Scientific Research and Essays Vol. 7(27), pp. 2409-2416, 19 July, 2012
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE12.237

Full Length Research Paper

On the efficiency of some semi- and nonparametric kernel estimators in the modeling of binary response data

Özge Akkuş[1]*, Serdar Demir[1] and Hüseyin Tatlıdil[2]

[1]Department of Statistics, Muğla University, 48000, Muğla, Turkey.
[2]Department of Statistics, Hacettepe University, Ankara, Turkey.

Accepted 28 May, 2012

The semiparametric estimation procedure of binary response data is composed of two steps. In the first step, various semiparametric estimators are used for estimating the model parameters, whereas nonparametric regression type estimators are required in the second step to obtain the probability estimates dependent on the estimated model parameters in the first step. In this study, we have investigated the efficiency level of the Klein and Spady estimator which is widely used in the current literature in first step of the semiparametric modelling and the classical Nadaraya-Watson kernel estimator used in the second step of the estimation procedure of binary response data when the parametric model assumptions are satisfied. We have also wanted to see the variation in the estimates when the adaptive Nadaraya-Watson kernel estimator has been used instead of the classical estimator. So far, there has neither been any simulation study nor a study comparing those methods, analytically in Statistics literature. Therefore, a comprehensive simulation study has been conducted and data sets from the logistic distribution have been generated to display that success in practice. Four different sample sizes have been considered to see the differences along with the variation in the sample sizes. All findings have been assessed in terms of both the mean averaged square error and the correct classification rate criteria for ordinary, Pearson and deviance residuals, respectively. Additionally, a real data set has been used to demonstrate the effectiveness of the simulation results with the results in practice. The simulation results indicate that the semiparametric Klein and Spady estimations give considerable close results to the parametric counterparts when parametric model assumptions are satisfied. This obviously means that there is no significance difference between the parametric and the semiparametric approaches in case of satisfaction of the parametric model assumptions. Another considerable finding is that there is no clear superiority between classical and adaptive Nadaraya-Watson estimators.

Key words: Nonparametric regression, kernel estimation, semiparametric estimation, binary data.

INTRODUCTION

A model with both an unknown function and an unknown finite-dimensional parameter is called “semiparametric”. Since the semiparametric approach does not need much of the assumptions required by the parametric alternative, it has gradually become popular for the limited dependent variable in Econometrics. Besides, researchers necessarily use the semiparametric model approach when the parametric model assumptions are violated. We mainly focus on the semiparametric modeling of binary response data here.

The semiparametric estimation procedure of binary response data is composed of two steps: In the first step, one of the semiparametric estimators such as the semiparametric least square estimator proposed by

*Corresponding author. E-mail: ozge.akkus@mu.edu.tr.

(Ichimura, 1993), the semiparametric maximum likelihood estimator of Klein and Spady (1993) or the density weighted average derivative estimator of (Powell et al., 1989) could be used for the estimation of the model parameters.

It is well known that the IC estimator outperforms when the dependent variable is continuous. However, PW estimator could or should be used for the model including only discrete explanatory variables, which is extremely a restrictive requirement. That is why we are interested in the performance of the semiparametric KS estimator which could be used for the data types of both discrete and continuous.

In the second step, the semiparametric estimation procedure is completed by the application that is one of the nonparametric regression estimation procedures of the binary dependent variable on the linear combination of the model parameters estimated in the first step. Since the classical Nadaraya-Watson (NW) estimator has a simple mathematical form, it is generally preferred to be used in the estimation process.

This study is composed of two parts. In the first part of the study, the efficiency degree of the traditional semiparametric KS estimator has been determined when the parametric model assumptions are satisfied in the first step of the semiparametric modelling of binary response data. Additionally, parametric Probit (PR) model estimates have also been obtained to see the deviations of the KS estimates from that of the true parametric PR estimates.

In the second part of the study, we show and propose that Adaptive Nadaraya-Watson (NWA) kernel estimator may be used in place of the NW estimator in the semiparametric estimation of binary response data. We show that the proposed NWA estimator presents pretty much similar results to the NW estimator.

So far, in Statistics literature, there has neither a simulation work nor a study comparing those methods, analytically. For this purpose, a comprehensive simulation study has been conducted and data sets from the logistic distribution have been generated. This means that, much of the results obtained from the generated data perform in favour of the parametric model estimates due to the known mathematical structure of the model.

