Research Article

Coverage Properties of a Neural Network Estimator of Finite Population Total in High-Dimensional Space

Festus A. Were [iD],[1] **George O. Orwa** [iD],[1] **Romanus O. Otieno** [iD],[1] **Nicholas Makumi** [iD],[1] **and Ramy Aldallal** [iD][2]

[1]*Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya*
[2]*College of Business Administration in Hotat bani Tamim, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia*

Correspondence should be addressed to Festus A. Were; were.festus2022@students.jkuat.ac.ke

The problem in nonparametric estimation of finite population total particularly when dealing with high-dimensional datasets is addressed in this paper. The coverage properties of a robust finite population total estimator based on a feedforward backpropagation neural network developed with the help of a superpopulation model are computed, and a comparison with existing model-based estimators that can handle high-dimensional datasets is conducted to evaluate the estimator's performance using simulated datasets. The results presented in this paper show good performance in terms of bias, MSE, and mean absolute error for the feedforward backpropagation neural network estimator as compared to other identified existing estimators of finite population total in high-dimensional datasets. In this regard, the paper recommends the use of the proposed estimator in estimating population parameters such as population total in the presence of high-dimensional datasets.

1. Introduction

Assume that there is a finite population of N unique and identifiable units; $U = \{1, 2, \ldots, N\}$. Let each population unit to have the variable of interest Y. It is assumed that auxiliary variable $X \in \mathbb{R}^d$ exists which is closely related with Y and is known for the entire population (i.e., $X_1, X_2 \ldots, X_N$). Researchers encounter the problem of estimating a population function (i.e., a function of $Y's$), for instance, the population total.

$$T = \sum_{i=1}^{N} Y_i. \tag{1}$$

While estimating the population total T, a sample s is picked so that the pair $(x_{i,j}, y_i)$, $i = 1, 2, \ldots, n$ and $j = 1, 2, 3, \ldots, d$, is obtained from the variables X and Y. These are then used in the design, estimation, or both stages. For these auxiliary variables, a superpopulation model [1, 2] can be used at the estimation stage of inference. It should be noted that all these methods are based on simple statistical

models that describe the underlying relationships between the survey and auxiliary variables (linear regression models). Hansen [3] showed that under the parametric superpopulation, model misspecification can lead to substantial errors in inference. To solve this problem, nonparametric regression involving robust estimators in finite population sampling has been proposed [4–6].

When applying nonparametric kernel-based regression estimators over a finite range in estimating finite population parameters, one of the most common problems that is encountered is the bias at the edges [7]. It is also known that kernel and polynomial regression estimators provide good estimates for the population totals when $x \in \mathbb{R}^d$ and $d = 1$ [6, 8].

Despite the fact that high-dimensional auxiliary knowledge can be accounted for in the aforesaid estimators, the problem of sparse regressors in the design space makes kernel methods and local polynomials unfeasible, as performance deteriorates significantly as the dimension increases [8–10]. The reason behind this poor performance is due to the curse of dimensionality. The "curse of

dimensionality" is a phenomenon induced by the sparsity of data in high-dimensional spaces leading to a decrease in the fastest attainable rates of convergence of regression function estimators towards their target curve as the dimension of the regressor vector grows. Friedman [11] provided an overview of the concept of the curse of dimensionality.

Given the challenge of the curse of dimensionality, one has to use different nonparametric estimators to retain a large degree of flexibility. Using recursive covering in model-based approach [12] and generalized additive modelling in a model-assisted framework [13] is one way to get around this curse of dimensionality when dealing with multivariate auxiliary information. These estimation methods come at a cost of reduced flexibility with the associated risk of increased bias [9–11, 14].

In this regard, a robust nonparametric estimator of finite population total based on a feedforward backpropagation neural network method is proposed in this paper to help in resolving the failures of previously identified estimation approaches. Despite the fact that kernel and local approximators have the same property as artificial neural networks (ANNs), they usually require a large number of components to achieve similar approximation accuracy [15]. As a consequence, ANNs are regarded as an efficient method of performing parametric and nonparametric functional analysis.

2. Neural Network Estimator of Finite Population Total

In describing this estimator, the procedure provided in [16] is followed. Let Y be the survey variable associated with an auxiliary variable X assumed to follow a superpopulation model under a model-based approach. A commonly used working model for the finite population is

$$y_i = m(x_i) + \varepsilon_i, \qquad (2)$$

such that $x_{ij} \in \mathbb{R}^d$, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ i.i.d with mean zero and x_{ij}, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, d$ are considered as the auxiliary information.

