Journal of AppliedMath 2023; 1(2.1): 109.
Special Issue – Mathematical Models and Applications
Original Research Article

Impact analysis of correlated and non-normal errors in nonparametric regression estimation: A simulation study

Javaria Ahmad Khan[1,*], Atif Akbar[1], Nasir Saleem[1], Muhammad Junaid[2]

[1] *Department of Statistics, Bahauddin Zakariya University, Multan 60000, Pakistan*

[2] *Department of English, Institute of Southern Punjab, Multan 60000, Pakistan*

*** Corresponding author:** Javaria Ahmad Khan, jakhan0@yahoo.com

ARTICLE INFO

Received: 2 June 2023
Accepted: 19 July 2023
Available online: 27 July 2023

doi: 10.59400/jam.v1i2.109

ABSTRACT: In nonparametric regression, the correlation of errors can have important consequences for the statistical properties of the estimators, but the focus is on the on the identification of the effect on Average Mean Squared Error (AMSE). This is performed by a Monte Carlo experiment where we use two types of correlation structures and examine them with different correlation points/levels and different error distributions with different sample sizes. We concluded that if errors are correlated, then the distribution of errors is important with correlation structures, but correlation points/levels have a less significant effect, comparatively. When errors are uniformly distributed, AMSE is the smallest, followed by any other distribution, and if errors follow the Laplace distribution, then AMSE is the largest, followed by other distributions. Laplace also has some alarming effects. More specifically, the kernel estimator is robust in the case of a simple correlation structure, and AMSEs attain their minimum when errors are uncorrelated.

KEYWORDS: impact analysis; correlated errors; non-normal errors; nonparametric regression

1. Introduction

In real life, various aspects of life, statistically known as variables, are interconnected. To examine the relationship among these variables, the most common tool is regression. In regression estimation, the relationship between response and explanatory variables is determined, while prediction is a central issue. To estimate the response, nonparametric regression can be applied when the model is unknown and the assumptions of the model are relaxed. For this purpose, a large sample size is required then compared to parametric methods. Nonparametric regression estimation methods are based on kernels, wavelets and splines[1]. Kernel regression is a nonparametric technique used to estimate response variables by conditional expectation. The purpose is to derive the nonlinear relation between the response variable and co-variate[1]. Ullah and Vinod[2] discussed different nonparametric kernel methods, i.e., Nadaraya-Watson (NW) kernel estimator[3,4], K-nearest neighbor (NN) estimator[5,6], Mack and Muller (MM) estimator[7], Ahmad Lin estimator[8] and Gasser-Muller (GM) estimator[9].

Like other theories, nonparametric regression estimation is also based on some assumptions, like that there is no measurement error, the mean of the disturbance term is zero, there is an equal variance of the disturbance terms, there is no autocorrelation, there is zero covariance between disturbance terms and explanatory variables, there is no perfect multicollinearity, etc. If these assumptions of the model are violated, then it may cause problems. If disturbance terms or errors are correlated, then the estimates will be biased[10].

If the measurement error problem is present in the data, then the estimates are biased[11,12]. Similarly, if errors have unequal variance, then the estimates will still be unbiased but less efficient[13]. Also, it is difficult to calculate the standard deviation of the forecast errors; usually confidence intervals lie on extreme points; those become too wide or too narrow[14].

The main purpose of our study is to observe the effect of correlated errors when nonparametric regression estimation is applied. Altman[15] showed that the performance of nonparametric regression estimation is the same for both cases, whether errors are correlated or not. We are going to examine the effect of correlated errors on non-normality. There is vast literature that provides different estimators or methods to tackle correlated errors, and their performance is proven good theoretically and by simulation studies, i.e., Muller and Stadtmuller[16] focused only on a fixed design case. Their estimator was based on squared differences of various spans of the data, and Smith et al.[17] used a Bayesian method through which transformation of the dependent variable can be performed. Park et al.[14] provided an estimator that is simpler to apply because it does not require any information about the error correlation. Su and Ullah[18] used a pre-whitening transformation of the dependent variable, which is estimated from the data using the technique of local polynomials. Their new established estimator's distribution had weak dependence conditions, and they showed that it was more efficient than the local polynomial estimator. Lee et al.[19] method is very efficient in many error structures because their proposed method is based on approximating the average squared of errors.

Similarly, Chiu[20], Hart[21], Herrmann et al.[22], Opsomer et al.[23] and De Brabanter et al.[10] provided modifications in bandwidth selection methods in the presence of correlated errors, and they proved that, under some restrictions, the proposed methods provided strong, consistent results.

Methods related to our work are presented in Section 2, and the finite sample properties of the estimators and their results and related discussions are summarized in Section 3.

