ESTIMATING NON-LINEAR REGRESSION PARAMETERS USING DENOISED VARIABLES

Supported by tetfund TERTIARY EDUCATION TRUST FUND

O.A. Fasoranbaku*, A. Oluwapelumi and A.O. Soyonbo
Department of Statistics, Federal University of Technology, Akure, Nigeria
*Corresponding author: oluwapelumialabi@yahoo.com

Received: June 06, 2016 **Accepted:** September 11, 2016

Abstract: The observed data from various fields are frequently characterized by measurement error and this has been a challenging problem to constructing consistent estimators of the parameters in a nonlinear regression model. In the study, simulated data under three (3) sample sizes (i.e. 32, 256 and 1024) were used, applying Epanechnikov kernel, Gaussian kernel, Wavelet and Polynomial Spline on noisy data. The study revealed the performances of denoised nonlinear estimators under different sample sizes and comparison was made using the mean squared error criterion. The result of the studies showed that the denoised nonlinear least squares estimator (DNLS) is the best under each sample size considered.

Keywords: Denoising, measurement error, Monte-Carlo, non-linear regression model, production function

Introduction

Statistical estimation can be regarded as a subfield of statistics, and lies at the core of a number of areas of science and engineering, including data mining, and signal processing. Each of these disciplines provides some information on how to model data and how best to exploit the hidden structure of the model of interest. In this work, we are interested in estimating nonlinear regression model (Nonlinear Cobb- Douglas production model). In nonlinear regression, observational data are modeled by a function which contains parameters that are not linear in nature. The data consist of independent variables (explanatory variables) and their associated observed dependent variables (response variables) which may contain measurement error or noise.

Variables are said to be noisy if they are not measured correctly because the measurement system generating are not perfect. In statistics, an error is not a mistake because variability is an inherent part of things being measured and of the measurement process. Error-in-variables (EIV) model are regression models that account for measurement errors in independent variables. Many economic data sets are contaminated by the mismeasured variables. The problem of measurement errors is one of the most fundamental problems in empirical economics. The presence of measurement errors causes biased and inconsistent parameter estimates and leads to erroneous conclusions to various degrees in economic analysis. A measurement error is called classical if it is independent of the latent true values; otherwise it is called non-classical. There have been many studies on the identification and estimation of linear, nonlinear, and even non parametric model with classical measurement errors (Cheng and Van Ness, 1999; Carroll *et al.*, 2006).

A natural approach to overcome this problem is to apply the smoothing techniques to handle the data for proper removal of the noisy observation (i.e. denoise the data). In statistics and image processing, to smoothen a data set is to create an approximating function that attempt to capture important patterns in the data, while leaving out noise or other fine scale structure or rapid phenomena. Smoothing extracts more information from the data as long as the assumption of smoothing is reasonable and provides flexible and robust analysis. There are several methods of smoothing techniques which can be used to screen out noise, such as: wavelets, developed by Donoho and Johnstone, (1994, 1995a and 1995b). Other methods are kernel, polynomial spline, etc. These appear often in applied fields such as marketing (Blattberg and Neslin, 1990), medicine and biology (Aldroubi and Unser, 1996), and image processing (Prasad and Lyengar, 1997).

Measurement errors in the independent variables of linear regression models lead to inconsistent coefficient estimates. To overcome this inconsistency problem many studies on denoisinghas been extended to least squares estimator, least absolute deviation estimator and M-estimator using kernel, wavelet and polynomial spline as smoothers. The study carried out by Cai *et al.* (2000) denoised both the dependent and explanatory variables; while Cui *et al*. (2002) suggested denoising only the explanatory variables. Furthermore, a series of papers (You and Zhou, 2007; You *et al.*, 2009; Zhou and Liang, 2009) adopted the approach of only denoising explanatory variables. Cui *et al*. (2010) denoised only the explanatory variables and showed that the denoised nonlinear least squares estimator is not robust to outliers. The study carried out by Fasoranbaku and Soyonbo (2015) showed that the denoised nonlinear least square estimator under the several smoothers (Epanechnikov, Gaussian, wavelet and polynomial spline) considered outperforms both the denoised nonlinear least absolute deviation estimator and nonlinear M-estimator. Soyombo and Fasoranbaku (2015) also used the known Epanechnikov Kernel smoother, to perform the denoising procedures, carry out simulation studies under some settings to determine the performance of the denoised non-linear estimators when the parameter values are varied. The results show that parameters of non-linear model are not sensitive and thus have no effect on the performance of denoised non-linear estimators.