Also let

$$T = \sum_{i \in s} y_i + \sum_{i \in r} y_i, \qquad (3)$$

be the finite population total where s are the sample units and r are the nonsampled units. Assume that y_i is given according to equation (2) with $x_i \in \mathbb{R}^d$, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ i.i.d. Consider estimating $m(x)$ based on a feedforward backpropagation neural network. The neurons which act as the basic building blocks can be considered as a nonlinear transformation of the input variables $x = (x_1, \ldots, x_d)\prime$.

Feedforward neural network that has least one layer of hidden units is considered to be a complex network and allow for information feedback can be specified. Without loss of generality, the paper will only concentrate on the structure presented in equation (4), which is commonly used

for a wide range of applications and has appealing features of being implemented in statistical softwares.

In the simplest case of one hidden layer with $H \geq 1$ neurons, the network can be written to represent the network function as follows:

$$f_H(x, \theta) = v_0 + \sum_{h=1}^{H} v_h \psi\left(w_{0h} + x^T w_h\right), \ x \in \mathbb{R}^d, \qquad (4)$$

with $w_h = (w_{1h}, \ldots, w_{dh}) \in \mathbb{R}^d$ and

$$\theta = \left(w_{01}, \ldots, w_{0H}, w_1^T, \ldots, w_H^T, v_0, \ldots, vh\right)^T \in \mathbb{R}^{M(H)}, \qquad (5)$$

where $M(H) = (d + 1)H + H + 1$ represents the vector of all parameters of weights of the network. $\psi: \mathbb{R} \mapsto \mathbb{R}$ is a given activation function. Regarding regression issues, sigmoid functions that resemble the distribution function of a genuine random variable, for example, typically produce good results. The logistic sigmoid and the bipolar sigmoid are two extensively used sigmoid functions that can be employed depending on the needed output. Whenever the goal is to approximate functions that map into probability space, the logistic function is preferred. The activation function is viewed as a smooth equivalent of the indicator function when the input signals are "squashed" between zero and one. As an illustration of the logistic function, consider the following:

$$\psi(u) = \frac{1}{1 + \exp(-u)}, \quad -\infty < u < \infty, \qquad (6)$$

which tends to one (zero) since its arguments approach infinity (negative infinity). As a result, based on the received input signals, the logistic activation function creates partially on/off signals.

For this work, $f_H(x; \theta)$ specifies a one-dimensional mapping from the input space $\mathbb{R}^d$ to the output space. $O = \left\{f_H(x; \theta), \theta \in \mathbb{R}^{M(H)}, H \geq 1\right\}$; for each continuous function m, any $\varepsilon > 0$, and any compact set $C \subseteq \mathbb{R}^d$, there exists a function $f_H \in O$ with uniform approximation qualities [17–19], for example,

$$\sup_{x \in C} \left| m(x) - f_H(x; \theta) \right| < \varepsilon. \qquad (7)$$

This suggests that any regression function $m(x)$ can be well estimated with a big sufficient number of neurons and the right parameters θ.

Therefore, a nonparametric estimate for $m(x)$ is obtained by first choosing H, which serves as a tuning parameter and determines the smoothness of the estimate. The parameter θ is estimated from the data by nonlinear least squares:

$$\widehat{\theta}_n = \arg \min_{\theta \in \mathfrak{R}^{M(H)}} D_n(\theta), \qquad (8)$$

with

$$D_n(\theta) = \sum_s \left(y_i - f_H(x; \theta)\right)^2. \qquad (9)$$

Under the right circumstances, $\widehat{\theta}_n$ converges in probability as $n \longrightarrow \infty$ and H, constant to the parameter vector

$\theta \in \Theta_H$ which equates to the best approximation of $m(x)$ by a function of type $f_H(x; \theta)$, $\theta \in \Theta_H$ with

$$\theta = \arg \min_{\theta \in \mathfrak{R}^{M(H)}} D(\theta) \text{ with } D(\theta) = E\{m(x) - f_H(x; \theta)\}. \tag{10}$$

Also, under some stronger assumptions, the asymptotic normality of $\widehat{\theta}_n$ and thus the estimator of $\widehat{m}(x) = f_H(x; \widehat{\theta}_n)$ also follow the regression function $m(x)$. Therefore, the immediate consequence of these is that $f_H(x; \widehat{\theta}_n) \longrightarrow f_H(x; \theta)$ for $n \longrightarrow \infty$.

