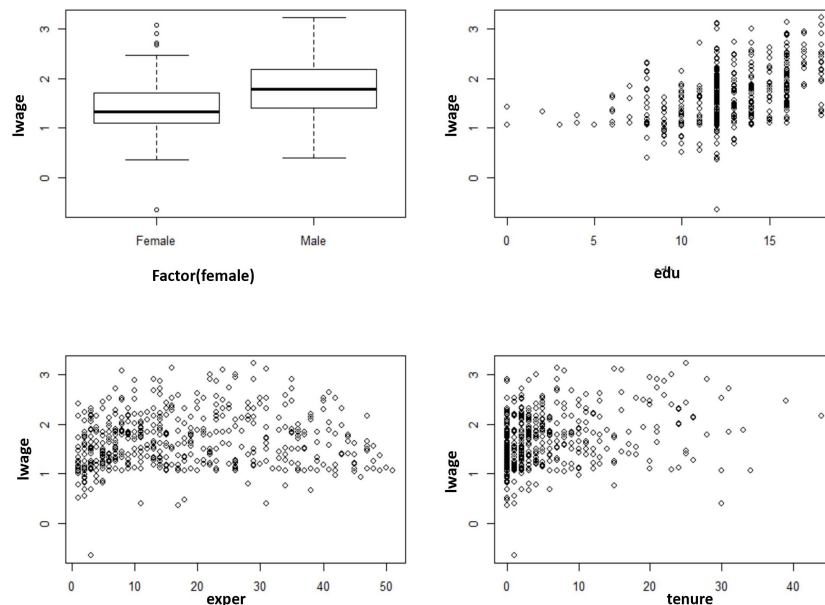


Figure 14. Scatterplots of the wage data variables

Another structure of the semiparametric regression modeling that may be assumed to model this data is the single index

Regression Data (526 observations, 4 variable(s)):

Regression Type: Local-Linear
 Bandwidth Selection Method: Least Squares Cross-Validation
 Formula: $\text{l wage} \sim \text{factor}(\text{female}) + \text{educ} + \text{exper} + \text{tenure}$
 Bandwidth Type: Fixed
 Objective Function Value: 0.157068 (achieved on multistart 4)

Exp. Var. Name: factor(female)	Bandwidth: 1.948785e-05	Lambda Max: 0.5
Exp. Var. Name: educ	Bandwidth: 4.530705	Scale Factor: 7.479181
Exp. Var. Name: exper	Bandwidth: 6.036449	Scale Factor: 1.088541
Exp. Var. Name: tenure	Bandwidth: 20.42454	Scale Factor: 16.85818

Continuous Kernel Type: Second-Order Gaussian
 No. Continuous Explanatory Vars.: 3

Unordered Categorical Kernel Type: Aitchison and Aitken
 No. Unordered Categorical Explanatory Vars.: 1
 Estimation Time: 14.78 seconds

