Pak.j.stat.oper.res. Vol.21 No. 4 2025 pp 545-554 DOI: https://doi.org/10.18187/pjsor.v21i4.4746

Pakistan Journal of Statistics and Operation Research

Empirical Performance of Nonparametric Regression with Heteroscedasticity

Javaria Ahmad Khan[1], Atif Akbar*[2], Muhammad Ejaz[3]

* Corresponding Author

1. Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan, jakhan0@yahoo.com
2. Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan, atifakbar@bzu.edu.pk
3. Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan, ejazamin@bzu.edu.pk

Abstract
Heteroscedasticity is a well-known violation of an assumption in parametric regression analysis. In such cases, to handle this problem, a generalized least squares method is used. In this article, we have manifested the robustness of nonparametric regression in the case of heteroscedastic errors. Nonparametric regression is a robust method that proceeds without requiring inflexible assumptions from the model. We empirically compared the performance of the generalized least squares method with multivariate nonparametric kernel regression. Multivariate nonparametric kernel regression is used with a Gaussian kernel and six bandwidths on China's per capita consumption expenditure. The performance of nonparametric regression with Bayesian bandwidth was found better on the basis of mean squared error. Simulation results are also presented, with their graphical representation, where nonparametric regression with different bandwidths at different heteroscedastic levels is observed, and we found that our proposed method performed best in both presence and absence of homoscedasticity.

Key Words: Nonparametric regression; Bandwidth; China per capita data; Heteroscedasticity; Simulation

Mathematical Subject Classification: 62G08, 62G35

1. Introduction

Regression analysis is a statistical process that is specifically used to estimate the association between income and expenditure, sales and purchases, etc. Not only this, regression analysis has large applications, i.e., for forecasting future opportunities, optimizing business processes, empirical support to management decisions, and also for identifying errors in judgment, etc. (Syla, 2013; Rusov et al., 2017). When the relationship between the response and explanatory variables is known, parametric regression models are used. Parametric models are inflexible and have some constraints, like the linear relationship among response and explanatory variable(s), uncorrelated errors, homoscedasticity of the errors, and errors following the normal distribution (Leeflang et al. 2000). When assumptions about parametric models are violated, then it provides misleading results (DiNardo and Tobias, 2001).

If the relationship is unknown, it is better to use nonparametric regression models (Hardle, 1990; Stuart et al., 1999). The nonparametric regression (NPR) approach is popular due to less restrictive assumptions and is applied with the help of information derived from the data (Stuart et al. 1999). Nonparametric kernel regression is based on kernels that are constrained symmetrical and nonnegative real-valued functions K, whose integration results in unity (Hardle, 1990). Bandwidth (b) or smoothing parameter plays a very important role in the performance of the kernel estimators. The optimal value of b minimizes the mean integrated squared error (MISE). Various methods for selection of b are featured, which can be categorized into two quite different approaches: Classical and Plug-in (Sheather et al. 1991; Loader, 1999; Wand and Jones, 1995).

The wide applications of regression can be observed in Agriculture (Ujjainia et al., 2020), Finance (Anghelache and Anghel, 2014), Management sciences (Henderson and Souto, 2018), Biology (Paul and Saha, 2007), Ecology (Akselrud, 2024), Engineering (Robinson, 2003), Insurance (Siddig, 2016), Medical Sciences (Grover et al. 2013), Mining (Mafudi and Suyono, 2018), Toxicology (Shaki, 2024), Physics (Krueger, 2011), and

Chemical Sciences (Khan and Akbar, 2019). Heteroscedasticity is the violation of homoscedasticity. This problem is encountered in data sets when the error term differs across values of an independent variable. The impact of violating the assumption of homoscedasticity is a matter of degree, increasing as heteroscedasticity increases. In this case, the use of the weighted least squares (WLS) model is suggested, but if there is a certain correlation among residuals, WLS gives misleading inference. Aitken (1934) suggested to use of a generalized least squares (GLS) model, which provides unbiased, efficient, consistent, and asymptotically normal estimators. Similarly, Rao (1973) suggested using minimum norm quadratic unbiased estimation (MINQUE). The efficiency losses of MINQUE are not substantial when the number of observations per sample is large, especially for a small number of independent samples. A rich literature is available, which justifies the use of a heteroscedasticity consistent covariance matrix estimator (HCCME). By using these estimators, testing inferences are performed, and results are asymptotically valid, regardless of whether or not the errors share the same variance (Aslam, 2014).