The estimation error $\widehat{\theta}_n - \theta$ can be broken down into two asymptotically independent parts: $\widehat{\theta}_n - \theta = (\widehat{\theta}_n - \theta_n) + (\theta_n - \theta)$, where the value

$$\theta_n = \arg \min_{\theta \in \mathfrak{R}^{M(H)}} \sum_{i=1}^{n} \{m(x) - f_H(x, \theta)\}^2, \tag{11}$$

minimizes the sample version of $D(\theta)$ [20]. $f_H(x; \theta)$ converges to the regression function $m(x)$ for $H \longrightarrow \infty$, due to the universal approximation property of neural networks. As H grows with n at an adequate rate, $f_H(x; \widehat{\theta}_n)$ becomes a consistent nonparametric estimator of $m(x)$. As a result of these findings, Were and Orwa [16] showed that the corresponding estimate of the finite population total is as follows:

$$\widehat{T}_{\mathrm{NN}} = \sum_{j \in s} y_j + \sum_{j \in r} \widehat{m}_n(x_j), \tag{12}$$

which is the proposed estimator for the finite population total where $\widehat{m}_n(x_j) = f_H(x; \widehat{\theta}_n)$.

As noted in [16], T_{NN} is a model-based estimator, so that all the inference is with respect to the model for the $y_i's$, not the survey design. This estimator is identical to that proposed in [5], except that the NN is replaced by a kernel-based regression. Lastly, this estimator can be used to estimate the population totals of a finite population as long as each of the unsampled elements has the same distribution as the sample.

It should be noted that

(1). Where certain conditions are satisfied and if the activation function $\psi(u)$ is Lipschitz continuous and strictly increasing, then it can be shown that the neural network estimate $\widehat{T}_{\mathrm{NN}}$ of the population total T given by (12) with $\widehat{m}_n(x) = f(x, \widehat{\theta}_n)$ and $\widehat{\theta}_n$ given by (8) is consistent in the following sense.

$$\frac{1}{N} |T - \widehat{T}_{\mathrm{NN}}| \longrightarrow 0 \text{ in probability}, \tag{13}$$

where $N, n \longrightarrow \infty$ with $(n/N) \longrightarrow \pi \in (0, 1)$,

provided that the number H_n and the bound Δ_n of the network weights satisfy $H_n, \Delta_n \longrightarrow \infty$ such that

$$\Delta_n = o(n^{1/4}),$$
$$H_n \Delta_n^4 \log n = o(n), \tag{14}$$
$$H_n \log n = o(\Delta_n^\alpha),$$

where α determines how fast the tail probability of the ε_i and y_i decreases. White [19] showed that the appropriate choice for Δ_n is such that $\Delta_n \longrightarrow \infty$ as $n \longrightarrow \infty$ and $\Delta_n = o(n^{1/4})$, i.e., $n^{1/4} \Delta_n \longrightarrow 0$ as $n \longrightarrow \infty$.

(2). Where certain conditions are met, it can be shown that the mean squared error defined by $E(T_{\mathrm{NN}} - T)^2$ where T denotes the true population total of the proposed estimators reduces to

$$\widehat{E}(T_{\mathrm{NN}} - T)^2 = \frac{1 - f}{n} \widehat{\mathrm{var}}(\varepsilon_i), \tag{15}$$

where estimate of $\mathrm{var}(\varepsilon_i)$ is given as

$$\widehat{\mathrm{var}}(\varepsilon_i) = \frac{v}{v - 2} \frac{1}{H} \sum_{i=1}^{H} \widehat{\sigma}_i^2. \tag{16}$$

For the details and complete proof of these properties, see [16].

3. Coverage Properties

In order to compute and understand the coverage properties of the proposed estimator and how it is compared against other existing nonparametric regression estimators, the proposed estimator's performance is compared to that of identified estimators: multivariate adaptive regression splines (MARS), generalized additive model (GAM), and local polynomial (LP), which can handle high-dimensional data through a simulation study.

Scenarios where the true function is the sum of two-dimensional linear function, two-dimensional quadratic function, and three-dimensional mixed function given below are considered:

2-dim linear: $m(x) = -1 + 2X_3 + 4X_6$.

2-dim quadratic: $m(x) = 5.5 - 6X_2 + 8(X_2 - 0.5)^2 - 3X_{10} + 32(X_{10} - 0.5)^3$.

3-dim mixed model: $m(x) = 8(X_2 - 0.5)^2 + \exp(2X_5 - 1) + \sin(2\pi(X_8 - 0.5))$.

For all of the simulation performed, data are generated according to model 2 where $\varepsilon \sim N(0, 1)$. The auxiliary variable vector $X \in \mathbb{R}^d$ was generated from *iid* uniform $(0,1)$ random vector. The errors ϵ were generated from i.i.d $N(0, 1)$ with noise level $\sigma = 0.1, 0.4$. tanh is used as the activation function in this neural network.