For the purpose of estimating the error model, this study investigate Cobb Douglas production model in economics. The model with additive error is written as:

$$P_t = \beta_1 L_t^{\beta_2} K_t^{\beta_3} + u_t \qquad (1)$$
$$(\beta_1 > 0), \quad (0 < \beta_2 < 1), \quad (0 < \beta_3 < 1)$$

Where: P_t is output at time t (the real value of all goods produced in a period of time)

L_t is the Labour input (the number of person hours in a period of time), K_t is the Capital input (the real value of Machinery and Building), β_1 is a Constant, (total factor productivity), β_2 and β_3 are the output elasticity of Labour and Capital (measure the respective contribution of L_t and K_t to the production process) and u_t is the stochastic disturbance term

Suppose that $\{(L_t, K_t, P_t) : 1 \le t \le n\}$ are unobservable "true" variables satisfying a nonlinear relationship,

measurements of (L_t, K_t, P_t) are collected to yield an observable data set of $\{(x_{t1}, x_{t2}, y_t) : 1 \le t \le n\}$ i.e. the true variables plus additive measurement errors such that;

$$x_{t1} = L_t + \delta_t, \quad x_{t2} = K_t + \varepsilon_t \text{ and } y_t = P_t + u_t \qquad (2)$$

Where: δ_t and ε_t are measurement errors.

To be in line with the usual nonlinear model, the model (1) becomes:

$$y_t = \beta_1 x_{t1}^{\beta_2} x_{t2}^{\beta_3} + u_t \qquad (3)$$

In estimating denoised non-linear regression parameters, the effect of the sample sizes on denoised estimators has not been considered. Therefore, this study would employ three different denoised estimators (i.e denoised non-linear least square (DNLS), denoised non-linear absolute deviation (DNLAD), and denoised non-linear moment (DNM)) to estimate non-linear regression parameters under three different sample sizes (32, 256, and 1024). The goal of this study is to provide best estimator under each sample size for particular smoothers.

Materials and Methods

Denoising procedures

The basic idea behind smoothing a data set is the creation of an approximating function that attempts to capture important patterns in the data while leaving out the noise, and is also referred to as "denoising". There are various methods to help restore a data set from measurement noise. In this study, the following smoothing method are used

Kernel denoising

In this section, the denoising kernel-type smoothing procedure is considered. First, only the x variable is denoised. Therefore Kernel-type smoothing procedure for the x_i using a weight function is

$$w(x,h) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right) \qquad (4)$$

To this end, let $k(;) \ge 0$ be a symmetric kernel supported on [-1,1] with $\int_{-1}^{1} K(x)dx = 1$ for some smoothing parameter h; where x is the value of the scalar variable for which one seeks an estimate, while X_i are the values of that variable in the data. K is a function of a single variable called the *kernel*. The kernel determines the *shape* of the function. The parameter *h* is called the bandwidth or smoothing constant and n is the sample size. It is important to note that the bandwidth strongly depends on the sample size, so when sample size increases, the bandwidth tends to shrink. The bandwidth controls the degree of smoothing and adjusts the size and form of the function.

Also, $$u = \left(\frac{x - X_i}{h}\right) \qquad (5)$$

For the purpose of this study, the two most kernels are utilized

i. Epanechnikov kernel:

$$K(u) = 0.75(1-u^2)I_{(|u|\le 1)} \; on \quad u \in (-1,1) \qquad (6)$$

ii. Gaussian Kernel: $$k(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \qquad (7)$$

The problem of selecting the smoothing parameter for kernel estimation has been explored by many authors and no procedure is yet been considered the best in every situation. Automatic bandwidth selection methods can be divided into two categories: classical and plug in method.

The accuracy of kernel smoothers is a function of the kernel K and the bandwidth h; the precision depends mainly on the smoothing parameter h. One of the most frequently used methods of bandwidth selection was introduced by Silverman (1986). The choice of bandwidth is crucial and is also a challenge. There are various methods for selecting bandwidth but there is no single best method. A common choice is the "Silverman's rule of thumb" (Sheather and Jones, 1991 and Wand and Jones,1995) for an optimal bandwidth:

$$h_{opt} = 0.9[\min(s, lQR)]/1.349n^{-\frac{1}{5}} \qquad (8)$$

Where: s is the sample standard deviation and IQR is the interquartile range (0.75 quartile minus 0.25 quartile). In this case, the Gaussian and Epanechnikov Kernel were chosen and the unknown parameter h will be estimated by the optimal bandwidth h_{opt} (Walter Zucchini, 2003).

Wavelet denoising

Wavelets are functions that satisfy certain requirements. The very name wavelet comes from the requirement that they should integrate to zero, "waving" above and below the $x-axis$. The diminutive connotations of wavelet suggest the function has to be well localized. Other requirements are technical and needed mostly to insure quick and easy calculation of the direct and inverse wavelet transform.