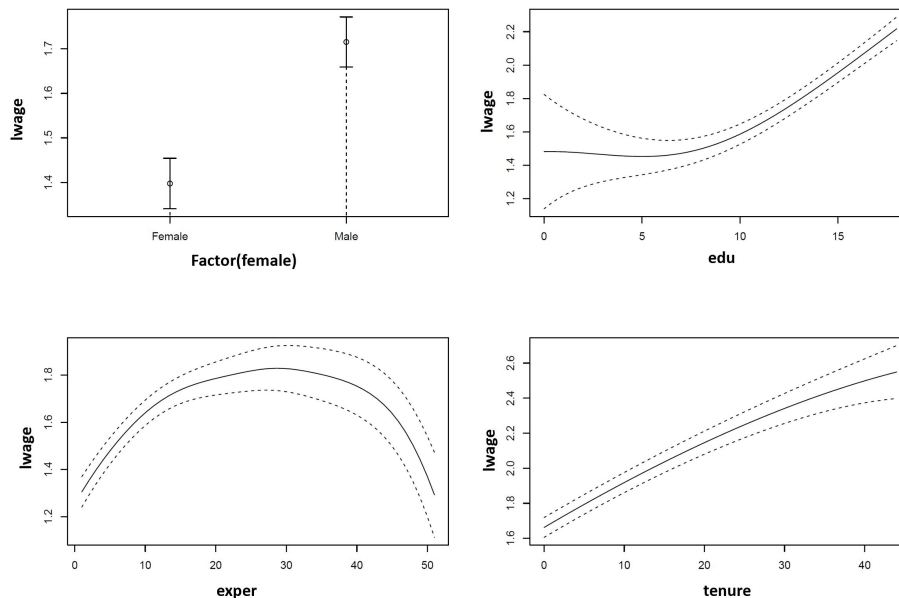


Figure 15. Smoothed estimated functions of wage data for the model
 $y = \beta_0 + \beta_1 \text{female} + f_1(\text{educ}, \beta_1) + f_2(\text{exper}, \beta_2) + f_3(\text{tenure}, \beta_3) + \epsilon$

model that takes the following form

$$y = f(\beta_1 \text{tenure} + \beta_2 \text{female} + \beta_3 \text{educ} + \beta_4 \text{exper}) + \epsilon, \quad (22)$$

where $f(\cdot)$ is an unknown function, and $\beta^T = (\beta_1, \dots, \beta_4)$ is the single index coefficients. The estimated parameters, R-square, and optimal bandwidth are displayed below and the smoothed fitted function, $\hat{f}(\cdot)$, is showed in Figure 16.

Single Index Model
 Regression Data: 526 training points, in 4 variable(s)

educ	exper	tenure	factor(female)
Beta: 1	0.002900096	0.008364508	0.2378936
Bandwidth: 0.08125926			
Kernel Regression Estimator: Local-Constant			

Residual standard error: 0.3904416
 R-squared: 0.4632725

Continuous Kernel Type: Second-Order Gaussian
 No. Continuous Explanatory Vars.: 1

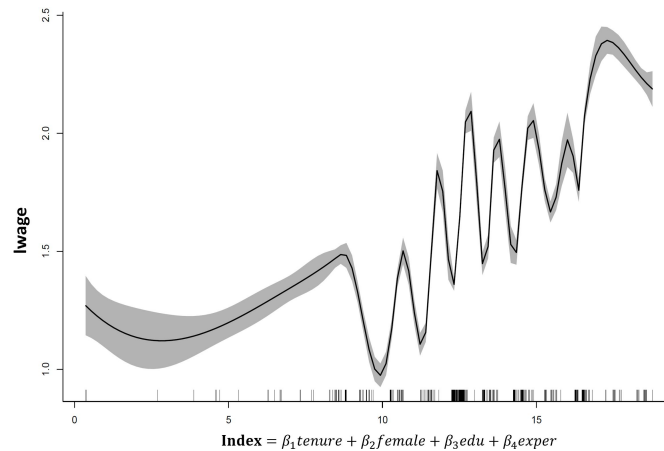


Figure 16. Smoothed estimated single index function of wage data for the model

$$\mathbf{y} = f(\beta_1 \text{tenure} + \beta_2 \text{female} + \beta_3 \text{edu} + \beta_4 \text{exper}) + \epsilon$$

6. Robust Nonparametric Estimation

Outliers may affect the kernel smoothing or spline smoothing estimation. So robust estimation is needed in this case. To estimate the unknown function at a point value x under the kernel smoothing method, the p th-degree polynomial is estimated using the weighted least squares method and the fitted value is calculated. Nadaraya-Watson estimator can be used as described in equation (13) and the quantity in equation (11) is minimized to obtain the estimate at the value x . To downweight the effect of the unusual values, the kernel weights need to be adjusted by incorporating the residuals in calculating the weights. In this case, a low weight will be assigned to large residuals. The general method for the robust linear model is m-estimation that is introduced by Huber (1964). This method can be developed for the nonparametric model case as follows.