1000 samples of sizes $4,000$ and $8,000$ were generated using simple random sampling from a population of size $10,000$. Because of the hypothesized relationship between the study variable and the auxiliary variable, which must be depicted in the simulation, the sampling is done with indices.

Tables 1–3 summarize the findings of this simulation investigation. The unconditional bias (UB), unconditional mean square error (UMSE), unconditional relative mean square error (URMSE), and unconditional mean absolute error (UMAE) for said estimators at different sample sizes are shown in Tables 1–3. The MAE reveals how near the

TABLE 1: UBIAS, UMSE, URMSE, and UMAE for a two-dimensional linear model.

		Bias	MSE	RRMSE	MAE
$n = 4000$	$\widehat{T}_{\mathrm{NN}}$	8.7982	151.4639	0.1900	0.0043
	$\widehat{T}_{\mathrm{MARS}}$	9.8620	153.1423	0.2170	0.0048
	$\widehat{T}_{\mathrm{GAM}}$	9.8700	152.9656	0.2172	0.0048
	$\widehat{T}_{\mathrm{LP}}$	10.0203	156.9519	0.2205	0.0049
$n = 8000$	$\widehat{T}_{\mathrm{NN}}$	3.7104	20.7253	0.060	0.0011
	$\widehat{T}_{\mathrm{MARS}}$	4.3909	29.5419	0.0779	0.0014
	$\widehat{T}_{\mathrm{GAM}}$	10.4348	30.3327	0.4512	0.0041
	$\widehat{T}_{\mathrm{LP}}$	13.1195	40.3936	0.2327	0.0080

TABLE 2: UBIAS, UMSE, URMSE, and UMAE for a two-dimensional quadratic model.

		Bias	MSE	RMSE	MAE
$n = 4000$	$\widehat{T}_{\mathrm{NN}}$	3.3743	20.5077	0.0596	0.0011
	$\widehat{T}_{\mathrm{MARS}}$	6.8289	76.5408	0.1206	0.0021
	$\widehat{T}_{\mathrm{GAM}}$	20.0105	643.8682	0.3534	0.0062
	$\widehat{T}_{\mathrm{LP}}$	18.1960	536.6546	0.3213	0.0057
$n = 8000$	$\widehat{T}_{\mathrm{NN}}$	1.9396	12.5319	0.0343	0.0006
	$\widehat{T}_{\mathrm{MARS}}$	4.0274	25.4001	0.0711	0.0013
	$\widehat{T}_{\mathrm{GAM}}$	12.4017	246.0122	0.2190	0.0039
	$\widehat{T}_{\mathrm{LP}}$	11.3112	200.8425	0.1998	0.0035

TABLE 3: UBIAS, UMSE, URMSE, and UMAE for a three-dimensional mixed model.

		Bias	MSE	RMSE	MAE
$n = 4000$	$\widehat{T}_{\mathrm{NN}}$	3.5196	18.9278	0.0583	0.0010
	$\widehat{T}_{\mathrm{MARS}}$	5.7422	52.2492	0.0951	0.0016
	$\widehat{T}_{\mathrm{GAM}}$	14.7975	353.5178	0.2450	0.0041
	$\widehat{T}_{\mathrm{LP}}$	16.8233	437.2852	0.2785	0.0046
$n = 8000$	$\widehat{T}_{\mathrm{NN}}$	1.8147	5.4731	0.0300	0.0005
	$\widehat{T}_{\mathrm{MARS}}$	3.3823	18.3560	0.0560	0.0009
	$\widehat{T}_{\mathrm{GAM}}$	8.8086	122.7989	0.1458	0.0024
	$\widehat{T}_{\mathrm{LP}}$	9.9900	151.4552	0.1654	0.0027

estimate being examined is to the true value, while the MSE and RMSE represent the estimator's precision. For example, if TNN's UMSE and URMSE are comparable, it will reasonably be considered "better" or "more desirable" than other estimators.

The deviation of the estimator's expected value from the true total value is known as the bias of a population total estimator. All of the estimators of the finite population total discussed here are biased, but $\widehat{T}_{\mathrm{NN}}$ is the least biased. $\widehat{T}_{\mathrm{NN}}$ can be seen as most efficient estimator of finite population total in all models and sample sizes, closely followed by $\widehat{T}_{\mathrm{MARS}}$. Because of their relatively large bias values, the generalized additive estimator and the local polynomial regression estimator overestimate the finite population total under all models.