2. Materials and methods

Consider a nonparametric regression model

$$Y_i = m(X_i) + \varepsilon_i \qquad (1 \le i \le n) \tag{1}$$

where Y is a dependent variable, X is explanatory variable, $m(x)$ is completely amorphous and ε_i is a normal and random error. To estimate nonparametric regression, Nadaraya-Watson kernel method is used.

Nadaraya-Watson kernel estimator

This method was proposed by Nadaraya[3] and Watson[4] to estimate the unknown function $m(x)$ as given in Equation (1). To do this, they proposed an estimator as given by:

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right) y_i}{\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right)} \tag{2}$$

where, k is a kernel and h is bandwidth.

Different types of kernels, i.e., Epanechnikov, Gaussian, Tri-weight etc. are available and it is important to note that the choice of the kernel does not affect the Mean Squared Error[24].

We are using Gaussian kernel in this study as given by

$$k(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}x^2}, -\infty < x < \infty$$

Silverman[25] described that a kernel $k(x)$ is a weighting function and it is nonnegative integrable function. Kernels are used in kernel regression to estimate the conditional expectation of a random variable but must satisfy the following conditions;

- $\int_{-\infty}^{+\infty} K(x)\mathrm{d}x = 1$
- $K(-x) = K(x)$for all values of x.

The smoothing parameter, window width, or bandwidth, denoted by h, is used to manage the roughness of the curve. Different methods of bandwidth are available in the literature, which can be categorized as classical or first-generation methods and plug-in or second-generation methods. Rule of thumb, least squares cross validation, biased cross validation, etc. are part of classical methods; similarly, the direct plug-in (DPI) method and solving the equation consist of plug-in method[26,27].

In our study, we have used the plug-in method for bandwidth selection. The basic idea behind the selection of bandwidth is to obtain a value of h that minimizes the mean integrated squared error.

The effect of correlated errors is examined by different researchers in which they propose different methods to tackle this problem, like Kim et al.[28] and De Brabanter et al.[10] using a bimodal kernel technique, Su and Ullah[18] utilizing pre-whitening transformation, Lee et al.[19] approximating the squared error, etc. In our work, we examined the effect of correlated errors with four different error distributions on two different correlation structures with different correlation points. We include symmetric (normal, uniform, Laplace, and *t*) error distributions.

To evaluate the performance of different correlation points, Average Mean Squared Error (AMSE) is used[1,29], i.e., given by

$$\mathrm{AMSE} = \frac{E\left(Y_i - \widehat{m}(x)\right)^2}{n}$$

3. Monte Carlo experiment

The main purpose of our study is to examine the effect of different correlated errors on normal and independent co-variate. To perform this, the Monte Carlo experiment is conducted. Initially, we generated the error via uniform, *t*, normal and Laplace distributions when the co-variate is normally and independently distributed. We also compared the behavior of two different correlation patterns and discussed their results. For this work, two kinds of error models are considered: (i) structure given by Park et al.[14], i.e., $\varepsilon_{i+1} = \varphi\varepsilon_i + (1-\varphi^2)^{1/2}\delta_i$; where δ_i and ε_i are i.i.d. $N(0, 1)$ and $\varphi = -0.8, -0.5, -0.3, 0, 0.3, 0.5, 0.8$ and (ii) simple structure, i.e., $\varepsilon_{i+1} = \varphi\varepsilon_i$; where δ_i and ε_i are i.i.d. $N(0, 1)$ and $\varphi = -0.8, -0.5, -0.3, 0, 0.3, 0.5, 0.8$. Then we generate response variables and apply NW estimators, and the performance of the NW estimators is observed by AMSE.

The performance of the estimator is evaluated over different sample sizes. We have used $n = 25, 50, 100, 200, 500$, and 1000. For each distribution with various sample sizes, we repeat the original experiment 5000 times.

The Monte Carlo study is outlined.

3.1. The model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where, $\beta_0 = 0$ and $\beta_1 = 1$ and $x_i \sim \mathrm{U}(-1, 1)$.

The ε_i are generated from six distributions. The setup for the generation of ε_i is given by:

- $\varepsilon_i \sim \mathrm{N}(0, 2.5)$ Normally distributed with $\mu = 0$ and $\delta^2 = 2.5$.
- $\varepsilon_i \sim \mathrm{U}(-1, 1)$ Uniformly distributed over the interval (−1, 1).
- $\varepsilon_i \sim \mathrm{t}(12)$ *t*-distributed with "12" degrees of freedom.
- $\varepsilon_i \sim \mathrm{Laplace}(2, 3)$ Laplace distributed with $\alpha = 2$ and $\beta = 3$.

