MDPI

Neutrosophic Mean Estimation of Sensitive and Non-Sensitive Variables with Robust Hartley–Ross-Type Estimators

Abdullah Mohammed Alomair [1,*] and Usman Shahzad [2,3]

1   Department of Quantitative Methods, School of Business, King Faisal University, Al-Ahsa 31982, Saudi Arabia
2   Department of Mathematics and Statistics, International Islamic University, Islamabad 44000, Pakistan;
    usman.stat@yahoo.com
3   Department of Mathematics and Statistics, PMAS-Arid Agriculture University Rawalpindi,
    Rawalpindi 46300, Pakistan
*   Correspondence: ama.alomair@kfu.edu.sa

Abstract: Under classical statistics, research typically relies on precise data to estimate the population mean when auxiliary information is available. Outliers can pose a significant challenge in this process. The ultimate goal is to determine the most accurate estimates of the population mean while minimizing variance. Neutrosophic statistics is a generalization of classical statistics that deals with imprecise, uncertain data. Our research introduces the neutrosophic Hartley–Ross-type ratio estimators for estimating the population mean of neutrosophic data, even in the presence of outliers. We also incorporate neutrosophic versions of several robust regression methods, including LAD, Huber-M, Hampel-M, and Tukey-M. Our approach assumes that the study variable is both non-sensitive and sensitive, meaning that it can cause discomfort to participants during personal interviews, and measurement errors can occur due to dishonest responses. To address potential measurement errors, we propose the use of neutrosophic scrambling response models. Our proposed neutrosophic robust estimators are more effective than existing classical estimators, as confirmed by a computer-based numerical study using real data and simulation.

1. Introduction

Precise numerical data are at the heart of classical statistics, and various researchers have developed estimators to calculate the mean of a finite population using auxiliary information. When there is high correlation between the study variable and auxiliary variable, using a ratio estimation method instead of only considering the study variable can significantly reduce the sampling error. This results in a smaller required sample size while maintaining precision, as noted by Cochran [1]. Ratio estimation techniques have been extensively researched, with different types and uses developed over time. Researchers have explored various transformations of known parameters and statistics as auxiliary variables. Recent studies have shown that utilizing diverse types of auxiliary information can improve the performance of ratio-type estimators. For example, some scholars have suggested using exponential-type ratio estimators and refining their performance. Others have investigated the estimation of the mean through exponential ratio-type estimators in the presence of non-response. One study proposed an estimator that utilizes complete information, which outperforms exponential ratio-type estimators. Additionally, a study has examined the implementation of a ratio-type estimator for multivariate k-statistics and explored the use of auxiliary information with the coefficient of variation. These advancements were discussed by Robson [2] in his research. Tahir et al. [3] developed neutrosophic

ratio estimators in simple random sampling. Vishwakarma and Singh [4] extended their work in ranked set sampling scheme. Yadav and Smarandache [5] and Kumar et al. [6] defined generalized families of neutrosophic ratio and exponential estimators.

While classical statistics assumes precise data, fuzzy logic provides a solution for data that may not have exact measurements. Fuzzy statistics is a useful tool for analyzing data with fuzzy, ambiguous, uncertain, or imprecise parameters or observations. However, it does not account for the degree of indeterminacy in the data. Neutrosophic logic is an extension of fuzzy logic that enables the measurement of both the determinate and indeterminate parts of the observations. It is used to analyze data with vague or uncertain observations, as noted by Smarandache [7].

When data have some degree of indeterminacy, neutrosophic statistics are used. This statistical methodology goes beyond the traditional approach and is employed in situations where the sample or data contains neutrosophy. Neutrosophic statistics are particularly useful when observations within the population or sample are ambiguous, uncertain, and indefinite, as explained by Smarandache [7].

1.1. Neutrosophic Statistics and Hartley–Ross-Type Estimators

Neutrosophic statistical methods are utilized to examine datasets that contain some level of uncertainty, also known as neutrosophic data. In this type of statistics, the sample size may not be accurately determined, as explained by Smarandache [8]. Smarandache's research [8] has shown that neutrosophic statistics are highly effective in analyzing systems of uncertainty. In the field of rock engineering, neutrosophic numbers have been used to investigate the scale effect and anisotropy of joint roughness coefficient. This has led to a more efficient method for overcoming information loss and generating adequately fitted functions, as demonstrated by Chen et al. [9]. Additionally, a new technique called neutrosophic analysis of variance has been introduced for analyzing neutrosophic data. The field of neutrosophic statistics is currently being advanced by exploring new areas such as neutrosophic interval statistics (NIS), neutrosophic applied statistics (NAS), and neutrosophic statistical quality control (NSQC).

