Variable Selection Semiparametric Partial Quantile Regression Model with Missing Data

Aws Adnan Al-Tai **Dr. Qutaiba N. Nayef Al-Kazaz**

University of Baghdad – College of Administration and Economics

Department of Statistics

Abstract:-

In this study, we made a comparison between the LASSO and SCAD methods, which are among the distinguished methods of dealing with models in partial quantile regression. The golden ratio method was used to estimate the smoothing parameter. (Nadarya & Watson Kernel) was used to estimate the nonparametric portion, in addition to using the generalized cross-validation method to estimate a penalty parameter. The above methods proved their efficiency in estimating the regression parameters, but the LASSO method according to criterion the mean squared error (MSE) was the best after estimating the missing data by the nearest neighbor method.

Keywords: Quantile regression, partial linear model, variable selection, LASSO, SCAD, missing data, nearest neighbor.

1. Introduction

The term semi-parametric regression is one of the widespread terms that combines the parametric regression model and the non-parametric regression. This term was used by Finnas & Hoen (1981) in the demographic field[12], Quantile regression was first proposed by Koneker and Bassett (1978)[8], Fan and Li (2001) suggested the oracle properties of SCAD in the aspect of variable selection (Smoothly Clipped Absolute Deviation)[6], The Lasso method, which was proposed by Tibshirani (1996), is one of the most famous penal methods used in estimating and selecting a linear regression model simultaneously[11].

The main objective of regression analysis is to reduce the observed data or summarize it to ensure its presentation.

For the relationship between each of the explanatory and response variables, and to analyze the regression line, give a conceptual, an approximation to that relationship by drawing or displaying that relationship according to the direction of the approximation

line, As the semi-parametric regression model is a model that combines parametric regression and nonparametric regression, one of the most famous semi-parametric models is the partial linear regression model and is symbolized by the symbol (PLM), In addition, it gives an easier explanation for the effect of each variable compared to a complete non-parametric regression, as well as better than the non-parametric model because it avoids the curse of dimensional problem that occurs when the number of variables is increased in the non-parametric model, Quantile Regression is one of the important regression methods (techniques) that have the ability to investigate the relationship between the response variable and the explanatory variables and in the entire conditional distribution of the response variable, by estimating the conditional percentiles $\left(\mathrm{Q_p(Y/X)}\right), 0 < \mathrm{p} < 1$. It differs in the distribution of the response variable rather than being limited to estimating the conditional expectation (E(y/x)) as in the normal mean regression, It is the percentile model of the conditional distribution of the response variable, which is expressed as a function in the observations of the explanatory variables. $\mathrm{Q_p(Y/X)} = \mathrm{X_i}\beta_\tau,\ 0 < \tau < 1$

2. Partial Quantile Linear Regression Model (PQLRM) [14]

Quantile regression was first suggested by (Koneker and Bassett 1978). It is considered as one of the important regression techniques that have the ability to investigate the relationship between the response variable and the explanatory variables and in the full conditional distribution of the response variable, by estimating the conditional function $\left(Q_p(Y/X)\right),\ 0 < p < 1$ the different in the distribution of the response variable rather than being limited to estimating the conditional expectation $\left(\mathrm{E}(y/x)\right)$ as in the normal mean regression, And since the partial quantile linear regression model is written according to the following formula:

$$\boldsymbol{Y_i} = \boldsymbol{X_i^T \beta_{\tau,i}} + \boldsymbol{g(T_i)} + \boldsymbol{\varepsilon_i}\ , \qquad \mathrm{i} = 1, \dots, \mathrm{n} \qquad ..(1)$$

Where X and T are explanatory variables

Y: The vector of the response variable of the degree (n× 1)

g(.): Unknown smooth function of degree (n×1)

β: Unknown parameter vector of degree (p×1)

ε: is the error term

In view of the difficulty of conducting the estimation process due to the absence of a clear behavior of the non-parametric part of the model, one of the smoothing methods must be used to get rid of the noise in the non-parametric part to show the true behavior of it, In this research, (N.W- Kernel Smoothing) will be used to later build the model correctly, and in order to perform the estimation process, we will use Penalty Methods, such as the (LASSO & SCAD) method. Which possesses the characteristics that qualify it to conduct the estimation process accurately, as the penalties make the necessary approximation to the model variables so that the estimation process expands, taking into account all the model variables and with the least possible bias.

3. Variable selection [13]

Variable selection plays an important role in the model building process. In the first stage of the process of building the statistical model in the case of high-dimensional data from a practical point of view, in the presence of a large number of explanatory (predictive) variables that are important and unimportant and have a strong or weak impact on the response variable (y). Accordingly, within the framework of regulation, and in order to overcome the aforementioned defects, many different Penalized Methods were introduced to achieve the selection of the variable and reduce the capabilities of some regression parameters and make others equal to zero with the treatment of the problem of multiconularity.

The following penal methods will be used in estimating the parameters of the semi-parametric partial quantile regression model:

3.1 LASSO method estimator [11]

The Lasso method, Least Absolute Shrinkage and Selection operator which was presented by, Tibshirani (1996), is one of the most famous penal methods used in estimating and selecting the variables of the linear regression model simultaneously. The Lasso method reduces the estimations of some regression parameters and makes others equal to zero. Thus, it estimates and selects variables in one step simultaneously. The estimator of the Lasso method is obtained by minimizing the penalized least squares function. Regression model is written as follows:

$$\hat{\beta}_{lasso} = argmin_{\beta} \left\{ \sum_{i=1}^{n} \left(yi - X\acute{i}\beta \right)^2 \right\} s.t \sum_{j=1}^{p} |\beta_j| \leq t \quad ..(2)$$

Where : $t \geq 0$ Tuning Parameter

The formula for the estimator can be expressed in the above equation in an equivalent form, as follows:

$$\hat{\beta}_{lasso} = argmin_{\beta} \left\{ \sum_{i=1}^{n} \left(yi - X\acute{i}\beta \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \quad ..(3)$$

Where: (λ) It is called the Penalty Parameter or the Regularization Parameter, $\lambda \sum_{j=1}^{p} |\beta|$ called the Penalty Function.