In addition, $\widehat{T}_{\mathrm{NN}}$ has lower mean square error, relative mean square error, and mean absolute error which is followed closed with the estimator $\widehat{T}_{\mathrm{MARS}}$. It is also observed that as the sample increases, all the estimators recorded a significant improvement in their performance in estimating the finite population totals. The local polynomial regression estimator with a significant reduction in bias and mean square error is noteworthy. This follows the argument by Stone [10]: to improve the efficiency of the local smoother in high-dimensional spaces, one has to use a large sample size. The neural network estimator still outperforms other estimators with significant reduction in biases, mean square errors, relative root mean square errors, mean absolute errors, and mean absolute percentage errors as sample sizes increases.

The results provided in Table 3 which provides the results for performance of the estimators for a three-dimensional mixed model are noteworthy. Compared to the two-dimensional case, the performance of all the estimators has marginally decreased as indicated by marginal increase in biases, mean

square errors, relative mean square errors, and mean absolute errors across all the estimators of finite population total. It is also observed that the generalized additive estimator and local polynomial regression still recorded poor performance in terms of biases, mean square errors, relative mean square errors, and mean absolute errors in estimating the finite population total. In the other case, $\widehat{T}_{\mathrm{NN}}$ has lower biases, mean square errors, relative root mean square errors, mean absolute errors, and mean absolute percentage errors which is followed closely by the estimator T_{MARS}.

Even with the increasing sample size, all the estimators record a significant improvement in their performance in estimating the finite population totals. For instance, a local polynomial regression estimator is noted to have a significance reduction in bias and mean square errors as the sample size increases. The neural network estimator still remains the estimator of choice compared to other estimators as sample sizes increases.

The estimator's conditional performance was assessed and compared to that of other finite population total estimators in high-dimensional space that have been identified. To do this, 1000 simple random samples were sorted using the sample means of $\overline{X}_s$ value criterion. The samples were then grouped into sets of twenty samples such that the first set is made of samples with the lowest sample means of $\overline{X}_s$ values, the second set consists of samples with means of $\overline{X}_s$ that are larger than the sample means of the first set, and so on until the last set that consists of samples with the largest sample means of $\overline{X}_s$ values. In each of the group, the bias, mean square error, relative mean square error, and mean absolute error were computed.

The results of group conditional bias (CB), conditional mean square error (CMSE), conditional relative mean square error (CRMSE), and conditional mean absolute error (CMAE) for the finite population total estimators $\widehat{T}_{\mathrm{NN}}$, $\widehat{T}_{\mathrm{MARS}}$, $\widehat{T}_{\mathrm{GAM}}$, and $\widehat{T}_{\mathrm{LP}}$ are plotted against group average values $\overline{\overline{X}}$ denoted as Xbar in the fifty groups of mean of $\overline{X}_s$.