The present study is motivated by the lack of research on the performance of NPR on heteroscedastic data when such data is modeled, particularly by GLS. It occurs more often in datasets that have a large range between the largest and smallest observed values. Usually, heteroscedasticity is detected by some specific tests, e.g., Park test (1966), Glejser test (1969, 2000), Breusch-Pagan test (1979), White test (1980), etc., and then heteroscedasticity is removed by different methods like logarithmized data, Box-Cox transformation, etc. Here we are going to model such data by the nonparametric regression method, and model the data without removing the heteroscedasticity. Here, we are going to inquire about its robustness via simulated and real data. This article is planned in five sections. In the Second section, NPR and GLS are introduced, and Section 3 explains the concerned methods for selecting the bandwidth. In Section 4, real data example and the Monte Carlo simulation experiments are presented, and Section 5 concludes.

2. Methodology

2.1 Multivariate Nonparametric regression model

The relationship between a response (y) and a set of explanatory variables ($\boldsymbol{x}=(x_1,x_2,...x_d)'$) can be described by the multivariate kernel regression model as,

$$y_i=m(\boldsymbol{x}_i)+\varepsilon_i,$$

where ε_i, are assumed to be independent and identically distributed (iid) for i=1, 2, ..., n, with mean zero and variance σ_m^2. The Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964) of $m(.)$ is given by

$$\hat{m}(x,b)=\frac{n^{-1}\sum_{i=1}^{n}K_b(x\text{-}x_i)y_i}{n^{-1}\sum_{j=1}^{n}K_b(x\text{-}x_j)}, \quad (1)$$

where $\boldsymbol{b}=(b_1, b_2, ...,b_d)'$ is a vector of bandwidths with all positive elements, and

$$K_b(x)=\frac{1}{(b_1, b, ...,b_d)}K\left(\frac{x_1}{b_1},\frac{x_2}{b_2},..., \frac{x_d}{b_d}\right)$$

with $K(.)$ denotes a multivariate kernel function. Hardle (1990) expressed equation (1) as;

$$\hat{m}(x,b)=\frac{1}{n}\sum_{i=1}^{n} w_{b,i}(x)y_i, \quad (2)$$

where

$$w_{b,i}(x)= K_b(x\text{-}x_i)/\hat{f}_b(x),$$

$$\hat{f}_b(x)=\frac{1}{n}\sum_{i=1}^{n}K_b(x\text{-}x_i).$$

2.2 Generalized least square (GLS)

The generalization of the ordinary least squares (OLS) estimator is known as the generalized least squares (GLS) estimator of the coefficients of a linear regression. This method is used as a remedial measure when there is heteroscedasticity in the data and the absence of serial correlation is violated. In such a situation, the OLS estimator does not remain BLUE (best linear unbiased estimator) so, the GLS estimator is BLUE. Here, we use the equation as

$$\tilde{y}=\tilde{X}\beta+\tilde{\varepsilon}, \tag{3}$$

where $\tilde{y}=\Sigma^{-1}y$, $\tilde{X}=\Sigma^{-1}X$ and $\tilde{\varepsilon}=\Sigma^{-1}\varepsilon$, with Σ is an invertible matrix (Baltagi, 2008).

3. Bandwidths

As we discussed before, in nonparametric estimation, the selection of bandwidth (b) is very crucial. The following are data-driven bandwidths that can deal with multivariate regression. From a classical type of bandwidth category, biased and least square cross-validation (BCV and LSCV), Bayesian, smoothed cross-validation (SCV), and Normal scale rule (NSR), and from the Plug-in category, Direct plug-in (DPI) (Loader, 1999; Jones et al., 1996).

3.1 Least squares and biased cross-validation (LSCV and BCV)

Sain et al., (1994) explicitly derived and compared multivariate versions of the least-squares cross-validation method (due to its unbiasedness, the LSCV selector is sometimes called the unbiased cross-validation (UCV) selector) developed by Bowman (1984) and Rudemo (1982) and a biased cross-validation method similar to that of Scott and Terrell (1987) for multivariate kernel estimation using the product kernel estimator. The multivariate versions are straightforward generalizations of the univariate form.

$$LSCV(b_1, b_2, \dots, b_d) = \frac{1}{\left(2\sqrt{\pi}\right)^d n b_1, , \dots, b_d} + \frac{1}{\left(2\sqrt{\pi}\right)^d n^2 b_1, , \dots, b_d}$$

$$\times \sum_{i=1}^{n} \sum_{j\neq i} \left[exp\left\{-\tfrac{1}{4}\sum_{k=1}^{d} \Delta_{ijk}^2\right\} - \left(2\times 2^{d/2}\right) exp\left\{-\tfrac{1}{2}\sum_{k=1}^{d} \Delta_{ijk}^2\right\}\right]. \tag{4}$$

$$BCV(b_1, b_2, \dots, b_d) = \frac{1}{\left(\sqrt{2\pi}\right)^d n b_1, , \dots, b_d} + \frac{1}{4n(n-1) b_1, , \dots, b_d}$$

$$\times \sum_{i=1}^{n} \sum_{j\neq i} \left[\left(\sum_{k=1}^{d} \Delta_{ijk}^2\right) - (2d+4)\left(\sum_{k=1}^{d} \Delta_{ijk}^2\right) + (d^2+2d\right] \times \prod_{k=1}^{d} \phi(\Delta_{ijk}). \tag{5}$$

where $\Delta_{ijk} = \left(x_{ik} - x_{jk}\right)/b_k$.