By multiplying the kernel weights that control the smoothness by assigning weights that are proportional to the distance between the target value x and the other values x_i ($i = 1, 2, \dots, x_n$) by the weights that assigned for the values based on their residuals, big weights for small residuals and small weights for large residuals, we obtain robust weights that take the following form:

$$K_{x_i}^{\text{robust}}(x) = K_{x_i}^d(x) * K_{x_i}^r(x), \quad i = 1, 2, \dots, n, \quad (23)$$

where $K_{x_i}^d(x)$ is the kernel weight that is assigned to x_i based on its distance from x , and $K_{x_i}^r(x)$ is the weight that is assigned to x_i based on its residual. $R_j(x)$ is expected to downweight the effect of unusual values on estimation. The weight based on residuals can be obtained from the derivative of some appropriately chosen smooth convex loss function such as Huber or bisquare function. Below are the forms of these two functions:

Huber function

$$\rho(r) = \begin{cases} \frac{1}{2}r^2 & |r| \leq c \\ c|r| - \frac{1}{2}c^2 & |r| > c \end{cases} \quad (24)$$

Bisquare function

$$\rho(r) = \begin{cases} \frac{c^2}{6} [1 - [1 - (\frac{r}{c})^2]^3] & |r| \leq c \\ 0 & |r| > c \end{cases} \quad (25)$$

How to obtain $K_{x_i}^r(x)$? It is a function of the rescaled residuals and takes the following form

$$K_{x_i}^r(x) = \frac{\rho' \{ [y_i - \hat{f}(x_i)] / \hat{s} \}}{[y_i - \hat{f}(x_i)] / \hat{s}} = \frac{\rho'(r_i)}{r_i}, \quad i = 1, 2, \dots, n, \quad (26)$$

where y_i is the i th observation of the response variable, $\hat{f}(x_i)$ is the estimated function at the value x_i , \hat{s} is a robust estimation of the variance, such as the interquartile range of the residuals, and $r_i = [y_i - \hat{f}(x_i)] / \hat{s}$ is the rescaled residual of the value x_i . When x_i is an outlying value, r_i would be large, so the weight $K_{x_i}^r(x)$ would be small.

To evaluate the performance of the robust weights compared to kernel weights, 100 observations are generated from $f(x) = \{1 + e^{-20(x-0.5)}\}^{(-1)}$ where x is generated from Uniform(0, 1) and two outliers or unusual data points are added at (0.8, 0.6) and (0.75, 0.62) manually to the data. Figure 17 shows the estimated function with the outliers that are added to the data using the kernel smoothing, spline smoothing, and kernel robust weights. It reveals that the kernel and spline smoothing are not robust within the interval that contains the outliers but the robust estimation is not affected by these added outliers.

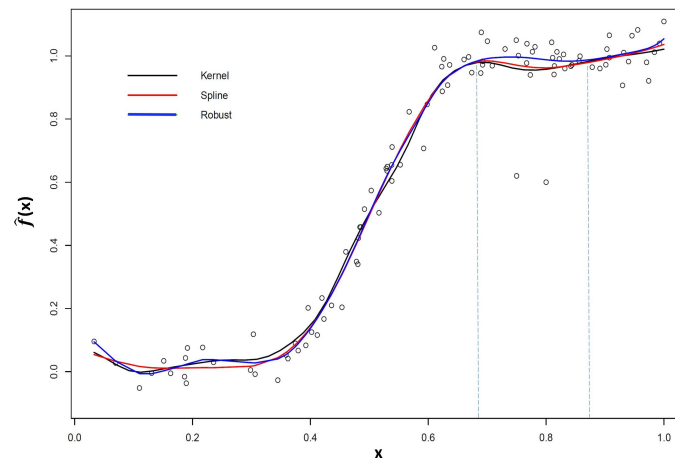


Figure 17. Simulated data from $f(x) = \{1 + e^{-20(x-0.5)}\}^{(-1)}$ along with two added outlying points, (0.8, 0.6) and (0.75, 0.62), and the smoothed function by kernel smoothing, spline smoothing and robust smoothing of the nonparametric regression function using the same bandwidth value, $h = 0.046$