The conditional findings for the estimators under the two-dimensional linear model, two-dimensional quadratic model, and three-dimensional mixed model are shown in Figures 1–3. The bias characteristics of the numerous estimators differ significantly in the majority of circumstances. A closer look at the plots reveals that $\widehat{T}_{\mathrm{NN}}$ and $\widehat{T}_{\mathrm{MARS}}$ have lower levels of bias overall, as seen by the displayed curves' proximity to the horizontal (no bias) line at 0.0 on the vertical axis. Consequently, despite the complex structure of
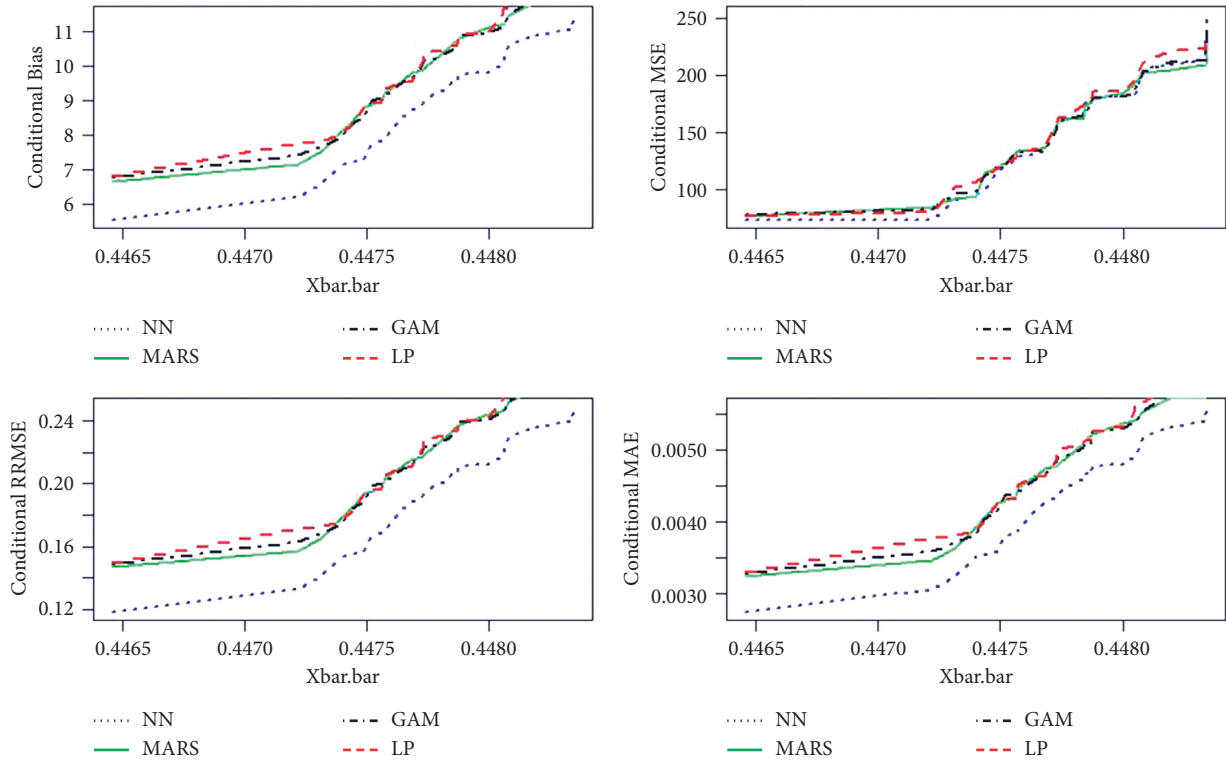
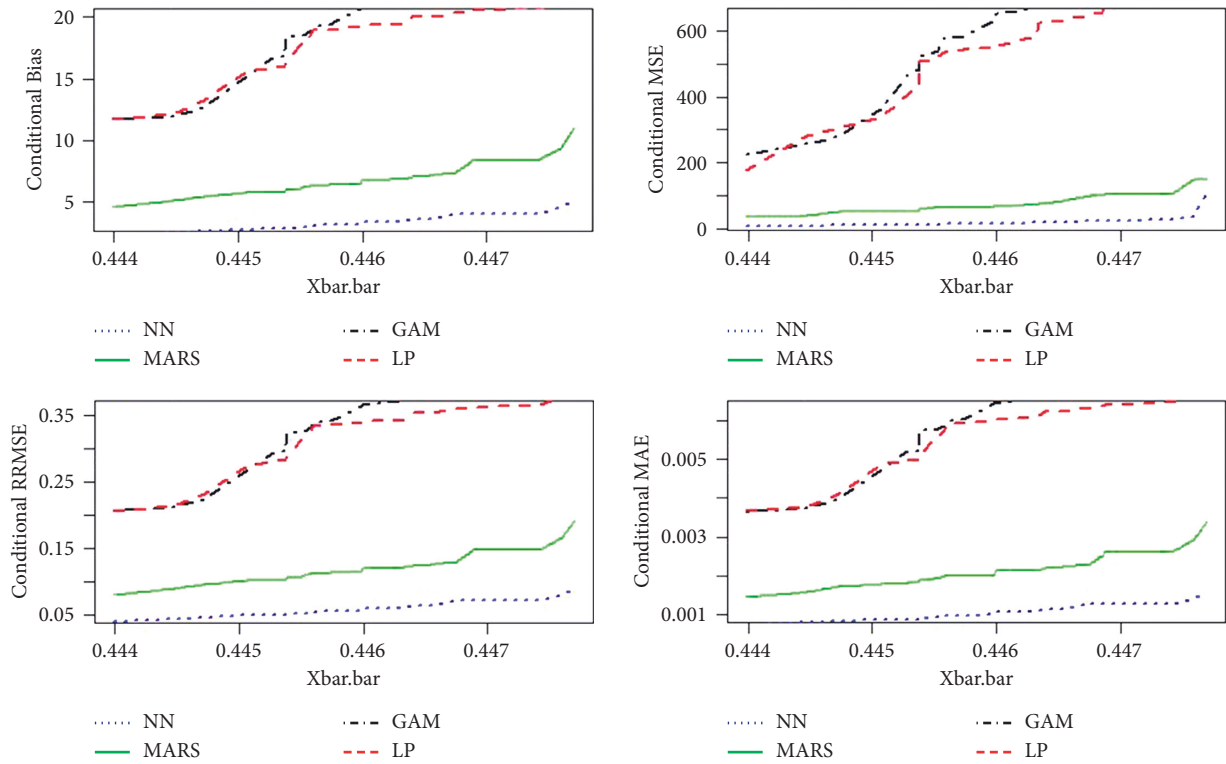Figure 1: CB, CMSE, CRMSE, and CMAE based on a two-dimensional linear model.