Keeping this in mind, in the first part of the simulation design we have investigated whether the KS estimator gives considerable close results to the parametric counterparts or not when the data is suitable for the parametric estimation. After that, parametric PR estimates are obtained to see the difference in results of an alternative parametric model apart from the logit estimates, as well.

The second part of the simulation study is designed to discuss the performance of the nonparametric NW and NWA kernel estimators for the modeling of a categorical dependent variable that are commonly used in the nonparametric modeling of a continuous dependent variable.

All findings have been assessed in terms of both the Mean of Averaged Square Error (MASE) and the Correct Classification Rate (CCR) criteria for Gaussian (GAU) and Epanechnikov (EPA) kernel functions and ordinary, Pearson and deviance residuals, separately.

METHODS

In binary dependent variable modeling, the mean function that is conditional on the vector of the explanatory variables *X* given in Equation (1) has been defined as the probability (also denoted by *P*) of belonging of an observations to the category “1” coded in the dependent variable,

$$m(x) = E(Y \mid X = x) = P[Y = 1 \mid X = x)] = G_{\varepsilon}(X^T \hat{\beta}) \quad \text{.....................} \quad (1)$$

where $\hat{\beta}$ is the vector of the estimated parameters, Y is a binary dependent variable and G is the distribution of the error term ε .

The parametric approach

In this approach both the structure of the distribution and its parameters are known. Under the linear index restriction ($X^T\hat{\beta}$), $\hat{\beta}$ could easily be obtained owing to the known mathematical structure of the model.

Parametric logistic regression model (LG) is obtained under the assumption that G represents the logistic distribution function. Similar to this, the PR model is obtained under the normal distribution assumption for the function G. The mathematical expressions of both LG and PR models are given in Equation (2) and (3), respectively,

$$E(Y \mid X = x) = P[Y = 1 \mid X = x)]_{(LG)} = \frac{exp(X^T\hat{\beta})}{1 + exp(X^T\hat{\beta})} \quad \text{................} \quad (2)$$

$$E(Y \mid X = x) = P[Y = 1 \mid X = x)]_{(PR)} = \Phi(X^T\hat{\beta}) \quad \text{.................} \quad (3)$$

In Equation (3), Φ denotes the standard cumulative normal distribution function.

Model parameters (β's) are derived by maximizing the logarithmic likelihood function given in Eq. (4) (McCullagh and Nelder, 1989).

$$log\, L_N(\beta) = N^{-1} \sum_{n=1}^{N} \left\{ \begin{array}{l} Y_n \, log\, G(X_n^T \beta) \\ +(1 - Y_n) \\ log[1 - G(X_n^T \beta)] \end{array} \right\} \quad \text{...................} \quad (4)$$

The semiparametric approach

In the semiparametric modeling of binary response data, no assumption is required related to the distribution of the error term and “*g*” has been substituted in place of *G* in Equation (1). The

$(X^T\hat{\beta})$ linear index assumption is still valid here. The general expression of the model is defined as follows:

$$E(Y \mid X = x) = P[Y = 1 \mid X = x)] = g(X^T\hat{\beta}) \quad \text{.................} \quad (5)$$

The semiparametric estimation procedure is summarized as follows:

Step 1: β is estimated using one of the semiparametric estimators mentioned in the Introduction part of the study.

Since there is no need of distributional assumption for G, an alternative estimator is required in the semiparametric approach. Such an alternative estimator was proposed by Klein and Spady (1993).

KS estimator

Let G_N be in the form of equation as follows;

$$G_N(\upsilon) = \frac{P_N g_N(\upsilon \mid Y = 1)}{P_N g_N(\upsilon \mid Y = 1) + (1 - P_N) g_N(\upsilon \mid Y = 0)} \quad \text{..................} \quad (6)$$

Where Y is dependent variable in dummy structure.

Klein and Spady showed that G_N which is nonparametric estimation of G could be estimated by the nonparametric regression of Y on $\upsilon = X^T\hat{\beta}_{KS}$.

The proportion of responses "1" in overall; the kernel estimation of the density function of $\upsilon = X^T\hat{\beta}_{KS}$ for responses "1" and the kernel estimation of the density function of $\upsilon = X^T\hat{\beta}_{KS}$ for responses "0" are given by Equations (7), (8) and (9), respectively.

$$P_N = \frac{\sum_{n=1}^{N} y_n}{N} \quad \text{....................} \quad (7)$$

$$g_N(\upsilon \mid Y = 1) = \frac{1}{(NP_N h_N)} \sum_{n=1}^{N} y_n K\left[\frac{\upsilon - X_n^T\hat{\beta}_{KS}}{h_N}\right] \quad \text{...............} \quad (8)$$

$$g_N(\upsilon \mid Y = 0) = \frac{1}{[N(1 - P_N)h_N]} \sum_{n=1}^{N} (1 - y_n) K\left[\frac{\upsilon - X_n^T\hat{\beta}_{KS}}{h_N}\right] \quad \text{..............} \quad (9)$$

After both replacing G_N (υ) in place of G in Equation (4) and maximizing the log likelihood function, the unknown β parameters are estimated by the KS method.