3.2 SCAD method estimator [6]

Fan and Li (2001) proposed the oracle properties of SCAD (Smoothly Clipped Absolute Deviation), which simultaneously estimate and select linear regression model variables.

The SCAD penalty function takes the following form:

$$p_{\lambda}(|\beta j|) = \begin{cases} \lambda|\beta j| & \text{if } 0 \leq |\beta j| < \lambda \\ \dfrac{(a^2-1)\lambda^2 - (|\beta j| - a\lambda)^2}{2(a-1)} & \text{if } \lambda \leq |\beta j| < a\lambda \\ \dfrac{(a+1)\lambda^2}{2} & \text{if } |\beta j| < a\lambda \end{cases} \quad ..(4)$$

Where: $a > 2, \lambda \geq 0$ The two parameters represent the tuning. The researchers (2001) Fan and Li suggested the value of the tuning parameter a = 3.7, because it gives the best performance in choosing the variable.

And through the derivative of the equation above with respect to β and equating the first derivative to zero, the estimator is obtained as follows:

$$\hat{\beta}_{scad} = argmin_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{n} (Yi - X\acute{\imath}\beta)^2 + n \sum_{j=1}^{p} p_{\lambda}(|\beta_j|) \right\} \quad ..(5)$$

4. penalty parameter

It is the parameter that controls the amount of reduction of the parameters and the selection of the subset of the variables included in the final model. If the value of (λ = 0), then this means that there is no penalty on the transactions, and thus the estimates of penal least squares are identical to the estimates of the ordinary least squares OLS.

And if the value of the penalty parameter is (λ→∞), then this means that the penalty is infinitely large, and then all the coefficients are forced to be zero.

4.1 Generalized Cross Validation [5]

The Cross Validation function is one of the most common functions in estimating the penalty parameter (λ), and its regular form can be defined as follows:

$$V_{(\lambda)} = \frac{1}{n} \sum_{k=1}^{n} \left(g_{n,\lambda}^{[k]}(t_k) - y_k \right)^2 \quad ..(6)$$

In order to obtain a generalized formula, the researchers suggested Graven & Wahba 1979 Minimize the function $g_{n,\lambda}^{[k]}(t_k)$ where this function can be estimated by replacing its data with points y_k whose data is known and by reducing we get an estimate of the function $g_{n,\lambda}^{[k]}$ and after solving this function we get:

$$V_{(\lambda)} = \frac{1}{n} \sum_{k=1}^{n} \left\{ \left(\sum_{j=1}^{n} a_{kj} y_j - y_j \right)^2 \Big/ (1 - a_{kk})^2 \right\} \quad ..(7)$$

Whereas:

$$\frac{\partial\, g_{n,\lambda}^{[k]}(t)}{\partial\, y_k} = a_{kk} \quad , \quad g_{n,\lambda}^{[k]}(t) = \sum_{j=1}^{n} a_{kj} y_j$$

By finding a solution to these equations through the smoothing model, we can get the final formula for the Generalized Cross Validation, which is written according to the following formula:

$$V_{(\lambda)} = \frac{1}{n} \sum_{k=1}^{n} \left(\sum_{j=1}^{n} a_{kj} y_j - y_k \right)^2 \Big/ \left(1 - \frac{1}{n} \sum_{k=1}^{n} a_{kk} \right)^2 \quad .(8)$$

5. Nadaraya &Watson kernel estimator [7]

It is named to the researchers who proposed this estimate Nadaraya &Watson 1968, According to the method of series weights, where it is used in non-parametric

regression functions, This estimator (N.W) has many characteristics, the most important of which is that it can be used in designs, whether the design is fixed or random, It also has a definite and continuous function whose integral is equal to one.

As for the kernel function used with the (N.W) estimator, it has several properties, including:

1- $\int k(v)dv = 1$
2- $\int vk(v)dv = 0$
3- $\int v^z k(v)dv = 0 \quad , \quad \forall\ z = 1,3,\dots,k-1$

And where (k) represents the degree of the kernel function, It has been confirmed in most applications that these conditions are fulfilled when (z = 2) that is, the kernel functions are of the second order, which are recognized either through derivation or integration.

The kernel estimator (N.W) can be derived in the case of a random design in the nonparametric regression functions according to the following formulas:

$$Y_i = g(T_i) + \varepsilon_i \qquad ..(9)$$

Where:

G(.) : undefined regression function (smooth function).