Hartley and Ross [10], Robson [2], Murty [11], and Smoo et al. [12] have studied various unbiased estimators for population mean. Hartley and Ross [10] have devised new ratio-type estimators for estimating the population mean, and their work has been further enhanced by other survey statisticians. When the variables have a negative correlation, Singh and Singh [13] have proposed unbiased Hartley–Ross estimators for the population mean. Additionally, Singh et al. [14] have developed modified ratio-type estimators based on Hartley and Ross estimators, incorporating additional information such as coefficient of variation, correlation, and more. Kadilar and Cekim [15] have drawn inspiration from Hartley–Ross's work and developed the Hartley–Ross-type regression estimator that utilizes auxiliary information. In this article, we will investigate the application of neutrosophic OLS and robust regression coefficients in Hartley–Ross-type neutrosophic mean estimators.

1.2. Research Gap

Previous research in survey sampling has primarily focused on data that are precise, certain, and unambiguous. However, such methods may generate a single, clear-cut result that could be inaccurate, overestimated, or underestimated, which poses a limitation in certain cases. Conversely, there are situations where data are of a neutrosophic nature, and in such cases, classical statistical methods are inadequate. Data with a neutrosophic nature are often characterized by uncertain and ambiguous observations, non-clear arguments, and vague interval values. Therefore, data collected from experiments or populations can be expressed as interval-valued neutrosophic numbers (INN), where the observation is assumed to fall within the boundaries of the given interval. In reality, indeterminate data is more common than determinate data, making it necessary to develop more neutrosophic statistical techniques to analyze such data.

Gathering data on numerous variables throughout life can be a costly endeavor, especially when the data are uncertain. Therefore, relying on traditional classical methods to determine the unknown true value of the population for ambiguous data can be both risky and expensive. Furthermore, if both the primary study variable and auxiliary variables have a neutrosophic nature, conventional Hartley–Ross-type estimation is inadequate. As a result, this study suggests the use of neutrosophic Hartley–Ross-type regression estimators.

A comprehensive examination of published studies [3–6] reveals that no research has been conducted in the domain of survey sampling to estimate an unknown population mean using Hartley–Ross-type regression estimation methods while auxiliary variables are present under neutrosophic data containing both sensitive and non-sensitive observations. This specific field of statistics necessitates further investigation and study. The present research serves as an introductory step in this area.

1.3. Scope of the Study

Neutrosophic statistical analysis is an approach that is capable of handling data with incomplete or indeterminate information, while also accommodating inconsistent beliefs. In some instances, when data are gathered using certain instruments, observations may exist within an uncertain range, with the possibility that the true measurement lies within that interval. In such cases of indeterminacy, classical statistical methods may not be adequate for analyzing the data. As an alternative, the method of neutrosophic statistics is utilized as a more flexible and general version of classical statistics. While numerous studies in the domain of survey sampling have been conducted under the framework of neutrosophy, the area of Hartley–Ross-type estimation remains a relatively new and underexplored field that warrants further attention.

To start, we assume that the neutrosophic study variable is non-sensitive, which makes the conventional estimation method appropriate. However, we also recognize the existence of sensitive variables that are highly personal, stigmatizing, or threatening, and they are likely to be observed in sampled units using non-standard survey techniques to enhance respondent cooperation. These techniques have been derived from the randomized response theory introduced by Warner [16] and have been extensively discussed in the works of Fox and Tracy [17], Chaudhuri and Mukerjee [18], and Chaudhuri [19]. Furthermore, we perform a numerical investigation on the sensitive variable to assess the performance of the proposed method by manipulating the values of the study variable using various randomized response models.