Step 2: The linear function of $(X^T\hat{\beta})$ is computed after the estimation of β.

Step 3: g in Equation (5) is estimated using one of the nonparametric regression estimation methods, that is ,used to regress Y on the estimated $(X^T\hat{\beta})$. As a result, the probabilities of observations belonging to the category "1" in the dependent variable are estimated.

NW estimator

The general form of the conditional expected value given by Equation (1) is defined as:

$$\hat{m}(x) = \int \frac{y\hat{f}(x,y)}{\hat{f}(x)} dy \quad \text{............} \quad (10)$$

where $\hat{f}(x,y)$ is the estimated joint probability density function of X and Y and $\hat{f}(x)$ is the estimated density function of X. Nadaraya (1964) and Watson (1964) proposed an estimator for the estimation of $m(x)$ based on the kernel functions. The kernel estimators of $f(x)$ and $f(x,y)$ are given as:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \quad \text{............} \quad (11)$$

$$\hat{f}(x,y) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} \ddot{K}\left(\frac{x - X_i}{h_1}, \frac{y - Y_i}{h_2}\right) \quad \text{.........} \quad (12)$$

where h is a bandwidth (smoothing) parameter which controls for the smoothing level of the kernel estimation; K is a symmetrical probability density function called "kernel function"; $\ddot{K}(.)$ is a bivariate kernel function which can be obtained using the multiplicative kernel functions defined as follows (see Hardle, 1990; Horowitz and Hardle, 1994).

$$\ddot{K}\left(\frac{x - X_i}{h_1}, \frac{y - Y_i}{h_2}\right) = K\left(\frac{x - X_i}{h_1}\right) K\left(\frac{y - Y_i}{h_2}\right) \quad \text{...............} \quad (13)$$

and h_1 and h_2 are fixed bandwidths.

The use of the EPA (Epanechnikov, 1969) and GAU kernel functions given below are commonly preferred in practice due to the simplicity of their functional forms.

$$K(u)_{EPA} = 3(1 - u^2)/4 \quad ; \quad |u \leq 1|$$

$$K(u)_{GAU} = exp(-u^2/2)/\sqrt{2\pi}; \quad -\infty < u < \infty$$

Using the same bandwidth parameters (h_1= h_2=h) in Equation (12) and replacing the density functions of $\hat{f}(x)$ and $\hat{f}(x,y)$ in Equation (10) with their kernel estimates in Equations (12) and (13), the Nadaraya-Watson kernel estimator of the regression function, $\hat{m}_{NW}(x)$, is obtained.

$$\hat{m}_{NW}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)} \quad \text{..........} \quad (14)$$

h plays a very important role in the performance of the kernel estimators. Various methods such as cross-validation, penalized functions, plug-in, bootstrap etc. have been developed to obtain the optimal bandwidth parameter h (Pagan and Ullah, 1999). In fact the Cross Validation (CV) method has become very popular in the estimations due to its simplicity in selecting optimal h.

The optimal bandwidth value of h is obtained by minimizing the CV function given by Equation (15) with a nonnegative weight function $w(X_i)$ (Hardle, 1990; Horowitz and Hardle, 1994).

$$CV(h)=n\sum_{i=1}^{n}[Y_i-\hat{m}(X_i)]^2 w(X_i) \quad \text{………………..} \quad (15)$$

CV includes the leave-one-out kernel estimator obtained by leaving the observations X_i and Y_i out of the data defined as follows:

$$\hat{m}_i(X_i)=\frac{\sum_{j\neq i}^{n} Y_j K\left(\frac{X_i-X_j}{h}\right)}{\sum_{j\neq i}^{n} K\left(\frac{X_i-X_j}{h}\right)} \quad \text{………………..} \quad (16)$$

The bandwidth that minimizes the CV function also minimizes the mean square error which is a performance criterion of an estimator (Hardle, 1990; Horowitz, 1998).

NWA estimator

The kernel estimator of the probability density function may not be efficient especially when a fixed bandwidth parameter is used for the multivariate case. Silverman (1986) proposed a procedure called the “adaptive kernel” (or sample point) estimator based on the varying bandwidth parameter.