7. Discussion and Conclusion

In this article, three types of modeling are discussed by examples: parametric, semiparametric, and nonparametric modeling. Nonparametric and semiparametric regression models are flexible compared to parametric models. Semiparametric is a hybrid of both parametric and nonparametric which allows having the best of both worlds: a model that is understandable and offering a fair representation of the data in the real life. However, semi/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. semi/nonparametric are flexible and avoid misleading results when we use the wrong model. The local linear estimator of the nonparametric function is less biased compared to the local constant estimator. One of the well-known semiparametric models is single index model. It assumes the link between the mean response and the explanatory variables is unknown and estimates it nonparametrically. This gives the single index model two main advantages over parametric and nonparametric models: (1) it avoids misspecifying the link function and its misleading results and (2) the reduction of dimension which is achieved by assuming the link function to be a univariate function applied to the projection of explanatory covariate vector onto some direction. Semi and nonparametric regression methods are not resistant to outliers, a robust method is introduced. Nonparametric regression analysis relaxes the assumption of linearity, substituting the assumption of a smooth regression function. The cost of relaxing the assumption of linearity is much greater computation and, in some instances, a more difficult to understand the result. The R code for all the graphs and analyses presented here, in this article, is available in the Appendix.

Acknowledgements

The author would like to thank the editor and the anonymous referees for their valuable comments and suggestions.

References

- Atkinson, K. (1989). *An Introduction to Numerical Analysis* (2nd ed.). John Wiley Pub.
- Dhekale, B. S., Sahu, P. K., Vishwajith, K. P., Mishra, P., & Narsimhaiah, L. (2017). Application of parametric and nonparametric regression models for area, production and productivity trends of tea (*Camellia sinensis*) in India. *Indian Journal of Ecology*, 44(2), 192-200.
- Fan, J., & Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*. New York, NY: Springer.
- Horowitz, J. L., & Hrdle, W. (1996). Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates. *Journal of the American Statistical Association*, 91(436), 1632-1640.
<https://doi.org/10.1080/01621459.1996.10476732>

- Huber, P. J. (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, 35, 73-101. <https://doi.org/10.1214/aoms/1177703732>
- Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, 58, 71-120. [https://doi.org/10.1016/0304-4076\(93\)90114-K](https://doi.org/10.1016/0304-4076(93)90114-K)
- Kheireddine, S., Sayah, A., & Yahia, D. (2015). General method of boundary correction in kernel regression estimation. *Afrika Statistika*, 10(1), 739-750. <https://doi.org/10.16929/as/2015.729.66>
- Li, J., Huang, C., Hongtu, Z., & for the Alzheimers Disease Neuroimaging Initiative. (2017). A functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association*, 112(519), 1169-1181.
- Lin, W., & Kulasekera, K. B. (2007). Identifiability of single index models and additive index models. *Biometrika*, 94, 496-501. <https://doi.org/10.1093/biomet/asm029>
- Loader, C. (1999). Bandwidth selection: classical or plug-in?. *The Annals of Statistics*, 27(2), 415-438. <https://doi.org/10.1214/aos/1018031201>
- Mahmoud, H. F. F., Kim, I., & Kim, H. (2016). Semiparametric single index multi change points model with an application of environmental health study on mortality and temperature. *Environmetrics*, 27(8), 494-506.
- Mahmoud, H. F. F., & Kim, I. (2019). Semiparametric spatial mixed effects single index models. *Computational Statistics & Data Analysis*, 136, 108-112. <https://doi.org/10.1016/j.csda.2019.01.008>
- Nadaraya, E. A. (1964). On estimating regression. *Theory of probability and its applications*, 9, 141-142. <https://doi.org/10.1137/1109020>
- Nocedal, J., & Wright, S. J. (1999). *Numerical Optimization*. New York NY: Springer. <https://doi.org/10.1007/b98874>
- Qin, J., Yu, T., Li, P., Liu, H., & Chen, B. (2018). Using a monotone single index model to stabilize the propensity score in missing data problems and causal inference. *Statistics in Medicine*, 38(8), 1442-1458.
- Rajaratnin, A., & Parmar, R. S. (2011). Application of parametric and nonparametric regression models for area, production and productivity trends of castor corn. *Asian Journal of Applied Sciences*, 4(1), 42-52.
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*. New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9780511755453>
- Sherman, R. P. (1994). Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators. *Annals of Statistics*, 22(1), 439-459. <https://doi.org/10.1214/aos/1176325377>
- Toma, A., & Fulga, C. (2018). Robust Estimation for the Single Index Model Using Pseudodistances. *entropy*, 20, 374. <https://doi.org/10.3390/e20050374>
- Wahba. (1977). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numerical Analysis*, 14, 651-667. <https://doi.org/10.1137/0714044>
- Wang, Y. (2011). *Smoothing splines: methods and applications*. FL: CRC Press, Boca Raton. <https://doi.org/10.1201/b10954>
- Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. London: Chapman and Hall. <https://doi.org/10.1007/978-1-4899-4493-1>
- Watson, G. S. (1964). Smooth regression analysis. *Sankhya, Series A*, 26, 359-372.
- Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach*. Boston, MA: Cengage Learning.
- Xia, Y., Li, W. K., Tong, H., & Zhang, D. (2004). A goodness-of-fit for singleindex models. *Statistica sinica*, 14, 1-39.