ε_i: Random error, $\mathrm{E}(\varepsilon_i) = 0$ ، $V(\varepsilon_I) = \sigma_\varepsilon^2$

$$\mathrm{g(t)} = \mathrm{E(Y/T = t)} = \int \mathrm{y\, f\,(y/t)dy} = \int \mathrm{y}\frac{\mathrm{f(t,y)}}{\mathrm{f(t)}}\mathrm{dy}$$

$$\hat{\mathrm{f}}(\mathrm{t,y}) = \frac{1}{\mathrm{n_{ht}h_y}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{K}\left(\frac{\mathrm{t_i - t}}{\mathrm{h_t}}\right)\mathrm{K}\left(\frac{\mathrm{y_i - y}}{\mathrm{h_y}}\right)$$

$$\frac{1}{\mathrm{n_{ht}}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{K_{ht}(t_i - t)\ K_{hy}(y_i - y)}$$

as well

$$\int \mathrm{y}\,\hat{\mathrm{f}}(\mathrm{t,y})\mathrm{dy} = \frac{1}{\mathrm{n_{ht}}}\int \mathrm{y}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{K_{ht}(t_i - t)K_{hy}(y_i - y)}$$

As

$$\int \mathrm{y\, K_{hy}(y_i - y)dy} = \mathrm{y_i}$$

$$\therefore \int \mathrm{y}\,\hat{\mathrm{f}}(\mathrm{t,y})\mathrm{dy} = \frac{1}{n_{ht}}\sum_{i=1}^{n} \mathrm{K_{ht}(t_i - t)y_i}$$

In addition, the denominator estimate can be found as follows:

$$\int \hat{\mathrm{f}}\,(\mathrm{t,y})\mathrm{dy} = \frac{1}{n_{ht}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{K_{ht}(t_i - t)}\int y\mathrm{K_{hy}}\,(\mathrm{y_i - y})\mathrm{dy}$$

$$= \frac{1}{n_{ht}} \sum_{i=1}^{n} K_{ht}(t_i - t)\ yi$$

Thus, the estimator (N.W) is as follows:

$$\hat{g}_{N.W}(t) = \frac{\sum_{i=1}^{n} K_{ht}(t_i - t)y_i}{\sum_{i=1}^{n} K_{ht}(t_i - t)} \quad ..(10)$$

The kernel estimator ($\hat{g}_{n.w}(t)$) can be written using the weights function, which is equal to:

$$W_{ht}(t, T_i) = \frac{K_{ht}(t_i - t)}{\sum_{i=1}^{n} K_{ht}(t_i - t)} \quad ..(11)$$

It can be written as follows:

$$\hat{g}_{N.W}(t) = \sum_{i=1}^{n} W_{ht}(t, T_i)y_i \quad ..(12)$$

and whereas:

$$\sum_{i=1}^{n} W_{ht}(t, T_i) = 1$$

The work of the (N.W) estimator on quantile semi-parametric regression functions can be summarized in the following equation:

$$g_n(t, \beta_\tau) = \sum_{i=1}^{n} W_{ni}(t)\left(Y_i - X_i^T \beta_\tau\right) \quad ..(13)$$

Where ($\{W_{ni}(t)\}_{i=1}^{n}$) denotes a series of weights and these weight functions can be normal if the following condition is met: $\sum_{i=1}^{n} W_{ni}(t) = 1$ Where: $W_{ni} > 0$

The weight function can be defined as:

$$W_{ni}(t) = \frac{K\left(\frac{t_i - t}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{t_i - t}{h_n}\right)} \quad ..(14)$$

Choosing the Smoothing Parameter

It is also called the bandwidth parameter and is denoted by the symbol (h), The choice of the kernel function and the smoothing parameter are necessary in estimating the regression function. The introductory parameter is an essential part in estimating the curve for non-parametric and semi-parametric regression and approximating that function to the original function and trying to obtain a balance between both the bias square and the variance so that the error is as small as possible.

6.1 Golden Ratio Method [2]

This method was suggested by the researchers (Al-Kazaz & Aldahham), where this method is based on two basic ideas:

The first basic idea showed that every two numerical values achieve the golden ratio if the ratio between the sum of these two numbers and the largest of them is equal to the ratio between the largest and the smallest of the two numbers. As the amount of the golden ratio is represented by the number (1.618033989).

The first method: The first method is calculated through the following equation:

$$\varphi = \frac{1+\sqrt{5}}{2} \quad ..(15)$$

The second method: The second method is calculated through the following equation:

$$\varphi = \cos(36^{\circ})$$

The second basic idea was inspired by the the book of allah Qur'an from Surat Al Imran in verse (96) about the location of the Kaaba on Earth, as the location of the Kaaba on Earth represents the golden ratio from north to south and from east to west.

Therefore, when dividing the value of the parameter (h) by the golden ratio, we get the optimal (h), as in the following equation:

$$\hat{h}_{GR} = \frac{\hat{\mathrm{h}}}{1.618} \qquad ..(16)$$

7. Missing Data [9]

In many cases, especially in the applied field, the information to be studied may not be available, and this may be due to many reasons, including the loss of that information for insufficient (unknown) reasons. It may occur as a result of the large amount of that value to be studied, Therefore, in practical practices, variables are usually exposed to loss, whether they are explanatory variables or the response variable, and the loss of data may lead to the bias of these data and this affects the quality of the data and knowledge of its performance and study. In general, the methods of dealing with lost data can be classified into two types:

1. **Case Deletion** : Without dealing with missing data
2. **Missing Data Imputation**: Where this method replaces the missing values in a data set with other possible values, and it has several benefits, including that the treatment of these missing data does not always depend on a specific method, and this allows researchers to choose a calculation method that is more appropriate with their applications.

Since each process of loss has a specific pattern (mechanism) according to which the loss takes place according to a certain probability, and as a result, the researchers (Little and Rubin) classified the mechanisms of loss into three types:

1. **Missing At Random**: If the cause of the loss is related to the values of other variables only and is independent of the missing value, then in this case the loss is random (MAR).
2. **Missing Completely At Random**: If the cause of the loss is independent of the missing value itself and the values of other variables in the sample, then the data is lost randomly (MCAR)
3. **Not missing at random**: If the cause of the loss is caused by the missing value itself, that is, the data is lost purposely, not randomly(Not MAR).