For a wavelet denoising procedure, the discrete wavelet transform is used. The procedure consists of three steps: (1) a linear wavelet transform (2) the shrinkage denoising which gives the denoised wavelet transform (3) linear inverse wavelet transform which gives the denoised estimate of the original data. In the first step, in order to obtain the vector of wavelet coefficients w, the $T \times 1$ vector of noisy data χ is multiplied by an appropriate $T \times T$ wavelet matrix *W* (whose elements depend on a specific wavelet family (e.g Daubechies, Asymentry, etc.).

$$w = W\chi \qquad (9)$$

The vector of wavelet coefficients consists of different sub-vectors, each of length $2^j \; (j = 1,...,J)$ which represents different resolution levels of the data. For instance, a dyadic length time series with monthly sampling frequency, the first resolution level captures frequency variation with duration of 2 – 4 months. Analogously, the second resolution level captures variations of 4 – 8 months, the level 3 resolution capture variations of 8–16 months and so on, up to level J. Since the data contain measurement errors (noise), this will also be transferred to specific wavelet coefficients. Donoho and Johnson (1994, 1995a, b) proposed a soft thresholding rule in order to remove the noisy wavelet coefficients and construct noise free estimates of the original data vector. In the second stage, the following thresholding rule is applied to the data

$$\hat{w} = \begin{cases} \text{sgn}(w)(|w| - \tau), & if \quad |w \geq \tau| \\ 0, & if \quad |w| \leq \tau \end{cases}$$

Where: $$\text{sgn}(w) = \begin{cases} +1 & if \quad w > 0 \\ 0 & if \quad w = 0 \\ -1 & if \quad w < 0 \end{cases}$$

$$\tau = \hat{\sigma}_\tau \sqrt{\{2\log(n)\}}$$

$\hat{\sigma}_\tau$ is the standard deviation of the wavelet coefficients. This rule pushes all coefficients towards zero, but when their magnitude is smaller than the threshold τ, which defines the level of noise in the data, they are set to zero; the resulting wavelet coefficients $\hat{w}$ are free from noise. In the third stage, the inverse of the wavelet transform is obtained in order to obtain noise free estimates of the original data vector χ as follows: $\hat{\chi} = W^{-1}\hat{w}$ (10)

with the property,

$$WW^T = I, \quad i.e.\ W^{-1} = W^T.$$

Then (10) is, equivalent to:

$$\hat{\chi} = W^T \hat{w} \qquad (11)$$

Polynomial spline denoising

Let $(x_i, y_i); x_1 < x_2 < \ldots < x_n, i \in z$ be a sequence of observations, modelled by a relation $Y_i = u(x_i)$, the smoothing spline estimate $\hat{u}$ of the function u is defined to be the minimiser of $\sum_{i=1}^{n}[y_i - \hat{u}(x_i)]^2 + \lambda\int_{x_1}^{x_n} \hat{u}''(x)^2 dx$ where λ is a positive smoothing parameter which controls the amount of smoothing of the data and is defined between 0 and 1, $\lambda = 0$ produces least squares straight line fit to the data, while $\lambda = 1$ produces a piecewise cubic polynomial fit that passes through the data points. The smoothing parameter λ is automatically selected in the "interesting range". The interesting range of λ is often near $\left[\frac{1}{1+\frac{h^3}{6}}\right]$

Where: h is the average spacing of the data points (Hambers and Hastie, 1992). It is useful to think of fitting a smoothing spline $\hat{u}$ in two steps:

1) First, derive the values $\hat{u}(x_i)\ i = 1,\ldots,n$
2) From these values, derive $u(x)$ for all.

Now treat the second step first. Given the vector $\hat{m} = \hat{u}(x_i), \ldots, \hat{u}(x_n)$ of fitted values, the sum of squares part of the spline criterion is fixed. It remains only to minimize $\int_{x_1}^{x_n} \hat{u}''(x)^2 dx$, and the minimizer is a natural cubic spline that interpolates the points $(x_i, \hat{u}(x_i))$ This interpolating spline is a linear operator, and can be written in the form $\hat{u}(x) = \sum_i^n \hat{u}(x_i) f_i(x)$, where $f_i(x)$ are a set of spline basis functions. As a result, the roughness penalty has the form: $\int_{x_1}^{x_n} \hat{u}''(x)^2 dx = \hat{m}^T A \hat{m}$, where the elements of A are $\int_{x_1}^{x_n} f_i''(x) f_j''(x) dx$ The basis functions, and hence the matrix A depend on the configuration of the predictor variables x_i,but not the responses Y_i or $\hat{m}$.