An adaptive kernel estimator of a probability density function which uses different bandwidth values for the data point of X_i, is given by:

$$\hat{f}_A(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h(X_i)}K\left(\frac{x-X_i}{h(X_i)}\right) \quad \text{……….} \quad (17)$$

where, $h(X_i)$ represents the varying bandwidth value for observation *i*. Using varying bandwidth parameters instead of the fixed bandwidths, the adaptive multiplicative kernel estimator of the variables $(X_1,X_2,...,X_d)$ (in *d*-dimensional space) could be defined as:

$$\hat{f}_A(x_1,...,x_d)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_{1i}\cdots h_{di}}\left\{\prod_{j=1}^{d}K\left(\frac{x-X_j}{h(X_j)}\right)\right\} \quad \text{………………...} \quad (18)$$

(Sain, 1994). The adaptive estimator of the bivariate density function given by Equation (19) is a special case of the function given in Equation (18) with two random variables.

$$\hat{f}_A(x,y)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h(X_i)h(Y_i)}K\left(\frac{x-X_i}{h(X_i)}\right)K\left(\frac{y-Y_i}{h(Y_i)}\right) \quad \text{…………} \quad (19)$$

A NWA kernel estimator with varying bandwidth parameter is obtained by replacing $\hat{f}_A(x)$ and $\hat{f}_A(x,y)$ into the numerator and denominator of Equation. (10). The resulting estimator is given as follows (Demir and Toktamis, 2010).

$$\hat{m}_{NWA}(x)=\int\frac{y\hat{f}_A(x,y)}{\hat{f}_A(x)}dy=\frac{\sum_{i=1}^{n}\frac{Y_i}{\lambda_i}K\left(\frac{x-X_i}{\lambda_i h}\right)}{\sum_{i=1}^{n}\frac{1}{\lambda_i}K\left(\frac{x-X_i}{\lambda_i h}\right)} \quad \text{……….} \quad (20)$$

In Equation (20), λ_i $(i=1,...,n)$ represents the local bandwidth factors. Silverman (1986) proposed an algorithm composing of three steps in the estimation of the adaptive estimators.

In the first step, a prior kernel estimator $\tilde{f}(X_i)$ with a fixed *h* is obtained. In the second step, λ_I is computed as:

$$\lambda_i=\left[\frac{\tilde{f}(X_i)}{\psi}\right]^{-\alpha} \quad \text{……………………………………………..............} \quad (21)$$

Where ψ is the geometric mean of the function $\tilde{f}(X_i)$ and α is the sensitivity parameter which satisfies the condition $0 \le \alpha \le 1$.

In the last step, the following adaptive kernel estimator of the density function of *X* is obtained by replacing $h\lambda_i$ instead of $h(X_i)$ in Equation (17).

$$\hat{f}_A(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h\lambda_i}K\left(\frac{x-X_i}{h\lambda_i}\right) \quad \text{…………} \quad (22)$$

The same procedure is applied to obtain the adaptive kernel estimator of the bivariate density function given in Equation (19) to reach the expression in Equation (20). The adaptive kernel estimation is equivalent to the kernel estimation with fixed bandwidth parameter when the sensitivity parameter α is equal to 0. Abramson (1982) and Silverman (1986) emphasized the point that α is considered to be 0.5 for obtaining better results.

Simulation study

This section contains the simulation planning. The linear index function given below has been determined following the study of (Proença and Silva, 2000) so that the minimal conditions could be satisfied for the semiparametric estimation. The identifiability condition of the model parameters proposed by (Manski, 1988) has been achieved by assigning the value 1 to the first coefficient of a continuous variable.

$$Index_i = X^T\beta = 1+X_{1i}+X_{2i}; \quad i=1,\dots,n$$

In order to reveal the effects of both continuous and discrete explanatory variables, X1 has been generated from standard normal distribution whereas a discrete variable X2 has been generated from Bernoulli distribution whose parameter is 0.75. The probabilities with respect to the index values have been computed using the ordinary logistic function presented below. This means that all simulated data are consistent with the model LG.