8. Partial Linear Quantile Regression Model in the case of complete and incomplete data [10]

In the partial linear quantitative regression model in Equation (2.1), and specifically if there are missing variables in the response variable (Yi), in a sample of size (n), since the two variables (T,X) have completely observed observations then An (index variable) can be set to express whether or not there is a data loss.

$$\delta_{i=}\begin{cases}0 & \text{if Yi is missing}\\ 1 & \text{otherwise}\end{cases}$$

And assuming that (Y) is missing at random
Where and under the condition of independence $P(\delta = 1 \mid Y, X, T) = P(\delta = 1 \mid X, T)$, Where the loss mechanism (MAR) receives the most attention in theory and practice because it describes the normal process condition.
And assuming that:

$$r = \sum_{i=1}^{n} \delta i \quad , \quad m = n - r$$

Where (m, r) is defined as the response group and the non-response group or the loss response variable Y, (respectively).
So, (S_r) represents the response state, and (S_m) represents the non-response state.
Assuming that (K) is a symmetric probability density function and ((h =h_n) the bandwidth that is decreasing towards zero with increasing sample size (n→ ∞).
whereas :

$$Y_i - X_i^T \beta = g(T_i) + \varepsilon_i \qquad i = 1,2, \dots, r \qquad \dots(17)$$

Assuming that (β) values are defined, we have a kernel estimator $\hat{g}(t)$ for $g(t)$, based on the complete observations data:

$$\hat{g}(T_i) = \frac{\sum_{j=1}^{n} \delta_j K\left(\frac{t_i - T_j}{h}\right)(Y_j - X_j \beta)}{\sum_{j=1}^{n} \delta_j K\left(\frac{t_i - T_j}{h}\right) + n^{-2}} \qquad .. (18)$$

Since the n^{-2} component is added to avoid the case that the denominator is zero: where K(.) is called the kernel function, which can be obtained from using the (Gaussian kernel) the standard normal density function and using $\hat{g}(t)$ instead of g(t) in equation (18) we get:

$$Y_i - X_i^T \beta \approx \frac{\sum_{j=1}^{n} \delta_j K\left(\frac{t_i - T_j}{h}\right)(Y_j - X_j \beta)}{\sum_{j=1}^{n} \delta_j K\left(\frac{t_i - T_j}{h}\right) + n^{-2}} \qquad , i \in r$$

Using the transformations, we get:

$$Z_i \approx U_i^T \beta \ , \qquad i \in S_r$$

whereas:

$$Z_i = Y_i - \frac{\sum_{j=1}^{n} \delta_j Y_j K\left(\frac{(T_i - T_j)}{h}\right)}{\sum_{j=1}^{n} \delta_j K\left(\frac{(T_i - T_j)}{h}\right) + n^{-2}}$$

$$U_i = X_i - \frac{\sum_{j=1}^{n} \delta_j X_j K\left(\frac{(T_i - T_j)}{h}\right)}{\sum_{j=1}^{n} \delta_j K\left(\frac{(T_i - T_j)}{h}\right) + n^{-2}} \qquad , i \in S_r$$

According to the theory of the penalty linear quantile regression model, the parameters of β can be estimated as follows:

- In the case of adding the penalty limit for the estimator of the (LASSO) method:

$$\hat{\beta}_{lasso,n} = \left(\sum_{i=1}^{n} \delta_i U_i U_i^T\right)^{-1} \left(\sum_{i=1}^{n} \delta_i U_i Z_i\right) + \lambda \sum_{j=1}^{p} |\beta_{\tau,j}|$$

- In the case of adding the penalty limit for the SCAD method estimator:

$$\hat{\beta}_{scad,n} = \left(\sum_{i=1}^{n} \delta_i U_i U_i^T\right)^{-1} \left(\sum_{i=1}^{n} \delta_i U_i Z_i\right) + \sum_{j=1}^{p} P_\lambda\left(|\beta_{\tau,j}|\right)$$

And by substituting into the equation (18) we get:

$$\hat{g}_n(t_i) = \frac{\sum_{j=1}^{n} \delta_j K\left(\frac{t_i - T_j}{h}\right)\left(Y_j - X_j \hat{\beta}_n\right)}{\sum_{j=1}^{n} \delta_j K\left(\frac{t_i - T_j}{h}\right) + n^{-2}} \qquad ..(19)$$

9. Method of nearest neighbor in the response variable Y [1,3]

The nearest-adjacent substitution method is one of the modern methods of finding the missing values, and for the purpose of knowing how this method works:

Suppose we have the response variable (y_i) which has missing some of its observations (Missing Observation), and the explanatory variable (x_i) which is complete Observation where:

$$\underbrace{y_1, y_2, \dots, y_{n-m}}_{\text{obs.}}, \underbrace{y_{n-m+1}, y_{n-m+2}, \dots, y_n}_{\text{missing}}$$

$$\underbrace{X_1, X_2, \dots, X_{n-m}, X_{n-m+1}, X_{n-m+2}, \dots, X_n}_{\text{obs.}}$$

$$i = 1,2, \dots, n \quad \text{and} \quad j = 1,2, \dots, n-m$$

$$1 = n-m+1, n-m+2, \dots, n$$

Where (m) represents the number of missing values in the response variable (y_i) and (n) represents the number of total observations. On this basis, the value of (y_j) is chosen and that $(1 \le j \le n-m)$ that corresponds to the lowest absolute difference between X_j, X_i and as shown With the following equation:

$$|X_j - X_i| = min_{1 \le j \le n-m} |X_j - X_i|$$

If there is no single value, the value of the average data is compensated.