Now back to the first step. The penalized sum of squares can be written as;

$$\|Y - \hat{m}\|^2 + \lambda \hat{m}^T A \hat{m}, where Y = (Y_1, \ldots, Y_n)^T$$

Minimizing over $\hat{m}$ gives $\hat{m} = (I + \lambda A)^{-1} Y.$

Linearization of non-linear function

The algorithm is based on Newton Raphson method of approximation. Let us consider (1), a nonlinear Cobb Douglas production model:

$$P_t = \beta_1 L_t^{\beta_2} K_t^{\beta_3} + u_t \qquad (12)$$

Let $f(L, K, \beta_1, \beta_2, \beta_3)$ represent the function, then the nonlinear Cobb Douglas production function becomes:

$$P_t = f(L_t, K_t, \beta_1, \beta_2, \beta_3) + u_t \qquad (13)$$

In Newton-Raphson method we find the values of β_j that maximize a twice differentiable con- cave function, the objective function $g(\beta)$. In this method, we approximate $g(\beta)$ at β^t by Taylor series expansion up to quadratic terms

$$g(\beta) \approx g(\beta^t) + G(\beta^t)(\beta - \beta^t) + \frac{1}{2}(\beta - \beta^t)' H(\beta^t)(\beta - \beta^t)$$

Where: $G(\beta^t) = \left[\frac{\partial g}{\partial \beta_i}\right]_{\beta^t}$ is the score vector and

$H(\beta^t) = \left[\frac{\partial^2 g}{\partial \beta_i \partial \beta_k}\right]_{\beta^t}$ is the Hessian matrix.

The value of P_t, L_t, K_t have been observed and we must estimate $(\beta_1, \beta_2, \beta_3)$, therefore equation 3.2 is re-write as, $P_t = (\beta_1, \beta_2, \beta_3) + u_t$ (14)

$$G(\beta\ ) = \left[\frac{\partial S(\beta)}{\partial \beta_1}, \frac{\partial S(\beta)}{\partial \beta_2}, \frac{\partial S(\beta)}{\partial \beta_3}\right]'$$

$$H(\beta\ ) = \begin{bmatrix} \frac{\partial^2 S(\beta)}{\partial \beta^2{}_1}, \frac{\partial^2 S(\beta)}{\partial \beta_1 \partial \beta_2}, \frac{\partial^2 S(\beta)}{\partial \beta_1 \partial \beta_3} \\ \frac{\partial^2 S(\beta)}{\partial \beta_1 \partial \beta_2}, \frac{\partial^2 S(\beta)}{\partial \beta_2{}^2}, \frac{\partial^2 S(\beta)}{\partial \beta_2 \partial \beta_3} \\ \frac{\partial^2 S(\beta)}{\partial \beta_1 \partial \beta_3}, \frac{\partial^2 S(\beta)}{\partial \beta_2 \partial \beta_3}, \frac{\partial^2 S(\beta)}{\partial \beta_3{}^2} \end{bmatrix}$$

This Hessian matrix is positive definite, the maximum of the approximation of $g(\beta)$ occurs when its derivative is zero

$$G(\beta^t) + H(\beta^t)(\beta - \beta^t) = 0 \qquad (15)$$

$$\beta = \beta^t - \left[H(\beta^t)\right]^{-1} G(\beta^t) \qquad (16)$$

This gives a way to compute β^{t+1}, the next value in iterations, and is defined as;

$$\beta^{t+1} = \beta^t - \left[H(\beta^t)\right]^{-1} G(\beta^t) \qquad (17)$$

The iteration procedures continue until convergence is achieved. Near the maximum the rate of convergence is quadratic as defined by $\left|\beta^{t+1} - \hat{\beta}_t\right| \le c\left|\beta^t - \hat{\beta}_t\right|^2$ for some $c \ge 0$ when $\beta_i^{\,t}$ is near $\hat{\beta}_t$ for all i. Thus we get estimates $\beta_i^{\,t}$ by Newton Raphson methods.

From the linearization result in equation (16) we can obtain estimate of $\beta_1, \beta_2, \beta_3$ as follow:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \beta_1^0 \\ \beta_2^0 \\ \beta_3^0 \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 S(\beta)}{\partial \beta^2{}_1}, \frac{\partial^2 S(\beta)}{\partial \beta_1 \partial \beta_2}, \frac{\partial^2 S(\beta)}{\partial \beta_1 \partial \beta_3} \\ \frac{\partial^2 S(\beta)}{\partial \beta_1 \partial \beta_2}, \frac{\partial^2 S(\beta)}{\partial \beta_2{}^2}, \frac{\partial^2 S(\beta)}{\partial \beta_2 \partial \beta_3} \\ \frac{\partial^2 S(\beta)}{\partial \beta_1 \partial \beta_3}, \frac{\partial^2 S(\beta)}{\partial \beta_2 \partial \beta_3}, \frac{\partial^2 S(\beta)}{\partial \beta_3{}^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial S(\beta)}{\partial \beta_1} \\ \frac{\partial S(\beta)}{\partial \beta_2} \\ \frac{\partial S(\beta)}{\partial \beta_3} \end{bmatrix}$$

Once a parameter vector is obtained, the estimates are likely going to be better than the old trial estimates, and so can be used in place of $(\beta_1^0, \beta_2^0, \beta_3^0)$ which known as initial parameters and the computation can be done again. The iteration can continue obtaining new and better estimates until the difference between successive parameter vectors is small enough to assume convergence.