Figure 2: CB, CMSE, CRMSE, and CMAE based on a two-dimensional quadratic model.

some of the plots, estimator $\widehat{T}_{NN}$ emerges as the least biased for practically every set of auxiliary variable means and distinct models.

Similarly, plotting conditional MSE vs. group means of auxiliary variable means reveals that the estimators behave in a similar way. The lowest MSE values are produced by $\widehat{T}_{NN}$
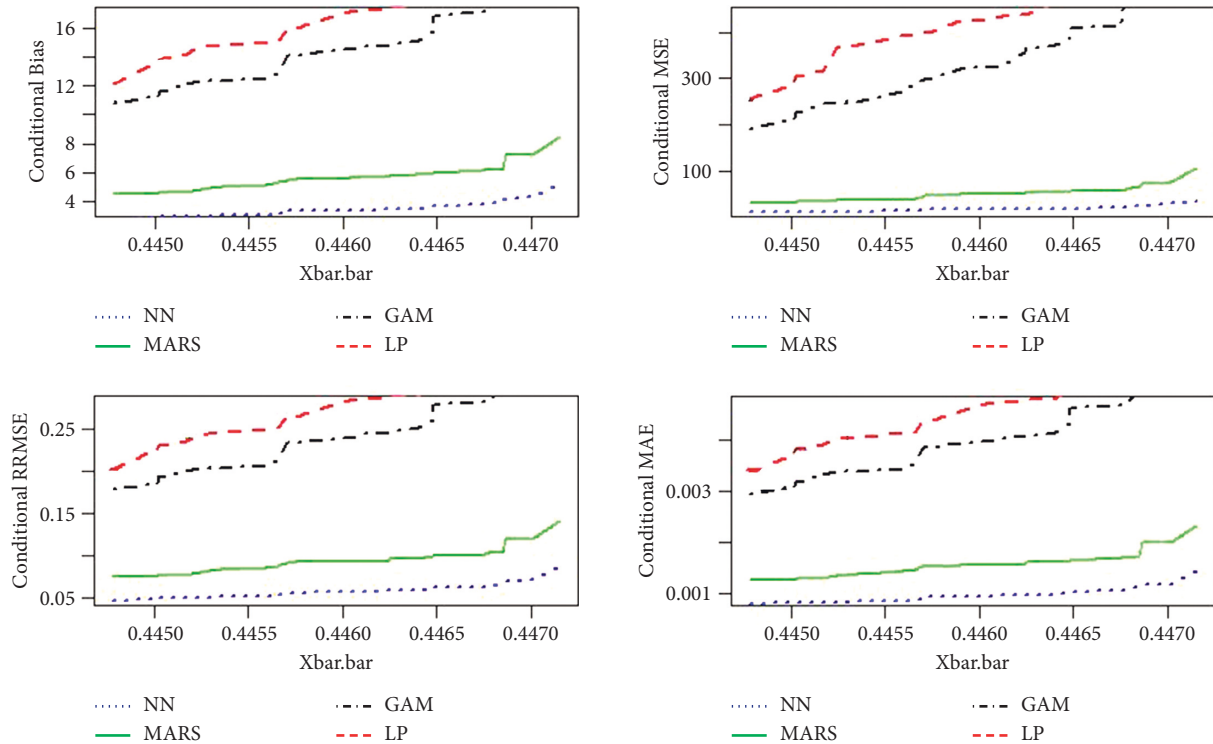
Figure 3: CB, CMSE, CRMSE, and CMAE based on a three-dimensional mixed model.

and $\widehat{T}_{\text{MARS}}$. $\widehat{T}_{\text{NN}}$, for instance, has the lowest MSE of any of the other estimators in the majority of circumstances. For bias, MSE, and MAE, $\widehat{T}_{\text{NN}}$ consistently outperforms all other estimators.

4. Conclusion

In this paper, the coverage properties of an estimator for finite population total based on a feedforward back-propagation neural network technique in nonparametric regression have been studied. The properties such as the bias, mean squared error, and mean absolute error have been computed for the case of high-dimensional datasets through a simulation, and the findings were compared with those of existing estimators such as multivariate adaptive regression splines (MARS), generalized additive model (GAM), and local polynomial (LP) which can handle high-dimensional data.