After obtaining the complete sample, the semi-parametric model is estimated as follows:

$$Y_i^{NN} = X_i' \hat{\beta}_{n,\tau} + \hat{g}_n(T_i) + \varepsilon_i \qquad ..(20)$$

Simulation [4]

Simulations were performed using (1000) replicate, three sample sizes with For each experience (n= 30, 50, 100), and as follows:

1- The explanatory variables of the parameter part (Xi) are generated in the following form

$$X_1 = 2 \times \bar{X}_1 \times U$$
$$X_2 = 2 \times \bar{X}_2 \times U$$

whereas: $0 < U < 1$

Where the following values ( ) were used as initial mean values in the generation process

2- The explanatory, nonparametric variables (ti) are normally distributed with mean (0) and variance (1).

3- The random errors are normally distributed with mean (0) and variance σ^2, where four values of error variance (2, 5, 7.2, 10.3) were used.

4- The dependent variable is generated through the models used in simulation experiments through the use of regression functions for the explanatory variables of the parametric part and the non-parametric part with an error term added.

The following model was used:

$$g(t) = 3.2t^2 - 1$$

The kernel function used is a standard normal density function, Gaussian Kernel, as follows:

$$K(.) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$$

Where data loss is random (MAR) according to the percentage of loss (10%, 20%, 30%), that is, if the cause of loss is related to the values of other variables only and is independent of the missing value

Table No. (1) Simulation results when P=2 , G=2, $\mu = 7.9, 5.9$ $\sigma = 5.9, 4.6$

MSE $\hat{Y}$									
	NO MISS			MISS 10%		MISS 20%		MISS 30%	
tao	n	LASSO	SCAD	LASSO	SCAD	LASSO	SCAD	LASSO	SCAD
0.3	30	0.4489	0.4504	3.8714	37.2984	2.6672	78.1925	2.7786	14.7843
	50	0.4191	0.4070	0.5732	120.3015	3.0920	124.3670	7.8433	72.3345
	100	0.4279	0.4480	0.6627	157.3845	1.6667	106.5214	3.5671	34.9065

Table No. (2) Simulation results when P=2, G=2 $\mu = 7.9, 5.9$ $\sigma = 5.9, 4.6$

MSE $\hat{Y}$									
	NO MISS			MISS 10%		MISS 20%		MISS 30%	
tao	n	LASSO	SCAD	LASSO	SCAD	LASSO	SCAD	LASSO	SCAD
0.6	30	1.1088	1.0882	0.3760	79.8201	1.5218	85.6489	2.3871	16.8712
	50	0.1601	0.1606	0.2721	25.8167	3.6893	19.5215	2.5585	721.4737
	100	0.3009	0.3135	2.1491	98.8130	2.4364	84.0024	2.9252	146.1568

Table No. (3) Simulation results when P=2, G=2, $\mu = 9.3, 6.7$ $\sigma = 3.2, 2.5$

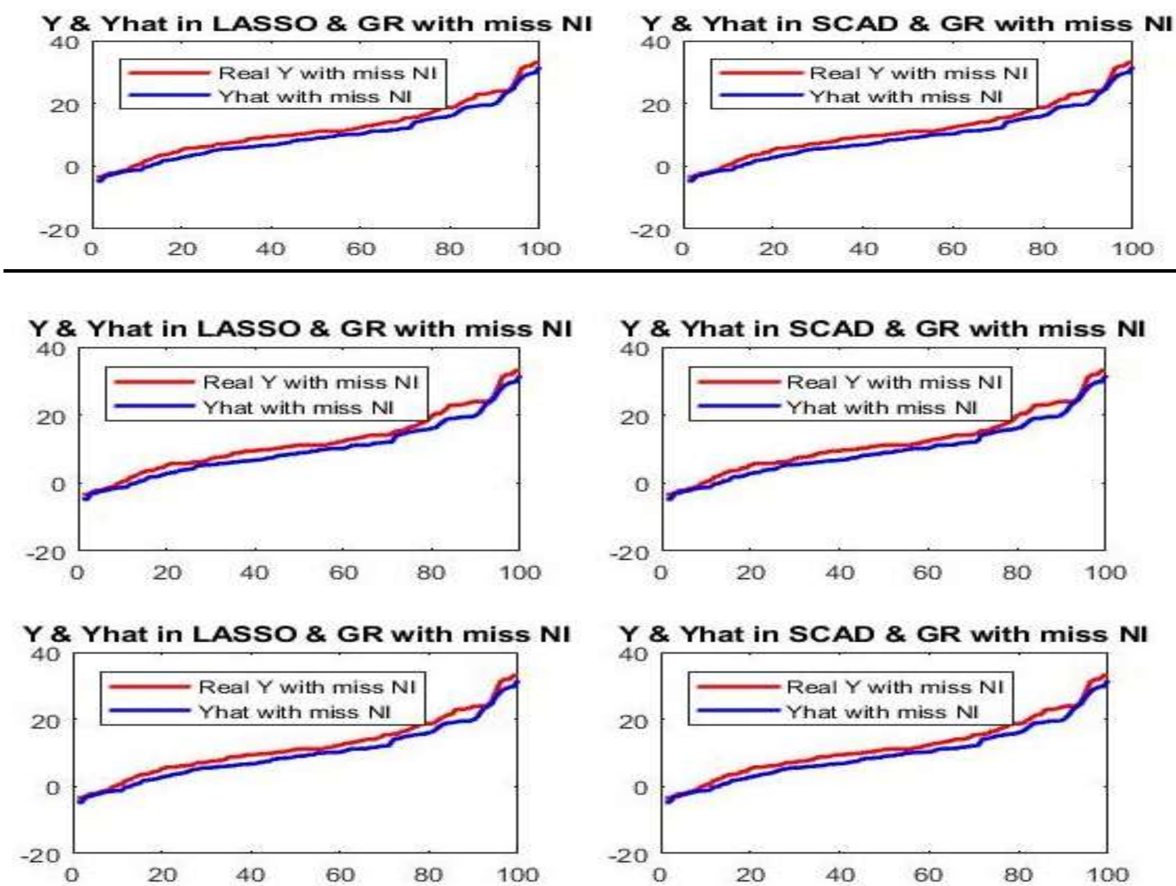
MSE $\hat{Y}$									
	NO MISS			MISS 10%		MISS 20%		MISS 30%	
tao	n	LASSO	SCAD	LASSO	SCAD	LASSO	SCAD	LASSO	SCAD
0.3	30	1.4575	1.4746	14.3059	30.8131	0.6500	42.2871	1.1201	97.1040