Denoised non-linear regression estimators

When the regressors in a non-linear regression model are subject to measurement errors, it becomes a problem to construct consistent estimators of the parameters. It is possible, however, to construct consistent estimators in a non-linear model like (1) by first applying the denoising techniques to the variables, then estimators like the least squares, least absolute deviation and M-estimator will be applied to these denoised variables to yield consistent estimators which are called

Denoised non-linear least squares (DNLS)

Now consider the non-linear regression model

$$P_t = f(L_t, K_t, \beta_1, \beta_2, \beta_3) + u_t \quad t = 1, \ldots, n \qquad (18)$$

Where: $f(., \beta_1, \beta_2, \beta_3)$ is a non-linear function. Since the true values of L_t, K_t are unknown, the denoised variables x_{t1}^*, x_{t2}^* will replace x_{t1}, x_{t2} to get a denoised non-linear least squares (DNLS) estimator. The random errors u_t are assumed to be uncorrelated and have mean 0 and unknown variance σ^2. The denoised least squares estimator of $(\beta_1, \beta_2, \beta_3)$ minimizes;

$$D_n = \sum_{t=1}^{n} [y_t - f(x_{t1}^*, x_{t2}^*, \beta_i)]^2 \quad i = 1, 2, 3 \qquad (19)$$

Denoised least absolute deviation

Least squares can be severely distorted by outlying observations which has led to the robust estimators that are unaffected by outlying observations. The least absolute deviation or L_1 method is a widely known alternative to the classical least squares or L_2 method for statistical analysis of linear regression models. Instead of minimizing the sum of squared errors, it minimizes the sum of absolute values of errors.

Still considering the non-linear regression model, since the true values of L_t, K_t are unknown, the denoised variables x_{t1}^*, x_{t2}^* will replace x_{t1}, x_{t2} to get a denoised non-linear least absolute deviation estimator. The denoised least absolute deviation estimator of $(\beta_1, \beta_2, \beta_3)$ is

$$L_n = \arg\min_{\beta_i} \sum_{t=1}^{n} \left| y_t - f(x_{t1}^*, x_{t2}^*, \beta_i) \right| \qquad (20)$$

Where: β_i is the solution of the parameters.

Denoised M-estimator

The least squares estimators of regression are known to be sensitive to outliers in the data. Robust estimators can be more efficient when the error distributions are non-Gaussian and can reduce errors in the data. M-estimators are a broad class of estimators, which are obtained as the minima of sums of functions of the data. M-estimators are arguably the most popular robust methods. To be more specific, M-estimator M_n is considered as

$$M_n = \arg\min_{\beta_i} \sum_{t=1}^{n} \rho\left[y_t - f(x_{t1}^*, x_{t2}^*, \beta)\right] \qquad (21)$$

Where: ρ is a loss function. The function ρ can be chosen in such a way to provide desirable properties of estimators (in terms of bias and efficiency) when the data are truly from the assumed distribution. Least-squares estimators are special M-estimators with $\rho(x) = x^2$, where .

$$x = \left[P_t - f(x_{t1}^*, x_{t2}^*, \beta_i)\right] \qquad (22)$$

Simulation studies

A Monte Carlo simulation is a problem solving techniques used to approximate the probability of certain outcomes by running multiple trials, using random variables.

In this work, an extensive Monte Carlo simulations is conducted to generate random data of sample sizes 32, 256

and 1024 to examine the performance of the denoised nonlinear estimators from the model

$$y_t = P_t + u_t \quad and \quad x_{t1} = L_t + \delta_t, \quad x_{t2} = K_t + \varepsilon_t \quad (23)$$

Where: $L_t \sim U(1,30)$, $K_t \sim U(10,200)$, $u_t \sim N(0,0.25)$, $\delta_t \sim N(0,0.16)$, $\varepsilon_t \sim N(0,0.16)$, $y_t = \beta_1 x_{t1}^{\beta_2} x_{t2}^{\beta_3} + u_t$ with standard parameter values $(\beta_1 = 1.01, \quad \beta_2 = 0.75, \quad \beta_3 = 0,25)$, which were derived from the theory of production by Charles Cobb and Paul Douglass with the following assumption: $\beta_1 > 0, \quad 0 < \beta_2 < 1, \quad 0 < \beta_3 < 1.$