3.2 Smoothed cross-validation (SCV)

Chacon and Duong (2011) presented this type of cross-validation. The smoothed cross-validation method is famous, due to its specialty that it reduces the variability of its non-smoothed counterpart. However, it shares a pilot bandwidth matrix with the plug-in. The choice of an optimal pilot bandwidth matrix is full of mathematical difficulties. But this method overcomes this problem by unconstrained pilot matrix. So, this method can be performed as follows:

$$SCV(b_1, b_2, \dots, b_d) = n^{-1}(b_1, b_2, \dots, b_d)^{-1/2} R(K) + n^{-2} \sum_{i,j=1}^{n} \left(\bar{\Delta}_{(b_1,b_2,\dots,b_d)} * \bar{L}_G\right)\left(X_i - X_j\right), \tag{6}$$

where L_G is a kernel with pilot bandwidth, $R(K) = \int_R K(x)^2 dx$, $\bar{\Delta}_{(b_1,b_2,\dots,b_d)} = \bar{K}_{(b_1,b_2,\dots,b_d)} - 2K_{(b_1,b_2,\dots,b_d)} + K_0$ and K_0 is a notation for the Dirac delta function.

3.4 Bayesian

Zhang et al., (2009) described a Bayesian approach for multivariate kernel regression. In which they considered that ε_i, for $i = 1, 2, \dots, n$, are *iid* and follow $N(0, \sigma_m^2)$ with σ_m^2 an unknown parameter. It follows that $\frac{y_i - m(x_i)}{\sigma_m} \sim N(0,1)$.

Let $\pi(\boldsymbol{b})$ and $\pi(\sigma_m^2)$ denote the prior densities of $\boldsymbol{b}$ and σ_m^2 respectively, and denote $(y_1, y_2, \dots, y_n)'$ as y. According to Bayes theorem, the posterior of $(b, \sigma_m^2)'$ is

$$\pi(\boldsymbol{b}, \sigma_m^2|\boldsymbol{y}) \propto \pi(\boldsymbol{h})\pi(\sigma_m^2)l((y_1, y_2, \dots, y_n|\boldsymbol{b}, \sigma_m^2).$$

They assume that the prior density of σ_m^2 is

$$\pi(\sigma_m^2) \propto \left(\frac{1}{\sigma_m^2}\right)^{p/2+1} exp\left\{-\frac{v/2}{\sigma_m^2}\right\},$$

where p and v are hyperparameters. The prior density of h_j is assumed to be

$$\pi(h_j) \propto \frac{1}{1+\lambda b_j^2},$$

Then the posterior of $\pi(\boldsymbol{b}, \sigma_m^2|\boldsymbol{y})$ becomes

$$\pi(\boldsymbol{b}, \sigma_m^2|\boldsymbol{y}) \propto \prod_{j=1}^{d} \pi(b_j) \left(\frac{1}{\sigma_m^2}\right)^{(n+p)/2+1} exp\left\{-\frac{\sum_{i=1}^{n}\left(y_i - \widehat{m}_{-i}(\boldsymbol{x_i}, \boldsymbol{b})\right)^2 + v}{2\sigma_m^2}\right\}. \quad (7)$$

3.5 Normal scale rule (NSR)

Chacon et al. (2011) discussed that for multivariate regression major obstacle is tackling the matrix analysis when treating higher-order multivariate derivatives. With an alternative vectorization of these higher-order derivatives, they exhibit a closed-form expression for a normal scale bandwidth matrix.

$$\widehat{\boldsymbol{B}}_{NS} = \left(\frac{4}{d+2r+2}\right)^{2/(d+2r+4)} \widehat{\boldsymbol{\Sigma}} n^{-2/(d+2r+4)}, \quad (8)$$

where $\widehat{\boldsymbol{\Sigma}}$ is an estimate of $\boldsymbol{\Sigma}$ which is the variance-covariance matrix.