	50	1.0592	1.0315	1.3939	152.4229	3.0263	44.2810	2.7232	208.2536
	100	0.2086	0.2042	0.2705	80.5524	0.5435	50.8221	9.3814	183.6034

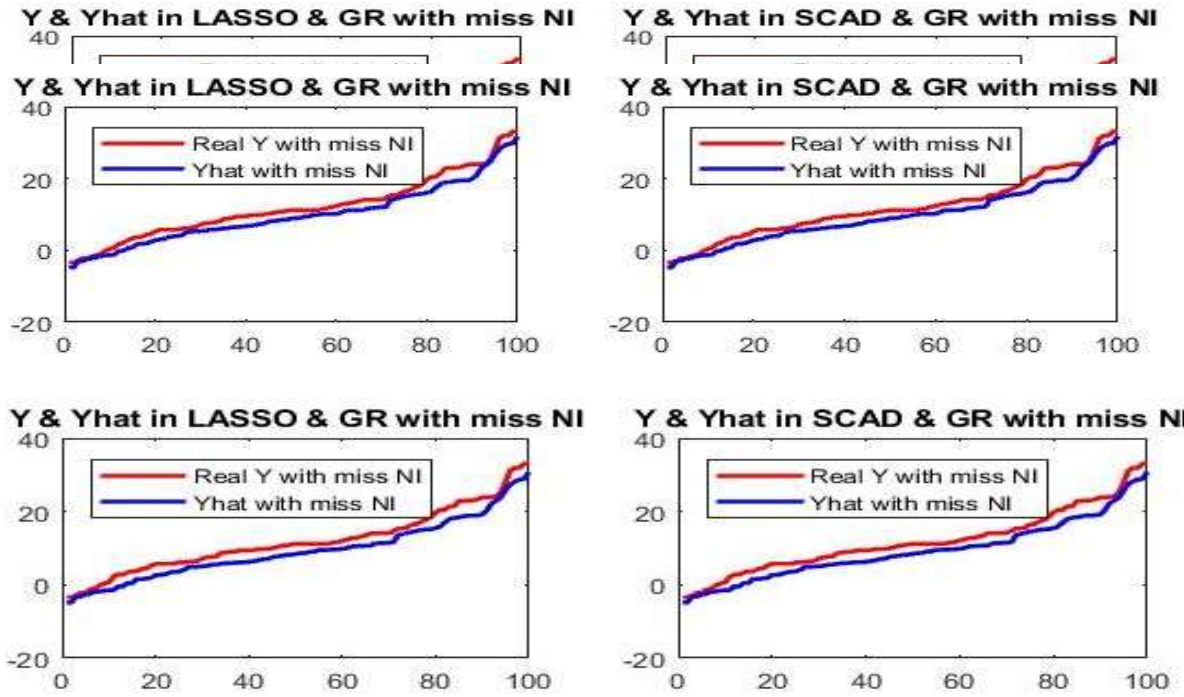
Table No. (4) Simulation results when P=2, G=2, $\mu = 9.3, 6.7$, $\sigma = 3.2, 2.5$

MSE $\hat{Y}$									
	NO MISS			MISS 10%		MISS 20%		MISS 30%	
tao	n	LASSO	SCAD	LASSO	SCAD	LASSO	SCAD	LASSO	SCAD
0.6	30	5.1807	5.0498	2.0230	33.1339	22.0560	136.0837	2.2186	30.2909
	50	0.1935	0.1943	6.5291	17.9492	10.6640	35.6970	16.3915	52.8751
	100	0.5071	0.5205	1.6033	21.0892	2.7362	51.8447	14.9335	55.6807