Four (4) different smoothers (i.e Epanechnikov Kernel, Gaussian Kernel, Wavelet and polynomial spline are used to denoise the explanatory variables (x_{t1} and x_{t2}) under the three (3) different sample sizes (i.e 32, 256 and 1024) and the new explanatory variables become x_{t1}^* and x_{t2}^*. The regression model of the denoised data is fitted as:

$$y_t = \beta_1 x_{t1}^{*\beta_2} x_{t2}^{*\beta_3} + u_t \quad (24)$$

Therefore, (24) is then applied to the estimators' one after the other (i.e DNLS, DNLAD and DNM estimators).
Sample sizes 32, 256, and 1024 are drawn repeatedly from the model (23). In each case, the MSE of the estimators are computed to compare the performance of the denoised nonlinear estimators, i.e. the MSE of the denoised nonlinear least squares (DNLS) estimator, denoised nonlinear least absolute deviation (DNLAD) estimator and denoised nonlinear M- estimator from 1,000 Monte Carlo samples. The analysis is carried out using R statistical package and the simulation results are summarized in the numerical tables below.

Results and Discussion

From the result of the analysis, Table 1 show the bandwidth, h (smoothing parameter), chosen for the Epanechnikov and Gaussian kernel to denoise the variables. Also Table 2 present the average, initial values, expected parameter values, bias and standard error estimates of denoised nonlinear estimators under three different sample sizes for each smoother considered. From this Table, it can be observed that the estimated parameter values of the three denoised estimators are close to the true parameter values (β_1=1.01, β_2= 0.75, β_3= 0.25). Therefore, the denoised nonlinear regression parameters are nearly unbiased.

Table 1: The bandwidth h for the variables

P			L			K		
32	**256**	**1024**	**32**	**256**	**1024**	**32**	**256**	**1024**
6.0771	2.9976	2.7091	4.2229	2.2752	1.4539	19.3420	13.2814	9.3080