$$P(Y_i=1/X=x)=\frac{exp(Index_i)}{1+exp(Index_i)}; \quad i=1,\dots,n$$

Dependent variable *Y* has been derived from the Bernoulli distribution with the parameter {$P(Y_i=1/X=x)$}. The MASE and CCR criterion have been interpreted in terms of the Ordinary, Pearson and Deviance residuals, separately. The mathematical definitions of these residuals are given as:

$$r_{Ordinary(i)}=y_i-\hat{P}(x_i), \quad r_{Pearson(i)}=\frac{y_i-\hat{P}(x_i)}{\sqrt{\hat{P}(x_i)\left[1-\hat{P}(x_i)\right]}},$$

$$r_{Deviance(i)}=Sign\left[y_i-\hat{P}(x_i)\right]\sqrt{2\left[y_i\log\frac{y_i}{\hat{P}(x_i)}+(1-y_i)\log\frac{1-y_i}{1-\hat{P}(x_i)}\right]}$$

Table 1. The results based on the **KS** estimator and GAU and EPA kernel functions.

N	Kernel Est.	$MASE_{GAU}$			CCR_{GAU}	$MASE_{EPA}$			CCR_{EPA}
		Ordinary	Pearson	Deviance		Ordinary	Pearson	Deviance	
25	NW	0.08161	0.55571	0.23009	88.80	0.09233	0.64298	0.25959	87.53
	NWA	0.08201	0.55061	0.23124	88.85	0.09423	0.64142	0.26454	87.27
100	NW	0.09874	0.70276	0.27710	86.64	0.10397	0.76374	0.29183	86.01
	NWA	0.10028	0.69598	0.28032	86.43	0.10621	0.76344	0.29660	85.60
250	NW	0.11474	0.82499	0.31956	84.54	0.11697	0.85229	0.32578	84.26
	NWA	0.11613	0.82929	0.32264	84.28	0.11855	0.86141	0.32941	83.95
500	NW	0.12199	0.88646	0.33788	83.52	0.12344	0.90203	0.34190	83.33
	NWA	0.12304	0.89538	0.34018	83.28	0.12456	0.91494	0.34442	83.05

Table 2. The results based on the **PR** estimator and GAU and EPA kernel functions.

N	Kernel Est.	$MASE_{GAU}$			CCR_{GAU}	$MASE_{EPA}$			CCR_{EPA}
		Ordinary	Pearson	Deviance		Ordinary	Pearson	Deviance	
25	NW	0.08469	0.56279	0.23738	88.28	0.09123	0.60993	0.25431	87.37
	NWA	0.08501	0.55744	0.23825	88.18	0.09251	0.60636	0.25757	87.21
100	NW	0.10091	0.71853	0.28230	86.18	0.10427	0.76013	0.29163	85.78
	NWA	0.10213	0.70926	0.28476	85.94	0.10599	0.75584	0.29517	85.40
250	NW	0.11484	0.84959	0.31949	84.42	0.11668	0.84959	0.32469	84.14
	NWA	0.11599	0.85712	0.32283	84.10	0.11800	0.85712	0.32766	83.88
500	NW	0.12169	0.88683	0.33687	83.33	0.12295	0.90152	0.34040	83.28
	NWA	0.12253	0.89394	0.33864	83.05	0.09123	0.60993	0.25431	83.07

Where y_i denotes the observed value of the dependent variable whereas $\widehat{P}(x_i)$ is the estimated probability with respect to the *x* of observation *i* $(i = 1,\ldots,n)$.

Two popular kernel functions, GAU and EPA, have been used in computations with sample sizes of 25, 100, 250 and 500. For each type of sample size 1000 replications have been performed.

It is well known fact that the interpretation of the ordinary residuals are not appropriate in modeling the categorical dependent variable in spite of their smallest MASE values in comparison with the other type of residuals. The reason why we have included the ordinary residuals in our study is to show that using this type of residual will lead us to misinterpret the results. In the sub-section this has been presented in detail. The MASE values of the deviance residuals are smaller than the Pearson residuals in any sample sizes and for both kernel functions presented in Tables 1 and 2. Since the deviance is the smallest residual type it should be evaluated in any stage of the simulation study.

Tables 1 and 2 summarize the simulation results on the basis of the KS and PR estimators with GAU and EPA kernel functions for averaged square residuals when the data fit the logistic distribution.

RESULTS

This part of the study summarizes all the simulation results obtained.

1. The first comparison is made on the efficiency of the semiparametric estimator KS and the parametric estimator PR when the simulated data sets are consistent with the parametric model LG. We conclude that the KS results overlap with the results of the PR under the satisfied assumptions of the parametric model. In other words, there is no significant difference between the parametric and the semiparametric approaches to the modeling of binary response data when satisfactions of the parametric model assumptions are achieved.
2. Another comparison is made on the performance of the NW and NWA estimators in terms of the MASE values considering the deviance residuals. We conclude that there is no clear superiority of the estimators on

Table 3. The results of the cancer data.