Appendix A

The R Code

```
library(SemiPar); library(np); library(car); library(stats)
library(graphics); data(fossil) # Load fossil data
attach(fossil)
fit = spm(strontium.ratio ~ f(age))
plot(fit)
points(age, strontium.ratio)
plot(age, strontium.ratio)
```



```

# -----
# Dataset cps71 from the np package, which consists of a random sample of
# wage data for 205 male individuals having common education (grade 13).
data(cps71)
attach(cps71)
str(cps71)          # gives information about the variables
cps71              # print the data
plot(age, logwage)
# -----
# Prestige of Canadian Occupations data set. This data have 6 variables:
# education, income, women, prestige, census, and type. Source: Canada
# (1971) Census of Canada.
data(Prestige)
attach(Prestige)
Prestige
plot(Prestige)
plot(education, prestige, xlab="Education", ylab="Prestige")
# -----
# ---- Fixing nonlinearity by using polynomial regression
# -----
par(mfrow=c(2,2))
plot(age, logwage, xlab="Age", ylab="log(wage)", main="")
linear.model=lm(logwage~age)
lines(age,fitted(linear.model), lty=1, lwd=2, col=4)
plot(linear.model) # check the assumptions
summary(linear.model)
plot(age, logwage, xlab="Age", ylab="log(wage)", main="")
Quad.model <- lm(logwage ~ age + I(age^2))
lines(age,fitted(Quad.model), lty=1, lwd=2, col=4)
plot(Quad.model)
cubic.model <- lm(logwage ~ age + I(age^2)+ I(age^3))
plot(age, logwage, xlab="Age", ylab="log(wage)", main="")
lines(age,fitted(cubic.model), lty=1, lwd=2, col=4)
plot(cubic.model)
Quart.model <- lm(logwage ~ age + I(age^2)+ I(age^3)+ I(age^4))
plot(age, logwage, xlab="Age", ylab="log(wage)", main="")
lines(age,fitted(Quart.model), lty=1, lwd=2, col=7)
plot(Quart.model)
summary(Quart.model)
anova(linear.model, Quart.model)
# -----
# ----- First method: Kernel regression
# -----
data(cps71)
attach(cps71)
Kernel.smooth <- npreg(logwage~age)
plot(Kernel.smooth, plot.errors.method="asymptotic",
     plot.errors.style="band", ylim=c(11,15.2),
     main="Estimated function and its 95% confidence band")
points(age, logwage)
summary(Kernel.smooth)
# ----- Another R function can be used to estimate
# ----- the unknown function using kernel regression
# ----- Effect of bandwidth on smoothing
plot(age, logwage)
lines(ksmooth(age, logwage, "normal", bandwidth=.5), col=1, lwd=2)
lines(ksmooth(age, logwage, "normal", bandwidth=1), col=2, lwd=2)
lines(ksmooth(age, logwage, "normal", bandwidth=2), col=3, lwd=2)
lines(ksmooth(age, logwage, "normal", bandwidth=5), col=4, lwd=2)
lines(ksmooth(age, logwage, "normal", bandwidth=10), col=5, lwd=2)
lines(ksmooth(age, logwage, "normal", bandwidth=100), col=6, lwd=2)
legend(20,15, c("h=0.5", "h=1"),lwd=2, col=1:2, bty="n")