3.6 Direct plug-in (DPI)

Chacon and Duong (2010) explained the crucial factor of multivariate regression, which requires an optimal bandwidth matrix with restricted parameterizations. They presented the first plug-in bandwidth selector with unconstrained parameterizations by introducing an alternative vectorization that gives elegant and tractable expressions.

$$\widehat{\boldsymbol{B}}_{PI,l} = \underset{\boldsymbol{B}\epsilon\mathcal{F}}{argmin} \left\{n^{-1}|\boldsymbol{B}|^{-1/2}(4\pi)^{-d/2} + \frac{1}{4}(vec^T\boldsymbol{B})^{\otimes 2}\widehat{\boldsymbol{\psi}}_{4,l}\right\}, \quad (9)$$

where $\boldsymbol{H}$ is the bandwidth matrix, $\mathcal{F}$ is the set of all symmetric and positive definite $d \times d$ matrices where d represents the length of regressors, $\otimes 2$ is the second Kronecker product of matrices and $\widehat{\boldsymbol{\psi}}_{4,l}$ fourth-order integrated density derivatives of $f(x)$.

4. Empirical Evaluation

4.1 Simulation

Now we are going to examine its performance at different levels of heteroscedasticity. The purpose of this section is to compare the performance of GLS with NPR by using different bandwidths through simulation, with varying sample sizes.

In this sub-section, we used the following model;

$$y_i = x_{1i} + x_{2i} + \sigma(x_i)\varepsilon_i, \quad i=1,2,\dots,n. \quad (10)$$

Where $x_{1i}, x_{2i} \sim N(0,1)$ and $\varepsilon_i \sim N(0,1)$. By following Lei and Chang-Lin (2008), we use $\sigma(x_i) = exp(\lambda x_1)$ as a variance function, where λ describes the intensity levels of heteroscedasticity. In this article, we use λ=0, 0.25, 0.50, 1, 2, 4. Simulation results are presented in Table 1 for n=25, 50, 100, 200, and 500 with 1000 replications, and they are graphically represented in Figure 1.

Table 1. AMSE of GLS and NPR

λ = 0							
	BCV	LSCV	NSR	DPI	SCV	Bayesian	LRM
25	0.4492	0.9326	**0.3438**	0.5257	0.7465	0.5119	0.8811
50	0.5076	**0.2115**	0.4346	0.3482	0.4995	0.9308	0.9391
100	0.5745	**0.3033**	0.4964	0.5051	0.6294	1.1479	0.9700
200	0.5949	0.5328	0.5213	**0.4831**	0.5986	1.3749	0.9854

500	0.6798	0.7109	**0.6150**	0.6407	0.6819	1.5307	0.9951
			λ = 0.25				
	BCV	LSCV	NSR	DPI	SCV	Bayesian	GLS
25	0.38378	**0.0521**	0.3178	0.1525	0.3949	1.2176	0.9592
50	0.5787	0.7156	**0.4938**	0.5073	0.6480	0.9026	0.9094
100	0.6406	0.8375	**0.5521**	0.6111	0.7405	1.1744	1.1076
200	0.6574	0.6308	0.5782	**0.5745**	0.6750	1.5159	1.1254
500	0.6874	**0.4638**	0.6103	0.5270	0.6217	1.6726	1.0734
			λ = 0.50				
	BCV	LSCV	NSR	DPI	SCV	Bayesian	GLS
25	0.4328	0.3395	0.3551	**0.2492**	0.4312	1.0740	1.0822
50	0.5787	0.8669	**0.4489**	0.4929	0.7411	1.0192	1.1551
100	0.6336	**0.3729**	0.5435	0.5302	0.6807	1.3000	1.2356
200	0.9821	1.0768	**0.8648**	0.8831	0.9823	2.1452	1.8524
500	0.9764	0.8129	0.8586	**0.7881**	0.8905	2.2254	1.6994
			λ = 1				
	BCV	LSCV	NSR	DPI	SCV	Bayesian	GLS
25	3.8423	10.0857	**2.8884**	3.8755	7.1877	7.1960	14.2174
50	1.3508	1.5289	1.0420	**0.7736**	1.8195	3.3697	3.9161
100	1.8559	1.8704	1.4152	**1.2019**	2.0777	6.2091	6.8970
200	2.1750	2.8988	**1.7870**	2.0598	2.4768	5.4114	5.2326
500	2.2833	**1.8943**	1.9659	1.9602	2.3141	7.5118	7.1871
			λ = 2				
	BCV	LSCV	NSR	DPI	SCV	Bayesian	GLS
25	0.9695	**0.1418**	0.7307	0.6413	1.5582	6.2920	5.6008
50	77.2308	**7.2274**	48.3309	19.1839	72.66355	163.2193	651.8484
100	20.8757	33.83195	**18.1412**	19.7735	24.6865	451.7273	13681.8600
200	110.6608	96.7910	**91.3962**	97.4926	124.6419	436.9494	479.0530
500	91.9671	78.6006	78.5181	**76.9889**	91.5080	3281.7380	4078.3850
			λ = 4				
	BCV	LSCV	NSR	DPI	SCV	Bayesian	GLS
25	413489.7	855056.1	222547.8	**117520.9**	593092.9	808602.4	3504016.0
50	39741.4	3601096.0	**31644.5**	49461.2	320831.1	16092317.0	26910577.0
100	43386.4	139206.3	**20444.0**	32008.1	72417.8	413760.7	443436.3
200	190107.9	33160.3	101489.0	**32192.5**	107144.0	63071113.0	96042820.0
500	671692.1	**385788.2**	420766.10	443146.6	677168.6	16151916 .0	17095630.0