From the results, the following observations and conclusions have been made:

(i) The neural network estimator estimates the finite population total better than all other robust estimators in high-dimensional case.

(ii) The performance of local polynomial estimator in the estimation of finite population becomes poor as the dimension of the data increases.

(iii) For all the estimators, as the sample sizes increases, biases, mean square errors, relative root mean square errors, mean absolute errors, and mean absolute percentage errors decrease for the four models considered.

(iv) For all the estimators, as the dimension increases, biases, mean square errors, relative root mean square errors, mean absolute errors, and mean absolute percentage errors decrease for all the four models considered.

To this end, the main conclusion is that the estimator of finite population total based on the feedforward back-propagation neural network has proved to yield results with great precision, and therefore it is recommended for estimating finite population total. It should be noted that the proposed estimator has been considered in case of simple random sampling without replacement (SRSWoR). An extension to other sampling techniques such as stratification may be done since they rely on SRSWoR, and it is hypothesized that efficiency will be improved compared to other existing estimators in the literature.

Data Availability

The data used are artificial data from simulation process using a specified model.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] R. L. Chambers and R. Dunstan, "Estimating distribution functions from survey data," *Biometrika*, vol. 73, no. 3, pp. 597–604, 1986.

[2] S. Wang and A. H Dorfman, "A new estimator for the finite population distribution function," *Biometrika*, vol. 83, no. 3, pp. 639–652, 1996.

[3] M. H. Hansen, "Some history and reminiscences on survey sampling," *Statistical Science*, vol. 2, no. 2, pp. 180–190, 1987.

[4] F. J. Breidt and J. D. Opsomer, "Local polynomial regression estimators in survey sampling," *Annals of Statistics*, vol. 28, no. 4, pp. 1026–1053, 2000.

[5] A. H Dorfman, "Nonparametric regression for estimating totals in finite populations," in *Proceedings of the Section on Survey Research Methods*, pp. 622–625, American Statistical Association Alexandria, Alexandria, VA, USA, 1992.

[6] R. Odhiambo Otieno and T. Mbithi Mwalili, "Nonparametric Regression Method for Estimating the Error Variance in Unistage Sampling," *East African Journal of Science*, vol. 2, no. 2, pp. 107–112, 2002.

[7] R. L. Chambers, A. H. Dorfman, and P. Hall, "Properties of estimators of the finite population distribution function," *Biometrika*, vol. 79, no. 3, pp. 577–582, 1992.

[8] G. E Montanari and M. G. Ranalli, "On calibration methods for design based finite population inferences," *Bulletin of the International Statistical Institute*, vol. 54, p. 60, 2003.

[9] P. J. Bickel and B. Li, "Local polynomial regression on unknown manifolds," *Lecture Notes-Monograph Series*, vol. 54, pp. 177–186, 2007.

[10] C. J. Stone, "Optimal global rates of convergence for nonparametric regression," *Annals of Statistics*, vol. 10, no. 4, pp. 1040–1053, 1982.

[11] J. H. Friedman, "Multivariate adaptive regression splines," *Annals of Statistics*, vol. 19, no. 1, 1991.

[12] A. Di Ciaccio and M. Ge, "A nonparametric regression estimator of a finite population mean," *Book of Short Papers*, CLADAG, pp. 173–176, 2001.

[13] J. D. Opsomer, F. J. Breidt, G. G. Moisen, and G. . Kauermann, "Model-assisted estimation of forest resources with generalized additive models," *Journal of the American Statistical Association*, vol. 102, no. 478, pp. 400–409, 2007.

[14] E. H. A. Rady and D. Ziedan, "Estimation of population total using local polynomial regression with two auxiliary variables," *Journal of Statistics Applications & Probability*, vol. 3, no. 2, pp. 129–136, 2014.

[15] A. R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 930–945, 1993.

[16] F. Were, O. Orwa, and R. Odhiambo, "Estimation of finite population totals in high dimensional datasets," in *Proceedings of the Analytical Methods to Model Nature, Jomo Kenyatta University of Agriculture and Technology*, Juja, Kenya, 2022.

[17] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.

[18] K. I. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, vol. 2, no. 3, pp. 183–192, 1989.

[19] H. White, "Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings," *Neural Networks*, vol. 3, no. 5, pp. 535–549, 1990.

[20] J. . Franke and M. H. Neumann, "Bootstrapping neural networks," *Neural Computation*, vol. 12, no. 8, pp. 1929–1949, 2000.