Estimator	Kernel Estimator	Kernel Function	MSE			CCR (%)
			Ordinary	Pearson	Deviance	
KS	NW	GAU	0.21968	0.93329	0.54508	63.08
	NWA		0.21983	0.93660	0.54550	63.08
	NW	EPA	0.21807	0.92696	0.54054	63.08
	NWA		0.21868	0.94057	0.54229	64.62
PR	NW	GAU	0.13766	0.71611	0.36676	80.00
	NWA		0.13743	0.69351	0.36393	80.00
	NW	EPA	0.13909	0.75839	0.37116	81.54
	NWA		0.14147	0.72477	0.37198	80.00

efficiencies. However, when the exact results are taken into consideration, the MASE values of the NW estimator are smaller than that of the NWA. Thus, we suggest that researchers should use the NW estimator due to its simpler mathematical structure in place of the estimator of NWA.

3. It is also evident that the CCR criterion does not present distinguishable information for the concerned comparison.
4. Since the MASE values obtained from the use of the GAU kernel function are smaller than that of the EPA kernel in almost all sample sizes under the deviance residuals, a definite conclusion cannot be presented about the comparison of the efficiency on the kernel functions used. We indeed know and expect that kernel functions do not have a significant effect on the results (Hardle et al., 2004).
5. Another remarkable point that should be emphasized is that CCR values get smaller as the sample size increases contrary to expectations. This is related to the asymptotic consistency of the estimators. The consistency of an estimator is dependent on some restrictive conditions and it is not so easy to satisfy such conditions for every data set studied in nonparametric regression approach (Hardle, 1990). The main problem of not satisfying such conditions is due to the difficulty of monitoring the attitude of the consistency of an estimator in the boundaries of the explanatory variables. In our case, in fact this problem has occurred especially due to the binary structure of the dependent variable. Moreover we know that the nonparametric estimations are more efficient especially for small sample size which is justified by the simulation results given in Tables 1 and 2. Therefore, interpreting the results related to the small sample size is rather meaningful.

The following section illustrates the effectiveness of the simulation results with the result in practice.

APPLICATION TO A REAL DATA

A cancer data including male patients of 65 is used to examine whether the simulation results given in previous section are supported by the results of a real data set in practice. Y is a binary variable coded as follows:

$$Y_i = \begin{cases} 0, \text{ if patient } i \text{ is alive.} \\ 1, \text{ if patient } i \text{ is dead.} \end{cases}$$

Three important factors that could affect the disease are determined. These are, the age of patient (X_1), the tumor size (X_2) and the situation of metastasis (X_3). X_1 and X_2 are continuous variables whereas X_3 is a binary variable coded as:

$$X_3 = \begin{cases} 0, \text{ if the metastasis is not stated.} \\ 1, \text{ if stated.} \end{cases}$$

We have not focused on to model and interpret the parameter estimates of the data here. We have only intended to give the results indicating the model quality and their interpretations. Table 3 gives the statistics that measure the quality of the estimated model for the data via the Mean Square Error (MSE) and the CCR criteria under three different residual types. The results are presented with respect to the GAU and EPA kernel functions and KS and PR estimators, separately.
The optimal h for the NW estimator using the GAU kernel function is computed as 10.39 whereas it is 18.02 for the EPA kernel function when the KS estimator is used in the first step of the semiparametric modeling. In the case of the PR estimator, the optimal bandwidth values are 0.16 and 0.34 for the GAU and EPA kernels, respectively. It is obviously seen in Table 3 that the real data results support the considerable part of the simulation results. These consistent parts are listed as:

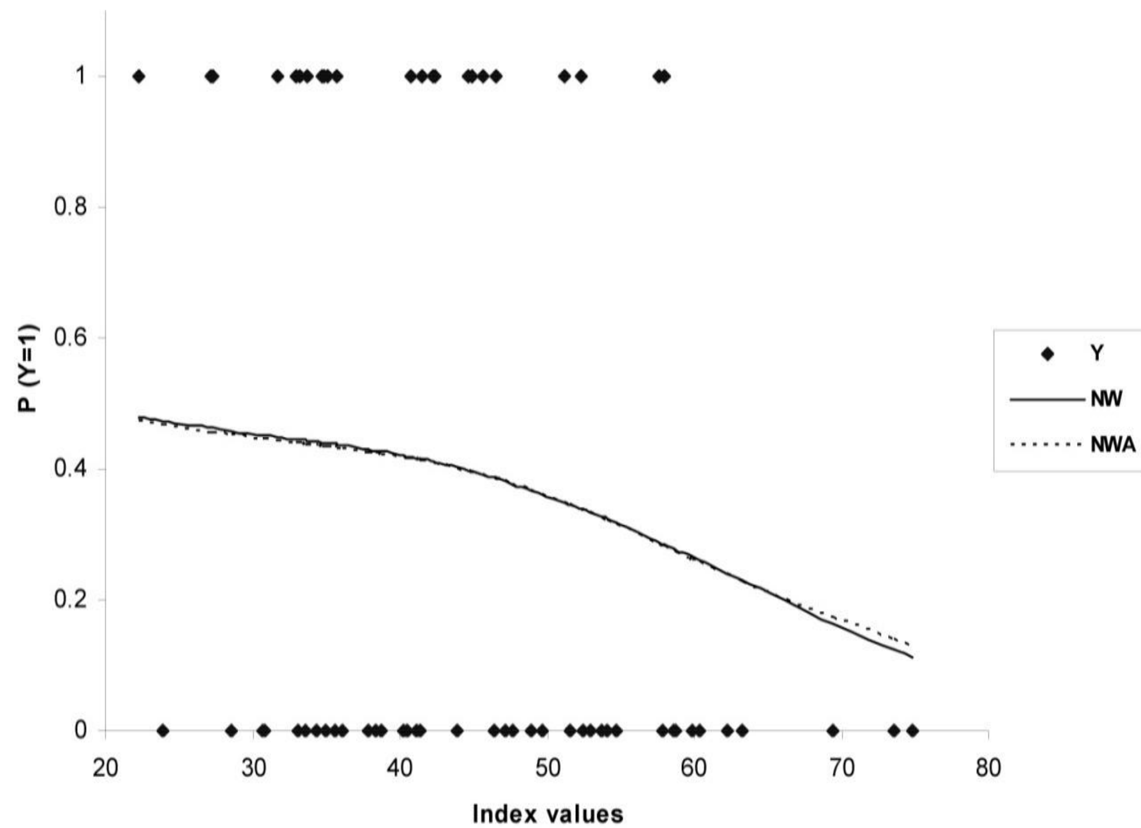

Figure 1. Predicted probabilities versus index values related to the NW and NWA estimators under the GAU kernel function and the KS estimator.

1. The deviance residuals have smaller MSE values in comparison with the Pearson residuals for both GAU and EPA kernel functions.
2. As we do not have a sufficient evidence to apply a parametric model to the data studied, we could only propose to use the semiparametric KS estimator in the estimation of β in the first step of the semiparametric estimation procedure.
3. In parallel with the simulation results, it was found that the MSE value of NW in any case is slightly smaller than NWA without taking any kernel function into consideration.

The graphical presentations of the probability of being dead: $P(Y=1)$ versus the estimated linear index values under different combinations of the NW, NWA and KS estimators and the GAU and EPA kernel functions are given in Figures 1 and 2. Both figures shows the similarity of the NW and NWA results under the KS estimator and two widely used kernel functions.

DISCUSSION

According to the simulation study, the best combination of the most efficient estimators has been determined both in the first and the second steps in the semiparametric estimation of binary response data when data sets are consistent with the parametric logit model.

Results have been assessed according to the MASE values under the deviance residuals and CCR values.

Results indicate that the semiparametric KS estimations are considerably close to the estimations of the parametric counterpart of it when the parametric model assumptions are satisfied. This obviously emphasizes the success of the semiparametric KS estimator. It cannot be concluded that NW estimator exactly outperforms than NWA or vice versa in all simulation scenarios in the second part of the semiparametric estimation. However, when the exact MASE values are taken into consideration, the use of NW estimator seems to be a little better. It was also concluded that the results are not affected by the kernel functions used.