In Table 1, AMSEs of NPR and GLS are presented. In this section, six different bandwidths are used for comparison. It is evident that the performance of NPR is better than the GLS for all levels of heteroscedasticity. AMSEs are increasing rapidly as levels of heteroscedasticity are growing, which is more evident in Figure 1, but NPR is less affected than GLS.

For the homoscedastic case (λ = 0), NPR is performing better with different bandwidths (for different sample sizes). For the moderate level of heteroscedasticity, LSCV and DPI exhibit acceptable performance. In the case of a severe level of heteroscedasticity (λ = 2, 4), the performance of NPR is much better than GLS with different bandwidths, i.e., DPI, NSR, and LSCV.

In our simulation scheme, no bandwidth performs uniformly best for all heteroscedastic levels, but generally, LSCV, NSR, and DPI perform better for all sample sizes. From Figure 2, performance of bandwidths can be examined. It is apparent that the AMSEs of all bandwidths increase as the level of heteroscedasticity increases, yet they still outperform LRM/GLS.

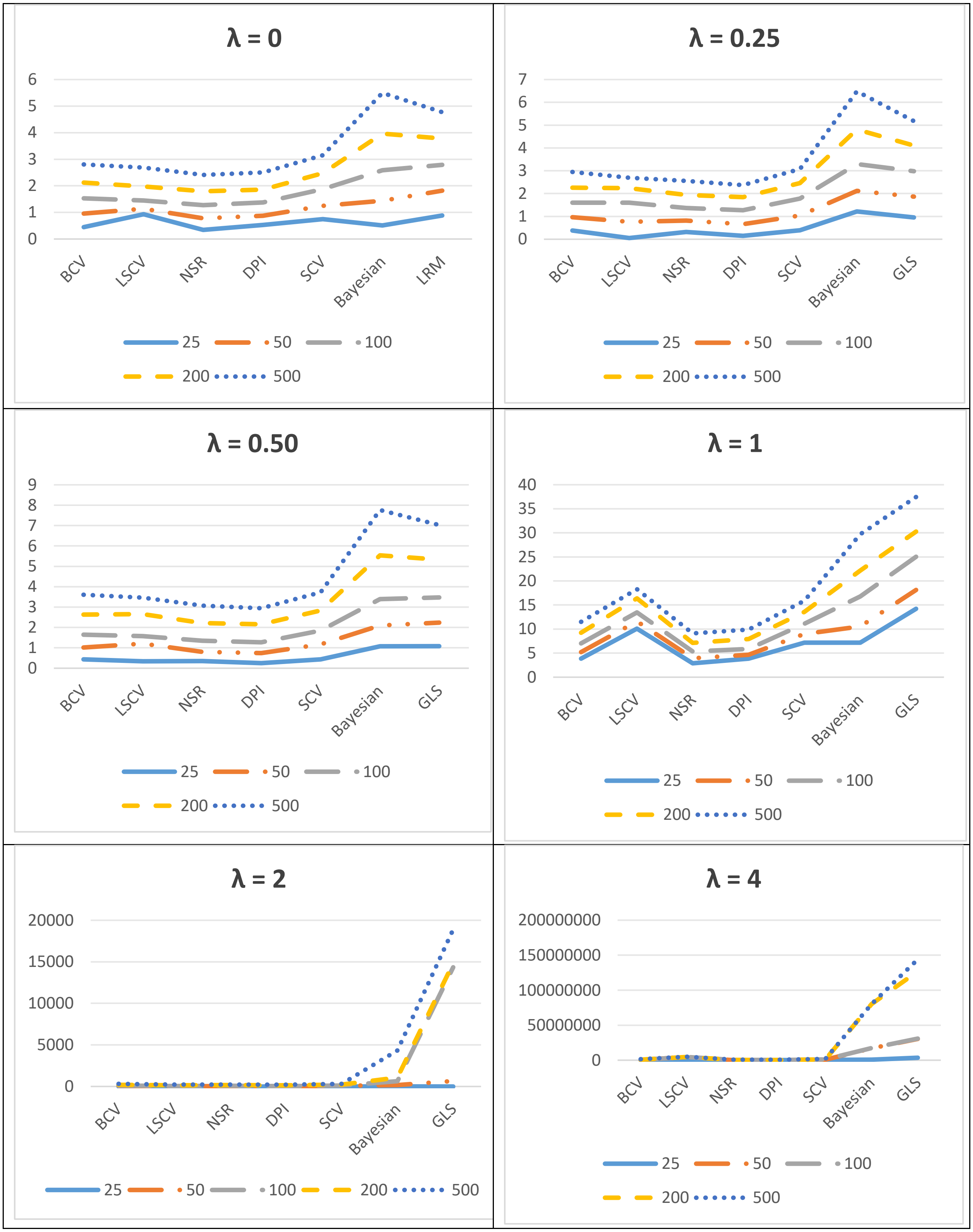

Figure 1: AMSEs according to sample size varying bandwidths

 DOI: https://doi.org/10.18187/pjsor.v21i4.4746

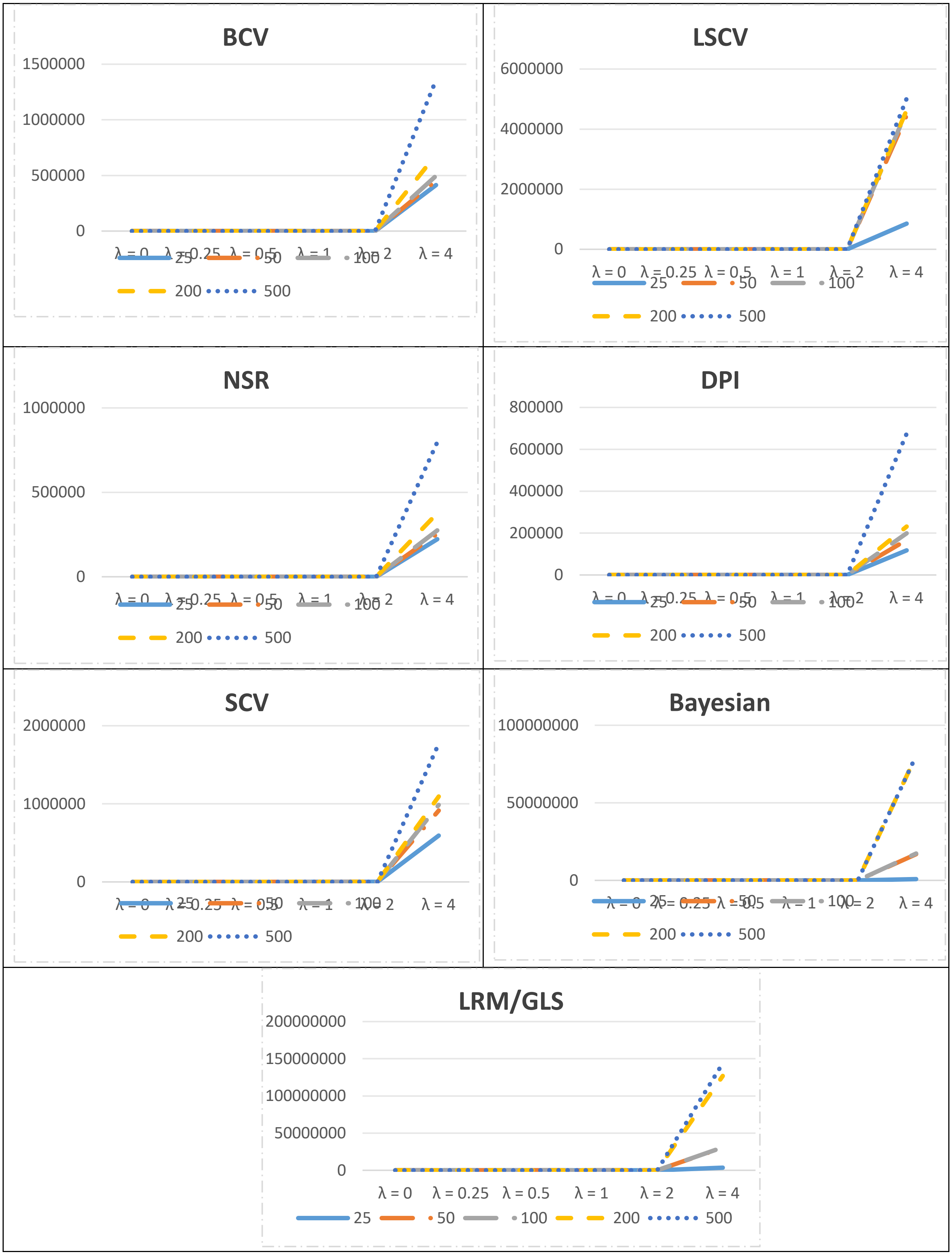

Figure 2: AMSEs according to sample size varying heteroscedasticity levels

4.2 China per capita data

Here, we will use a real data set to illustrate the results. A data set is based on per capita consumption expenditure (PCCE), agricultural business income (ABI), and other income (OI) for 31 different regions in China in 2009. This data is also used by Su et al. (2012), who mentioned that the data has a problem of heteroscedasticity, and this is also evident in Table 2.