In this experiment plugin bandwidth is used with Gaussian kernel.

3.2. Nonparametric estimation with correlated errors with different errors

To examine the effect of different correlated errors on AMSE, we have used four different distributions of errors and results are summarized in **Tables 1** and **2**.

Table 1. AMSE for NW estimation varying sample sizes, correlation and distribution of errors—Case I.

Distributions of errors	Sample size	$\varphi = -0.8$	$\varphi = -0.5$	$\varphi = -0.3$	$\varphi = 0.0$	$\varphi = 0.3$	$\varphi = 0.5$	$\varphi = 0.8$
$\varepsilon_i \sim N(0, 2.5)$	25	4.4018	2.3599	1.5229	1.2419	1.5039	2.4599	4.4502
	50	4.3491	2.2998	1.4999	0.9999	1.4898	2.3677	4.3891
	100	4.2512	2.2471	1.4329	0.9681	1.4337	2.2471	4.2349
	150	4.1524	2.1932	1.3979	0.9412	1.3986	2.1868	4.1578
	200	3.9916	2.1060	1.3366	0.9050	1.3334	2.0992	4.0129
	500	3.7140	1.9617	1.2381	0.8270	1.2387	1.9656	3.7567
$\varepsilon_i \sim U(-1, 1)$	25	0.6091	0.8598	0.9982	1.1980	0.9921	0.8999	0.6085
	50	0.5854	0.8428	0.9614	1.0144	0.9627	0.8516	0.5836
	100	0.5536	0.806	0.9122	0.9712	0.9107	0.8076	0.5540
	150	0.5377	0.7876	0.8876	0.9496	0.8847	0.7856	0.5392
	200	0.5132	0.7529	0.8469	0.9042	0.8450	0.7497	0.5122
	500	0.4722	0.6976	0.7812	0.8391	0.7852	0.7101	0.4722
$\varepsilon_i \sim t(12)$	25	1.0272	1.1008	1.1001	1.0404	1.0929	1.1234	1.2998
	50	1.0684	1.0667	1.0474	1.0172	1.0305	1.0696	1.1609
	100	1.0966	1.0181	0.9886	0.9680	0.9871	1.0180	1.0975
	150	1.1530	0.9938	0.9592	0.9434	0.9654	0.9922	1.0658
	200	1.3687	0.956	0.9202	0.9046	0.9231	0.9512	1.0270
	500	0.9402	0.8779	0.859	0.8372	0.8508	0.8783	0.9475
$\varepsilon_i \sim \mathrm{Laplace}(2, 3)$	25	12.1525	5.4988	2.7008	1.1451	2.7122	5.4011	12.3001
	50	12.0599	5.3973	2.5887	1.0113	2.5894	5.2780	12.1037
	100	11.6341	5.1285	2.4615	0.9681	2.4666	5.0830	11.6117
	150	11.3749	4.9941	2.3979	0.9465	2.4052	4.9855	11.3657
	200	10.9866	4.8644	2.3191	0.907	2.3049	4.8167	10.9550
	500	10.2591	4.4858	2.1580	0.8371	2.1233	4.5362	10.1988

From **Table 1**, for correlation structure, $\varepsilon_{i+1} = \varphi\varepsilon_i + (1 - \varphi^2)^{1/2}\delta_i$, which is adopted from Park et al.[14], AMSEs for all error distributions are decreasing as sample size increases, and when there is no correlation, the behavior of AMSE is almost stable for small and large sample sizes. Also, it has been noted that there is no effect on the direction of the correlation.

When errors are uniformly distributed, AMSEs are smaller than for any other distribution, whether there is a correlation or not. When there is no correlation, AMSEs are smaller and decrease with sample size. It is interesting to note that in our first case, an increase in the level of correlation (φ) results in decrease of AMSEs.

When errors follow *t*-distribution, all AMSEs are very close to each other and it seems that our first correlation structure makes AMSEs robust against levels of correlation. Maybe it is due to large degrees of freedom. Like other distributions, the AMSEs have also shown a shown a decreasing trend. In the case of Laplace errors, AMSEs are very large and bear the same trend.

Table 2. AMSE for NW estimation varying sample sizes, correlation and distribution of errors—Case II.