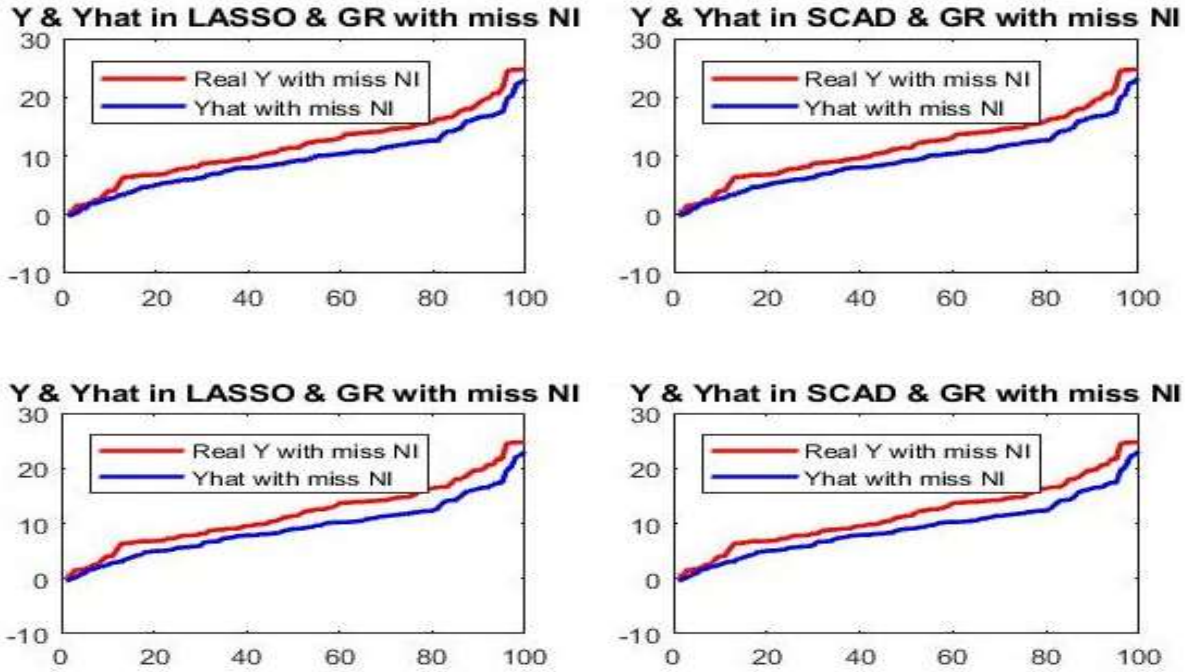
Figures when, n=100 $\tau = 0.3$ P=2 , G=2 $\sigma = 2$, miss 10%,20%&30%