Table 2. MSE of China per capita data consumption expenditure (PCCE)

Methods	Before Transformation	After Transformation
Breusch-Pagan test	Chi-square =11.53282 p = 0.00068	Chi-square = 0 .8332667 p = 0.36133
GLS/LRM	51589.9	6.221653e-13
BCV	650471.8	4.007885e-12
LSCV	649914.6	4.005497e-12
NSR	650470.8	4.007876e-12
SCV	650457.3	4.007843e-12
DPI	650457.8	4.007834e-12
Bayesian	**21.6**	**1.897534e-15**

In Table 2, Breusch-Pagan test is applied to examine the heteroscedasticity of data. It can be examined that in the presence of heteroscedasticity, performance of NPR is outstanding as compared to GLS in form of Mean Squared Error (MSE). NPR is applied with above mentioned bandwidths, where Bayesian bandwidth selection method performs very well. Further, Box-Cox transformation is applied to remove the heteroscedasticity and all methods are applied again. It can be examined that after transformation, heteroscedasticity is removed and after removing heteroscedasticity we again examined the MSE of data and found that NPR with Bayesian bandwidth is better than GLS, this time too.

5. Discussion and Conclusion

Regression is a famous tool for inquiring about performance indicators in any field of life. It is well known fact that before the implementation, parametric regression required different assumptions to be satisfied. If the equal variance assumption is violated, the ordinary least squares (OLS) estimators will no longer be Best Linear Unbiased Estimators (BLUE), and resultantly, the regression predictions will be inefficient too, and as a result, the tests of hypotheses (t-test, F-test) are no longer valid (Gujrati, 2008). Different approaches are used to overcome these problems. NPR is also a famous technique and imposes fewer restrictions due to its robustness.