Distributions of errors	Sample size	$\varphi = -0.8$	$\varphi = -0.5$	$\varphi = -0.3$	$\varphi = 0.0$	$\varphi = 0.3$	$\varphi = 0.5$	$\varphi = 0.8$
$\varepsilon_i \sim N(0, 2.5)$	25	4.3267	1.6343	0.5988	0.0042	0.5999	1.6045	4.1288
	50	4.1351	1.5970	0.5722	0.0040	0.5727	1.5971	4.0078
	100	3.8943	1.5166	0.5436	0.0032	0.5449	1.5187	3.8966
	150	3.8226	1.4764	0.5293	0.0021	0.5310	1.4795	3.8010
	200	3.6807	1.4107	0.5064	0.0020	0.5037	1.4203	3.6478
	500	3.4184	1.3257	0.4612	0.0016	0.4667	1.3252	3.4460
$\varepsilon_i \sim U(-1, 1)$	25	0.2132	0.0982	0.0301	0.0012	0.0389	0.0911	0.2220
	50	0.2095	0.0815	0.0292	0.0010	0.0301	0.0851	0.2179
	100	0.2048	0.0797	0.0284	0.0008	0.0284	0.0796	0.2048
	150	0.1984	0.0767	0.0272	0.0007	0.0272	0.0768	0.1982
	200	0.1892	0.0718	0.0253	0.0007	0.0252	0.072	0.1875
	500	0.1698	0.0651	0.0226	0.0006	0.0227	0.0653	0.1718
$\varepsilon_i \sim t(12)$	25	0.8001	0.2999	0.1199	0.0920	0.1267	0.2989	0.7825
	50	0.7765	0.2921	0.1110	0.0840	0.1107	0.291	0.7621
	100	0.7425	0.2893	0.1037	0.0732	0.1035	0.289	0.7449
	150	0.7255	0.2798	0.1001	0.0721	0.0999	0.2808	0.7261
	200	0.6938	0.2663	0.0937	0.0620	0.0937	0.2662	0.6897
	500	0.6373	0.2393	0.0853	0.0516	0.0854	0.2463	0.6267
$\varepsilon_i \sim \text{Laplace}(2, 3)$	25	9.874	3.8868	1.3851	0.7420	1.3704	3.8604	9.8507
	50	10.533	4.1612	1.4765	0.6540	1.4787	4.122	10.6648
	100	11.022	4.3059	1.539	0.5432	1.5336	4.2805	11.0588
	150	11.277	4.3887	1.5747	0.5321	1.5792	4.3848	11.2318
	200	11.777	4.4866	1.6461	0.4820	1.6332	4.5698	11.5555
	500	12.019	4.5322	1.789	0.4216	1.6992	4.8001	11.7211

It can be seen from **Table 2** that, in the correlation structure; $\varepsilon_{i+1} = \varphi\varepsilon_i$, AMSEs for all error distributions are decreasing as sample size increases for both positive and negative correlation points, and the same is the case for zero correlation for all types of error distributions, whether they are normal or non-normal; moreover, there is no effect of the direction of the correlation. Whatever the distribution of errors, AMSEs decreased when the intensity of the correlation of errors increased. When errors are uniformly distributed, the AMSE is the smallest, followed by any other AMSE. In case when errors follow the Laplace distribution, AMSE are larger and the interesting thing is that there whatever the distribution of errors, AMSEs decreased when intensity of correlation of errors increased. When errors are uniformly distributed, the AMSE is the smallest, followed by any other AMSE. In case when errors follow the Laplace distribution, AMSE are largest and the interesting thing is that there whatever the distribution of errors, AMSEs decreased when intensity of correlation of errors increased. When errors are uniformly distributed, the AMSE is the smallest, followed by any other AMSE. In cases where errors follow the Laplace distribution, AMSE is the largest, and the interesting thing is that there is an increasing trend with an increase in sample size.

4. Conclusions

From the above discussion, it is concluded that the higher values of correlation affect the performance of the smoother, and as the correlation approaches zero, the AMSEs are very low. The structure of correlation also matters a lot with the distribution of error. Non-normality also affects the performance of the estimator. When errors follow t and uniform distributions, the smoother performance is very good for high correlations and even supersedes the normal errors. This may lead to the utilization of t and normal errors in cases of correlated errors. The case of Laplace is very drastic, as the performance is very poor and the AMSEs are increasing with the increase in sample size and are not recommended for use.