```

```

legend(28,12.5, c("h=2", "h=5", "h=10","h=100"),lwd =2, col = 3:6, bty="n")
# ----- Smoothing splines
fit <- spm(logwage ~ f(age, degree=3), spar.method="ML")
plot(fit, col=1, shade=FALSE, ylab="logwage")
summary(fit)
plot(fit, se=FALSE)
plot(fit, drv=1, col=1, lwd=2, shade=FALSE, xlab="age", ylab="")
# -----
# ----- Second method: Spline regression
# -----
# ----- Study the effect of lambda on smoothing
plot(smooth.spline(age, logwage, spar = .1), xlab="age", ylab="logwage",
     type="l", xlim=c(20,70), ylim=c(11.5, 14.5), lwd=2)
points(age, logwage)
lines(smooth.spline(age, logwage, spar = .2), type="l", lwd=2, col=2)
lines(smooth.spline(age, logwage, spar = .5), type="l", lwd=2, col=3)
lines(smooth.spline(age, logwage, spar = 1), type="l", lwd=2, col=4)
lines(smooth.spline(age, logwage, spar = 2), type="l", lwd=2, col=5)
legend(20,14.7, c("lambda=0.1", "lambda=.2"),lwd =2, col = 1:2, bty="n")
legend(30,12.1, c("lambda=.5", "lambda=1"), lwd =2, col = 3:4, bty="n")
legend(50,12.1, c("lambda=2"), lwd =2, col = 5, bty="n")

# ----- Fit the function using the optimal soothing value
plot(age, logwage)
lines(smooth.spline(age, logwage), col=1, type="l", lwd=2)

spline=smooth.spline(age, logwage)
plot(spline, type="l", xlim=c(20,70), ylim=c(11, 15), lwd=2)
points(age, logwage)
spline
R2=cor(fitted(spline), logwage); print(R2)
# -----
# ----- Multiple Case
# ----- Kernel Regression -----
data(wage1); wage1
lwage=wage1$lwage; female=wage1$female; married=wage1$married; edu=wage1$edu
exper=wage1$exper; tenure=wage1$tenure
data=data.frame(lwage, female, married, edu, exper, tenure)
par(mfrow=c(2,2))
plot(female, lwage, type="p"); plot(edu, lwage); plot(exper, lwage)
plot(tenure, lwage)
bw <- npregbw(lwage ~ factor(female) + educ +
              exper + tenure, regtype = "ll", bwmethod = "cv.ls",
              data = wage1, tol = 0.1, ftol = 0.1)
npplot(bw, plot.errors.method = "asymptotic", common.scale = FALSE)
summary(bw)
# ----- nonparamtric Smoothing Spline -----
data(Prestige)
attach(Prestige)
library(mgcv)
plot(income, prestige)
plot(education, prestige)
mod.gam <- gam(prestige ~ s(income) + s(education))
mod.gam
plot(mod.gam)
# ----- Single index model -----
bw <- npindexbw(formula=lwage ~ edu+ exper +tenure + factor(female)
               , data=data)
model.index <- npindex(bws=bw, gradients=TRUE)
summary(model.index)
index=model.index$index; est=fitted(model.index)
plot(model.index)