In this article, we showed the robustness of NPR under heteroscedasticity. We showed the NPR performance with a real data example. Data is based on China's per Capita income and expenditure. Firstly, we apply the Breusch-Pagan test to show that the data is heteroscedastic. We applied GLS and NPR with considered bandwidths and a Gaussian kernel, and we found that the performance of NPR is better with Bayesian bandwidth than GLS in terms of MSE (Table 1). Then we removed the heteroscedasticity by using the Box-Cox transformation method. After removing heteroscedasticity, the considered methods are applied again. In this case, again, NPR performs better with Bayesian bandwidth than LRM. Then we design a simulation study to compare GLS and NPR methods at different levels of heteroscedasticity. We observed that NPR outperformed GLS with all considered bandwidths than GLS. No bandwidth performs optimally in every scenario, but NSR, LSCV, and DPI perform superior than others.

Fortunately, significance tests are generally robust to mild heteroscedasticity, so OLS estimates remain reliable in most cases. Nonetheless, severe heteroscedasticity can lead to problematic distortions (Berry and Feldman, 1985). From this whole study, it can be concluded that it is unnecessary to eliminate heteroscedasticity and to know about the severity of heteroscedasticity. NPR does not require such a process and performs well with all levels of heteroscedasticity. Therefore, we recommend a nonparametric approach as the most effective solution, given its robustness and speed. It remains unaffected by the degree of heteroscedasticity and yields the minimum mean squared error (MSE).