Table 2: Denoise only explanatory variables

Sample size	Parameters	DNLS			DNLAD			DNM		
		β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Average expected parameter values, bias and standard error under Epanechnikov kernel smoother										
32	Initial Value	114.4507	-0.057427	1	114.4507	-0.057427	1	114.4507	-0.057427	1
	Estimates	1.0192	0.7474	0.2501	1.0235	0.7456	0.2504	1.0192	0.7475	0.2500
	S.E	0.026	0.0057	0.004	0.0364	0.0083	0.0055	0.0291	0.0064	0.0045
	Bias	0.0092	-0.0026	0.0001	0.0135	-0.0044	0.0004	0.0087	-0.0022	0.0000
256	Initial Value	113.499	-0.015045	1	113.499	-0.015045	1	113.499	-0.015045	1
	Estimates	1.0182	0.7478	0.2500	1.0211	0.7464	0.2503	1.0178	0.7479	0.2500
	S.E	0.0089	0.002	0.0014	0.0112	0.0027	0.0016	0.0104	0.0023	0.0016
	Bias	0.0082	-0.0022	0.0000	0.0112	-0.0036	0.0003	0.0078	-0.0021	0.0000
1024	Initial Value	113.952	-0.015896	1	113.952	-0.015896	1	113.952	-0.015896	1
	Estimates	1.0182	0.7478	0.2500	1.0230	0.7458	0.2503	1.0180	0.7478	0.2500
	S.E	0.0044	0.001	0.0007	0.0055	0.0013	0.0008	0.0052	0.0012	0.0008
	Bias	0.0081	-0.0022	0.0000	0.0130	-0.0042	0.0003	0.0008	-0.0022	0.0000
Average expected parameter values, bias and standard error under Gaussian kernel smoother										
32	Initial Value	112.9249	-0.011786	1	112.9249	-0.011786	1	112.9249	-0.011786	1
	Estimates	1.0148	0.7488	0.2500	1.0183	0.7464	0.2509	1.0157	0.7488	0.2498
	S.E	0.0257	0.0057	0.004	0.0364	0.0083	0.0056	0.0287	0.0064	0.0044
	Bias	0.0048	-0.0012	0.0000	0.0083	-0.0036	0.0009	0.0057	-0.0018	0.0002
256	Initial Value	113.8599	-0.015571	1	113.8599	-0.015571	1	113.8599	-0.015571	1
	Estimates	1.0161	0.7483	0.2500	1.0191	0.7471	0.2502	1.0157	0.7485	0.2500
	S.E	0.0089	0.0020	0.0014	0.0112	0.0027	0.0016	0.0104	0.0023	0.0016
	Bias	0.0061	-0.0017	0.0001	0.0091	-0.0029	0.0002	0.0057	-0.0015	0.0000
1024	Initial Value	113.7859	-0.005594	1	113.7859	-0.005594	1	113.7859	-0.005594	1
	Estimates	1.0169	0.7481	0.2500	1.0219	0.7461	0.2503	1.0168	0.7482	0.2500
	S.E	0.0044	0.0010	0.0007	0.0055	0.0013	0.0008	0.0052	0.0012	0.0008
	Bias	0.0069	-0.0019	0.0000	0.0119	-0.0042	0.0003	0.0068	-0.0022	0.0000
Average expected parameter values, bias and standard error under Wavelet smoother										
32	Initial Value	112.8807	-0.000928	1	112.8807	-0.000928	1	112.8807	-0.000928	1
	Estimates	1.0112	0.7497	0.2500	1.0159	0.7476	0.2504	1.0107	0.7499	0.2500
	S.E	0.0261	0.0058	0.0041	0.0363	0.0082	0.0055	0.0285	0.0064	0.0044
	Bias	0.0012	-0.0013	0.0000	0.0059	-0.0024	0.0004	0.0007	-0.0001	0.0001
256	Initial Value	113.6514	-0.003086	1	113.6514	-0.003086	1	113.6514	-0.003086	1
	Estimates	1.0107	0.7498	0.2500	1.0108	0.7488	0.2505	1.0109	0.7497	0.2500
	S.E	0.0089	0.0020	0.0014	0.0117	0.0028	0.0017	0.0103	0.0023	0.0016
	Bias	0.0007	-0.0002	0.0000	0.0004	-0.0012	0.0005	0.0009	-0.0013	0.0000
1024	Initial Value	113.5752	-0.005034	1	113.5752	-0.005034	1	113.5752	-0.005034	1
	Estimates	1.0109	0.7497	0.2500	1.0160	0.7476	0.2503	1.0103	0.7498	0.2500
	S.E	0.0044	0.0010	0.0007	0.0055	0.0013	0.0008	0.0052	0.0012	0.0008
	Bias	0.0009	-0.0003	0.0000	0.0060	-0.0026	0.0003	0.0003	-0.0002	0.0000
Average expected parameter values, bias and standard error under Polynomial spline smoother										
32	Initial Value	114.1800	-0.019859	1	114.1800	-0.019859	1	114.1800	-0.019859	1
	Estimates	1.0109	0.7497	0.2501	1.0155	0.7477	0.2501	1.0113	0.7496	0.2501
	S.E	0.0258	0.0058	0.0040	0.0361	0.0083	0.0056	0.0288	0.0064	-0.0044
	Bias	0.0009	-0.0003	0.0001	0.0055	-0.0023	0.0001	0.0013	-0.0004	0.0001
256	Initial Value	113.7807	-0.001800	1	113.7807	-0.001800	1	113.7807	-0.001800	1
	Estimates	1.0108	0.7497	0.2500	1.0140	0.7483	0.2502	1.0107	0.7499	0.2500
	S.E	0.0088	0.0020	0.0014	0.0117	0.0027	0.0016	0.0103	0.0023	0.0016
	Bias	0.0008	-0.0003	0.0000	0.0004	-0.0017	0.0002	0.0004	-0.0001	0.0000
1024	Initial Value	113.4900	-0.011259	1	113.4900	-0.011259	1	113.4900	-0.011259	1
	Estimates	1.0107	0.7498	0.2500	1.0161	0.7475	0.2503	1.0109	0.7497	0.2500
	S.E	0.0044	0.0010	0.0007	0.0054	0.0013	0.0008	0.0052	0.0012	0.0008
	Bias	0.0007	-0.0002	0.0000	0.0061	-0.0025	0.0003	0.0009	-0.0003	0.0000

DNLS = Denoised nonlinear least squares; DNLAD = Denoised nonlinear least absolute deviation; DNM = denoised nonlinear M- estimator

Tabe 3: Mean square error (Epanechnikov and Gaussian Kernel)

Estimators	Parameters	Epanechnikov Kernel			Gaussian Kernel		
		32	**256**	**1024**	**32**	**256**	**1024**
DNLS	β_1	0.0007606	0.0001465	0.0000866	0.0006835	0.0001164	0.0000670
	β_2	0.0000393	0.0000088	0.0000058	0.0000339	0.0000069	0.0000046
	β_3	0.0000160	0.0000020	0.0000005	0.0000160	0.0000020	0.0000005
DNLAD	β_1	0.0015072	0.0002509	0.0001993	0.0013939	0.0002083	0.0001719
	β_2	0.0000883	0.0000203	0.0000193	0.0000819	0.0000157	0.0000193
	β_3	0.0000304	0.0000027	0.0000007	0.0000322	0.0000026	0.0000007
DNM	β_1	0.0009225	0.0001690	0.0000910	0.0008562	0.0001407	0.0000733
	β_2	0.0000458	0.0000097	0.0000063	0.0000442	0.0000075	0.0000063
	β_3	.0000203	0.0000026	0.0000006	0.0000194	0.0000026	0.0000006