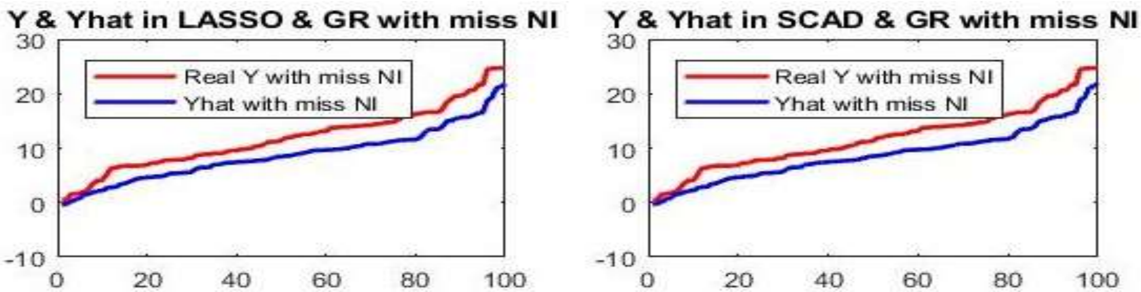


Figures when n=100 $\tau = 0.6$ P=2 , G=2, $\sigma = 5$ miss 10%,20%&30%

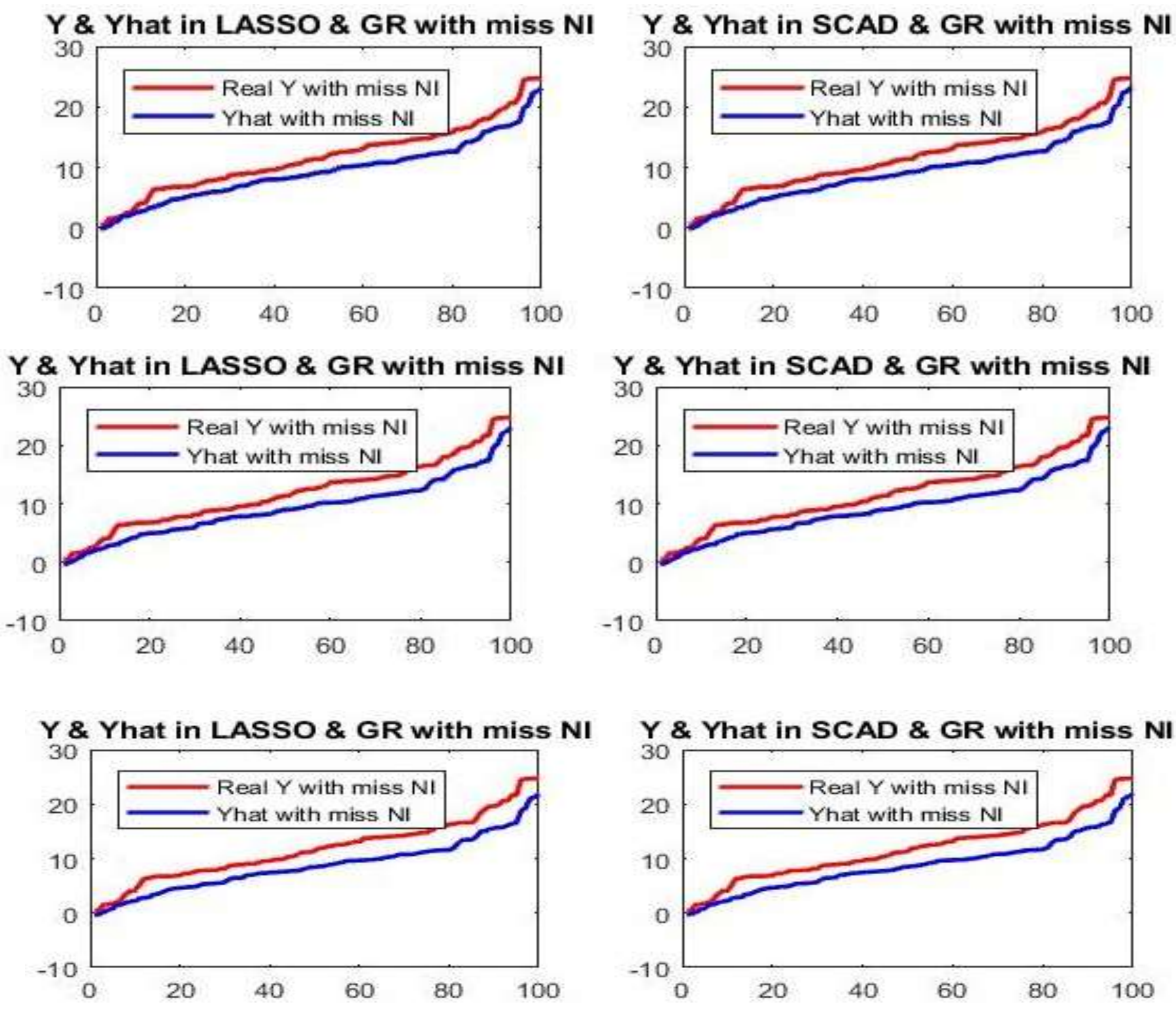


Figures when n=100 $\tau = 0.3$ P=2 , G=2, $\sigma = 7.2$, miss 10%,20%&30%





Figures when n=100 $\tau = 0.6$ P=2 , G=2, $\sigma = 10.3$, miss 10%,20%&30%



Conclusions

The estimation process using LASSO and SCAD method is relatively simple and fast and is very suitable for partial models, on other hand the add of missing to the data had

a significant impact appears in the increasing on the amount of MSE in the estimated model, where a large increase is observed on the amount of the MSE compared to the presence or absence of missing, as It is observed that the amount of MSE decreases with the increase in the sample size. We conclude from this comparison that there is a clear convergence between the estimation process by the two methods, with no missing, with a preference for the LASSO method when missing in the data occurs. A significant impact of the missing on the SCAD method appeared with an increase in the amount of the MSE.

References

1. Al-Kazaz, Q. N. & Hmood, M. Y. (2009). Comparing Some Methods For A single Imputed A missing Observation In Estimating Nonparametric Regression Function. Journal of Economics and Administrative Sciences, 15(53), 223-223
2. Al-Kazaz, Q. N. N., & Aldahham, M. Y. (2012, September). A proposal method for selecting smoothing parameter with missing values. In 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE) (pp. 1-5). IEEE.
3. Al-Kazaz, Q. N. N., (2007). "A comparison of robust Bayesian Approaches with other Methods for Estimating Parameters of Multiple Linear Regression Model with missing Data" dissertation introduced to Council of the College of Administration and Economics Baghdad University.
4. Hmood, M.Y. and Mohamed, M., 2014. A comparison of the Semiparametric estimators model using different smoothing methods. *Journal of Economic and Administrative Sciences*, *20*(75), pp.376-394.
5. Craven, P., & Wahba, G. (1978). Smoothing Noisy Data with Spline Functions. Estimating the Correct Degree of Smo... *Numerische Mathematik*, *31*, 79.
6. Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), 1348-1360.
7. Hmood, M. Y. (2000). "Comparing Kernel Estimators for estimating Regression function" thesis introduced to Council of the College of Administration and Economics Baghdad University.
8. Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. Econometrica: journal of the Econometric Society, 33-50
9. Liang, H., Wang, S., Robins, J. M., & Carroll, R. J. (2004). Estimation in partially linear models with missing covariates. Journal of the American Statistical Association, 99(466), 357-367
10. Qin, Y., Zhang, S., Zhu, X., Zhang, J., & Zhang, C. (2007). Semi-parametric optimization for missing data imputation. Applied Intelligence, 27(1), 79-88.Sherwood, B. (2016). Variable selection for additive partial linear quantile regression with missing covariates. Journal of Multivariate Analysis, 152, 206-223.
11. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.
12. Wellner, J. A., Klaassen, C. A., & Ritov, Y. A. (2006). Semiparametric models: a review of progress since BKRW (1993).

13. Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. Statistica Sinica, 801-817
14. Zhao, P. X. (2015). Quantile Regression for Partially Linear Models with Missing Responses at Random. In Applied Mechanics and Materials (Vol. 727, pp. 1013-1016). Trans Tech Publications Ltd.

المستخلص

فـي هـذه الدراسـة، قمنـا بـإجراء مقارنـة بـين طـريقتين LASSO وSCAD، وهمـا مـن الأسـاليب المميـزة للتعامـل مـع النمـاذج فـي الانحـدار التجزيئـي الجزئـي. تـم اسـتعمال طريقـة النسـبة الذهبيـة لتقـدير معلمـة التمهيـد (h). وتـم اسـتعمال (Nadarya & Watson Kernel) لتقـدير الجـزء اللامعلمـي، بالإضـافة إلـى اسـتخدام طريقـة العبـور لتقـدير معلمـة الجـزاء (λ). أثبتـت الطـرق المـذكورة أعـلاه كفاءتهـا فـي تقـدير معـاملات الانحـدار، لكـن طريقـة LASSO وفقًـا لمعيـار متوسـط مربعـات الخطـأ (MSE) كانـت الأفضـل بعـد تقـدير البيانـات المفقـودة بواسـطة طريقـة المجـاور الأقـرب (N.N.).