References

1. Aitken, A. C. (1934). On least squares and linear combinations of observations. Proceedings of the Royal Society of Edinburgh, 55, 42-48.
2. Akselrud, C. I. A. (2024). Random forest regression models in ecology: Accounting for messy biological data and producing predictions with uncertainty. Fisheries Research, 280, 107161.
3. Anghelache, C. and Anghel, M. G. (2014). Using the regression model in the analysis of financial instruments portfolios. Procedia Economics and Finance, 10, 324-329.
4. Aslam, M. (2014). Using heteroscedasticity-consistent standard errors for the linear regression model with correlated regressors. Communications in Statistics – Simulation and Computation, 43, xx-xx.
5. Baltagi, B. H. (2008). Generalized Least Squares. In Econometrics, Springer: Berlin, Heidelberg.
6. Berry, W. D. and Feldman, S. (1985). Quantitative Applications in the Social Sciences: Multiple Regression in Practice. SAGE Publications Inc: Thousand Oaks, CA.
7. Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 71, 353-360.
8. Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica, 47(5), 1287-1294.
9. Carroll, R. J. and Ruppert, D. (1988). Transformation and Weighting in Regression. Chapman and Hall: New York.
10. Chatterjee, S. and Hadi, A. S. (1988). Sensitivity Analysis in Linear Regression. Wiley: New York.
11. Cook, R. D. (1977). Detection of influential observation in linear regression. Technometrics, 19, 15-18.
12. Cook, R. D. and Weisberg, S. (1982). Residuals and Influence in Regression. Chapman and Hall: New York.
13. Cribari-Neto, F. (2004). Asymptotic inference under heteroscedasticity of unknown form. Computational Statistics & Data Analysis, 45(2), 215-233.
14. Davidson, R. and MacKinnon, J. G. (1993). Estimation and Inference in Econometrics. Oxford University Press: New York.
15. Draper, N. R. and Smith, H. (1998). Applied Regression Analysis. 3rd ed. Wiley: New York.
16. Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1, 59-82.
17. Fox, J. (2016). Applied Regression Analysis and Generalized Linear Models. 3rd ed. SAGE Publications: Los Angeles.
18. Goldberger, A. S. (1991). A Course in Econometrics. Harvard University Press: Cambridge, MA.
19. Greene, W. H. (2012). Econometric Analysis. 7th ed. Prentice Hall: Upper Saddle River, NJ.
20. Gujarati, D. N. and Porter, D. C. (2009). Basic Econometrics. 5th ed. McGraw-Hill: New York.
21. Hinkley, D. V. (1977). Jackknife and bootstrap. Biometrika, 64(3), 331-340.
22. Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1, 221-233.
23. Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T. C. (1985). The Theory and Practice of Econometrics. 2nd ed. Wiley: New York.
24. Koenker, R. and Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33-50.
25. Koenker, R. (2005). Quantile Regression. Cambridge University Press: New York.
26. Long, J. S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. The American Statistician, 54(3), 217-224.
27. MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics, 29(3), 305-325.
28. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2012). Introduction to Linear Regression Analysis. 5th ed. Wiley: Hoboken, NJ.
29. Mafudi, M., & Suyono, E. (2018). Firm specific attributes and corporate social responsibility disclosure in manufacturing and mining companies: evidence from Indonesia. Journal of Economic & Management Perspectives, 12(4), 14-27.

30. Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3), 703-708.
31. Neter, J., Wasserman, W. and Kutner, M. H. (1989). Applied Linear Regression Models. 2nd ed. Irwin: Homewood, IL.
32. Rao, C. R. (1973). Linear Statistical Inference and Its Applications. 2nd ed. Wiley: New York.
33. Robinson, P. M. (1987). Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. Econometrica, 55(4), 875-891.
34. Sen, P. K. and Singer, J. M. (1993). Large Sample Methods in Statistics: An Introduction with Applications. Chapman and Hall: New York.
35. Shaki, F., Amirkhanloo, M., & Chahardori, M. (2024). The Future and Application of Artificial Intelligence in Toxicology. Asia Pacific Journal of Medical Toxicology, 13(1).
36. Theil, H. (1971). Principles of Econometrics. Wiley: New York.
37. Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley: Reading, MA.
38. Ujjainia, S. Gautam, P., & Veenadhari, S. (2020). Crop yield prediction using regression model. International journal of innovative technology and exploring engineering, 9, 269-273.
39. Vogelsang, T. J. (2012). Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. Journal of Econometrics, 170(2), 209-233.
40. Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society, 54(3), 426-482.
41. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48(4), 817-838.
42. White, H. (1982). Maximum likelihood estimation of misspecified models. Econometrica, 50(1), 1-25.
43. White, H. (1994). Estimation, Inference and Specification Analysis. Cambridge University Press: New York.
44. Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. 2nd ed. MIT Press: Cambridge, MA.
45. Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics, 14(4), 1261-1295.
46. Xu, K. L. and Phillips, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variances. Journal of Econometrics, 142(1), 265-280.
47. Zhang, X., Wan, A. T. K. and Wong, A. (2012). Tests for structural change in partially linear panel data models. Econometric Theory, 28(2), 413-444.