Author contributions

Conceptualization, JAK and AK; methodology, JAK; software, JAK; validation, JAK and AK; formal analysis, JAK and AK; investigation, JAK and AK; resources, JAK, NS and MJ; data curation, JAK; writing—original draft preparation, JAK; writing—review and editing, JAK; visualization, JAK; supervision, AK, NS and MJ. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Hardle W. *Applied Nonparametric Regression*. Cambridge University Press; 1990.
2. Ullah A, Vinod HD. General nonparametric regression estimation and testing in econometrics. *Handbook of Statistics* 1993; 11: 85–116. doi: 10.1016/S0169-7161(05)80039-2
3. Nadaraya EA. On estimation regression. *Theory of Probability & Its Applications* 1964; 9(1): 141–142. doi: 10.1137/1109020
4. Watson G. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A* 1964; 26(4): 359–372.
5. Royal RM. *A Class of Nonparametric Estimators of a Smooth Regression Function* [PhD thesis]. Stanford University; 1966.
6. Mack YP. Local properties of *k*-NN regression estimates. *SIAM Journal on Algebraic and Discrete Methods* 1981; 2(3): 311–323. doi: 10.1137/0602035
7. Mack YP, Muller HG. Derivative estimation in nonparametric regression with random predictor variable. *Sankhyā: The Indian Journal of Statistics, Series A* 1989; 51(1): 59–72.
8. Ahmad IA, Lin PE. Fitting a multiple regression function. *Journal of Statistical Planning and Inference* 1984; 9(2): 163–176. doi: 10.1016/0378-3758(84)90017-X
9. Gasser T, Muller HG. Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics* 1984; 11(3): 171–185.
10. De Brabanter K, De Brabanter J, Suykens JAK, De Moor B. Kernel regression in the presence of correlated errors. *Journal of Machine Learning Research* 2011; 12: 1955–1976.
11. Maca JD, Carroll RJ, Ruppert D. *Nonparametric Kernel and Regression Spline Estimation in The Presence of Measurement Error.* Wirtschaftswissenschaftliche Fakultät, Humboldt Universität, Berlin; 1997.

12. Wang XF, Wang B. Deconvolution estimation in measurement error models: The R package decon. *Journal of Statistical Software* 2011; 39(10): i10. doi: 10.18637/jss.v039.i10
13. Li Y, Wong H, Lp W. Testing heteroscedasticity by wavelets in a nonparametric regression model. *Science in China Series A: Mathematics* 2006; 49: 1211–1222. doi: 10.1007/s11425-006-2016-2
14. Park BU, Lee YK, Kim TY, Park C. A simple estimation of error correlation in non-parametric regression models. *Scandinavian Journal of Statistics* 2006; 33(3): 451–462. doi: 10.1111/j.1467-9469.2006.00506.x
15. Altman NS. Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association* 1990; 85(411): 749–759. doi: 10.2307/2290011
16. Muller HG, Stadtmuller U. Detecting dependencies in smooth regression models. *Biometrika* 1988; 75(4): 639–650. doi: 10.2307/2336305
17. Smith M, Wong CM, Kohn R. Additive nonparametric regression with autocorrelated errors. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 1998; 60(2): 311–331.
18. Su L, Ullah A. More efficient estimation in nonparametric regression with nonparametric autocorrelated errors. *Econometric Theory* 2006; 22(1): 98–126. doi: 10.1017/S026646660606004X
19. Lee YK, Mammen E, Park BU. Bandwidth selection for kernel regression with correlated errors. *Statistics* 2010; 44(4): 327–340. doi: 10.1080/02331880903138452
20. Chiu ST. Bandwidth selection for kernel estimate with correlated noise. *Statistics & Probability Letters* 1989; 8(4): 347–354. doi: 10.1016/0167-7152(89)90043-6
21. Hart JD. Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society. Series B (Methodological)* 1991; 53(1): 173–187. doi: 10.1111/J.2517-6161.1991.TB01816.X
22. Herrmann E, Gasser T, Kneip A. Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika* 1992; 79(4): 783–795. doi: 10.2307/2337234
23. Opsomer J, Wang Y, Yang Y. Nonparametric regression with correlated errors. *Statistical Science* 2001; 16(2): 134–153.
24. Benedetti JK. On the nonparametric estimation of regression functions. *Journal of the Royal Statistical Society. Series B (Methodological)* 1977; 39(2): 248–253. doi: 10.1111/j.2517-6161.1977.tb01622.x
25. Silverman BW. *Density Estimation for Statistics and Data Analysis,* 1st ed. Chapman and Hall; 1986.
26. Wand MP, Jones MC. *Kernel Smoothing*. Chapman and Hall/CRC; 1995.
27. Loader CR. Bandwidth selection: Classical or plug-in? *Annals of Statistics* 1999; 27(2): 415–438. doi: 10.1214/aos/1018031201
28. Kim TY, Park BU, Moon MS, Kim C. Using bimodal kernel for inference in nonparametric regression with correlated errors. *Journal of Multivariate Analysis* 2009; 100(7): 1487–1497. doi: 10.1016/j.jmva.2008.12.012
29. Wu H, Zhang JT. *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches.* John Wily and Sons; 2006.