```

```

plot(index, est)
fit=spm(est ~ f(index), spar.method="REML")
plot(fit)
# -----
data(Prestige)
attach(Prestige)
fit1=spm(prestige ~ education + f(income)) # using spline smoothing
fit1=spm(prestige ~ f(education) + f(income)) # using spline smoothing
plot(fit1)
par(mfrow=c(1,2))
plot(fit1, ylab="prestige")

fit2 = npindex(prestige ~ education + income, gradients=TRUE, data =Prestige)
plot(fit2, plot.errors.method="bootstrap",
     plot.errors.center="bias-corrected",
     plot.errors.type="quantiles")
# -----
#----- Robust Derivative Estimation technique
x=runif(100,0,1); y=(1+exp(-20*(x-0.5)))^-1 + rnorm(100,0,0.05)
ydash=- (1+exp(-20*(x-0.5)))^-2*(-20)*exp(-20*(x-0.5))
ss=data.frame(x, y, ydash)
newdata <- ss[order(ss[,1]),]
x=newdata[,1]; y=newdata[,2]; ydash=newdata[,3]
x[99]=0.8; y[99]=0.6 ; x[98]=0.75; y[98]=0.62
ss=data.frame(x, y)
newdata <- ss[order(ss[,1]),]
x=newdata[,1]; y=newdata[,2]
# ----- Kernel smoothing get affected by outliers
Kernel.smooth <- npreg(y~x)
plot(Kernel.smooth, main="", ylim=c(-0.1,1.1))
points(x,y)
summary(Kernel.smooth)
# ----- Spline soomth also get affected by outliers
cars.spl <- sm.spline(x, y)
lines(cars.spl, col = "red")

# ----- Robust kernel estimation
n=length(y) #sample size
ch=1.345
band=.046
delta=rep(1,n)
whiillr=c(); seyhat=c(); yhat=c(); w=c()
for (ii in 1:8){
  der=c(); est=c()
  for (i in 1:length(x))
  {
    u=x-x[i]
    h0j<-exp(-(u/band)^2)
    sumh<-sum(h0j)
    h0j<-h0j/sumh
    w0j=h0j*delta
    wxbar=sum(w0j*x)/sum(w0j)
    wssx=sum(w0j*(x-wxbar)^2)
    w0jllr=(w0j/sum(w0j)) + ((x[i]-wxbar)*w0j*(x-wxbar))/wssx
    whiillr[i]=w0jllr[i]
    seyhat[i]=sum(w0jllr^2)
    w[i]=delta[i]
    one=rep(1, length(x))
    X=cbind(one,u)
    W=diag(w0j)
    beta=solve(t(X)%*%W%*%X)%*%t(X)%*%W%*%y
    est[i]=beta[1]
  }
}

```

```
    der[i]=beta[2]
  }
  yhat=est
  r=y-yhat
  m=median(r)
  mad=(median(abs(r-m)))/.6745    #mad is the mad estimate of scale
  rs=r/mad                        #rs are the rescaled residuals
  for (j in 1:n){
    if (abs(rs[j]) <= ch){delta[j]=1} else {delta[j]=ch/abs(rs[j])}
  }
  all=cbind(x, y, yhat ,rs ,delta)
  print(all[1:10,])
}
dferror=sum(w)-sum(whiillr)      #error degrees of freedom
sigma2=sum(w*(y-yhat)^2)/dferror #estimate of sigma squared
tvalue=2                         #invcdf .975 tvalue
seyhat=sqrt(sigma2*seyhat)       #seyhat contains the se(yhat)
lclm=yhat-tvalue*seyhat
uclm=yhat+tvalue*seyhat
points(x, typ="l", yhat, col="blue", lwd=2)
#lines(x,lclm, col="blue", lwd=2)
#lines(x,uclm, col="blue", lwd=2)
```

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).