Table 4: Mean square error (Wavelet and Polynomial Spline)

Estimators	Parameters	Wavelet			Polynomial Spline		
		32	256	1024	32	256	1024
DNLS	β_1	0.0006827	0.0000797	0.0000202	0.0006665	0.0000781	0.0000204
	β_2	0.0000353	0.0000040	0.0000011	0.0000337	0.0000041	0.0000011
	β_3	0.0000168	0.0000020	0.0000005	0.0000160	0.0000020	0.0000005
DNLAD	β_1	0.0013525	0.0001371	0.0000663	0.0013335	0.0001212	0.0000664
	β_2	0.0000730	0.0000093	0.0000085	0.0000742	0.0000102	0.0000139
	β_3	0.0000304	0.0000031	0.0000007	0.0000315	0.0000026	0.0000007
DNM	β_1	0.0008127	0.0001069	0.0000271	0.0008311	0.0001066	0.0000279
	β_2	0.0000410	0.0000070	0.0000015	0.0000411	0.0000053	0.0000015
	β_3	0.0000194	0.0000026	0.0000006	0.0000194	0.0000026	0.0000006

Tables 3 and 4 show the estimated mean squared errors (MSE) of the denoised nonlinear estimators (DNLS, DNLAD and DNM) under the three (3) sample sizes. When the MSE under Epanechnikov, kernel, Gaussian kernel Wavelet and Polynomial spline smoothers are compare, it is obvious that DNLS is the best under the three sample sizes considered and at the same time the denoised estimators perform better under large sample size (1024).

Conclusion

This study estimate non-linear regression parameters under different sample sizes. The Epanechnikov Kernel, Gaussian Kernel, Wavelet and Polynomial Spline smoothers are used to denoise only the explanatory variables under the three (3) different sample sizes (i.e. 32, 256, and 1024). The performance of the denoised nonlinear estimators is compared based on the mean squared error criteria to determine their efficiency. The simulation studies carried out for sample sizes 32, 256,and 1024 with 1,000 Monte Carlo samples, show that the denoised nonlinear least squares (DNLS) estimator which has the smallest MSE is the best (most efficient) estimator among all the three (3) denoised nonlinear estimators under the four smoothers considered. Besides, the denoised nonlinear estimators (i.e. DNLS, DNLAD and DNM) performed better under the large sample size 1024.

References

Aldroubi A & Unser M 1996. *Wavelets in Medicine and Biology*. CRC Press, Florida, 30(2): 91-122.

Carroll RJ, Ruppert D & Stefanski LA 1995. *Measurement Error in Nonlinear Models*. Chapman and Hall.

Cheng CL & Van Ness JW 1999. *Statistical Regression with Measurement Error*. Arnold, London, pp. 3-11.

Cui HJ & Hu T 2010.On nonlinear regression estimator with denoised variables. *Computational Stat. & Data Ana.*, 55: 1137-1149.

Cui HJ, He XM & Zhu LX 2002. On regression estimators with denoised variables. *Statistical Sinica*, 12: 1191-1205.

Fasoranbaku OA & Soyombo AO 2015. A comparative study of the performance of denoised non-linear regression estimators under different smoothers.

Gonin R & Money AH 1989.Nonlinear L^P-norm Estimation. Marcel Dekker, New York, p. 34.

Green PJ & Silverman BW 1994. Non Parametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman and Hall.

Hastie TJ & Tibshirani RJ 1990. *Generalized Additive Models*. Chapman and Hall.

Hossain M, Majumder A & Basak T 2012. An application of non-linear cobb-douglas production function to selected manufacturing industries in Bangladesh. *Open J. Stat.*, 2(4): 460-468.

Jennrich RJ 1969. Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Stat*., 40: 633-643.

Keynes JM 1936. *The General Theory of Employment, Interest and Money*.

Prasad L & Lyengar SS 1997. *Wavelet Analysis with Application to Image Processing*. CRC Press, Florida.

Seber GAF & Wild CJ 1989. Nonlinear regression. *Int. J. Adv. Scientific and Technical Res.*, 5(3): 86. Available online: http://www.rspublication.com/ijst/index.html. rspublicationhouse@gmail.com,

Soyombo AO & Fasorobaku OA 2015. Non-linear model: Sensitivity of its parameters on denoised regression estimators. *Int. J. Adv. Scientific & Technical Res*., 5: 2249-9954.

William HG 2012. *Econometric Analysis*. Prentice Hall, New York, pp. 181-218.

Wonnacoth RJ & Wonnacoth TH 1979. *Econometrics*. Second edition. John Wiley & Sons, New York.