It was revealed that both the simulation and the real data results are in favor of the usage of the KS estimator in the first step and the NW estimator in the second step. It can be concluded that NWA estimator could be used for not only modeling a continuous dependent variable but also for a binary response. However we suggest that

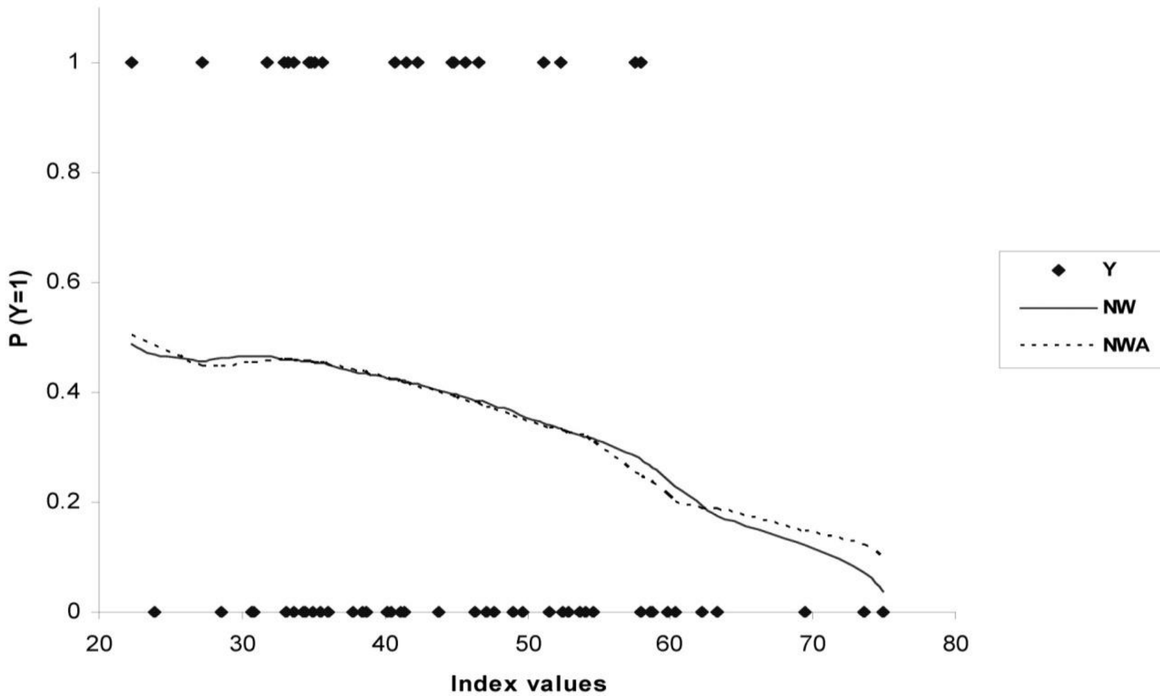

Figure 2. Predicted probabilities versus index values related to the NW and NWA estimators under the EPA kernel function and the KS estimator.

a further modification is necessarily needed for the NWA estimator to compete the opponent estimators.

REFERENCES

Abramson I (1982). On bandwidth variation in kernel estimates-a square-root law. Ann. Stat. 10:1217-1223.

Demir S, Toktamış Ö (2010). On the adaptive Nadaraya-Watson kernel regression estimators. Hacet. J. Math. Stat. 39(3):429-437.

Epanechnikov VK (1969). Non-parametric estimation of a multivariate probability density. Theor. Probab. Appl. 14:153-158.

Hardle W, Müller M, Sperlich S, Werwatz A (2004). Nonparametric and semiparametric models. Springer&Verlag, New York.

Hardle W (1990). Applied Nonparametric Regression. Cambridge University Press, Cambridge.

Horowitz JL (1998). Semiparametric Methods in Econometrics. Springer&Verlag, New York.

Horowitz JL, Hardle W (1994). Testing a parametric model against a semiparametric alternative. Economet Theor. 10:821-848.

Ichimura H (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. J. Econom. 58:71-120.

Klein W, Spady RH (1993). An efficient semiparametric estimator for binary response models. Econometrica 61:387-421.

Manski CF (1988). Identification of binary response models. J. Am. Stat. Assoc. 83: 729-738.

McCullagh P, Nelder JA (1989). Generalized Linear Models: Monographs on Statistics and Applied Probability 37, Chapman and Hall, London.

Nadaraya EA (1964). On estimating regression, Theor. Probab. Appl. 10:186-190.

Pagan A, Ullah A (1999). Nonparametric Econometrics. Cambridge University Press, Cambridge.

Powell JL, Stock JH, Stoker TM (1989). Semiparametric estimation of index coefficients. Econometrica 57(6):1403-1430.

Proença I, Santos Silva JMC (2000). Parametric and semiparametric specification tests for binary choice models: A comparative simulation study. CEMAPRE/ISEG, Working, p. 7/00.

Sain SR (1994). Adaptive kernel density estimation. Ph.D. diss. Rice University.

Silverman BW (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Watson GS (1964). Smooth regression analysis. Sankhya 26(A):359-372.