This article presents a fresh perspective on mean estimation in the context of neutrosophic data, which incorporates uncertainty not only in the form of imprecision and vagueness but also indeterminacy. The remaining sections of the article are organized as follows: Section 2 introduces the neutrosophic versions of the OLS-based and regression-based Hartley–Ross-type mean estimators. In Section 3, a new class of Hartley–Ross-type neutrosophic robust regression estimators is proposed. The usefulness of these methods in sensitive research is discussed in Section 4. A numerical example is provided in Section 5 to demonstrate the effectiveness of the proposed estimators. Finally, in Section 6, the article concludes by highlighting the potential of neutrosophic robust regression methods in addressing the challenges posed by the complex and uncertain nature of real-world data.

2. Adapted OLS Based Neutrosophic Hartley–Ross Type Estimators

Parametric regression methods are widely used in statistics, with ordinary least squares (OLS) being one of the most commonly used techniques. OLS aims to estimate the model's parameters while minimizing the sum of residual squares (r_i), and is known for its mathematical simplicity and elegance. The Gauss–Markov theorem requires several assumptions that must be satisfied for OLS to be considered the most suitable estimator for linear regression coefficients. OLS estimators are the best linear unbiased estimator (BLUE) because they have the smallest variance among all unbiased estimators. Readers who are interested in this topic can refer to the work of Al-Noor and Mohammad [20].

We consider a neutrosophic variable, denoted by $Y_{e_1,e_2} \in [Y_{e_1}, Y_{e_2}]$, on a finite population U_{e_1,e_2} consisting of N identifiable units. To estimate the unknown mean of Y_{e_1,e_2}, denoted by $\bar{Y}_{e_1,e_2} \in [\bar{Y}e_1, \bar{Y}e_2]$, a neutrosophic auxiliary variable $X_{e_1,e_2} \in [X_{e_1}, X_{e_2}]$ is used. Note that additional information may be available for the entire population through knowledge of X_{e_1,e_2} or $\bar{X}_{e_1,e_2} \in [\bar{X}_{e_1}, \bar{X}_{e_2}]$, or on a sample from U_{e_1,e_2} in the absence of population information related to study variable. By adapting Hartley and Ross [10], Singh and Singh [13], and Kadilar and Cekim [15], the class of OLS regression coefficient $(b_{xy_{e_1,e_2}})$ based neutrosophic Hartley–Ross-type sample mean estimators for the study variable:

$$T_{je_1,e_2} = \bar{X}_{je_1,e_2}\left[\bar{r}_{xy_{je_1,e_2}} + \frac{b_{xy_{e_1,e_2}}(\bar{X}_{e_1,e_2} - \bar{x}_{e_1,e_2})}{\bar{x}_{je_1,e_2}}\right]; \; For \; j = 1, 2, ..., 8. \tag{1}$$

where
$$\bar{r}_{xy_{je_1,e_2}} = \frac{\sum_{k=1}^{n} r_{k.xy_{je_1,e_2}}}{n}, \; r_{k.xy_{je_1,e_2}} = \frac{\sum_{k=1}^{n} y_{k_{je_1,e_2}}}{x_{k_{je_1,e_2}}}$$
$$x_{k_{1e_1,e_2}} = x_{ke_1,e_2} + C_{xe_1,e_2} \text{ and } \bar{X}_{1e_1,e_2} = \bar{X}_{e_1,e_2} + C_{xe_1,e_2}$$
$$x_{k_{2e_1,e_2}} = x_{ke_1,e_2} + \beta_{2xe_1,e_2} \text{ and } \bar{X}_{2e_1,e_2} = \bar{X}_{e_1,e_2} + \beta_{2xe_1,e_2}$$
$$x_{k_{3e_1,e_2}} = x_{ke_1,e_2}\beta_{2xe_1,e_2} + C_{xe_1,e_2} \text{ and } \bar{X}_{3e_1,e_2} = \bar{X}_{e_1,e_2}\beta_{2xe_1,e_2} + C_{xe_1,e_2}$$
$$x_{k_{4e_1,e_2}} = x_{ke_1,e_2}C_{xe_1,e_2} + \beta_{2xe_1,e_2} \text{ and } \bar{X}_{4e_1,e_2} = \bar{X}_{e_1,e_2}C_{xe_1,e_2} + \beta_{2xe_1,e_2}$$
$$x_{k_{5e_1,e_2}} = x_{ke_1,e_2}C_{xe_1,e_2} + \rho_{e_1,e_2} \text{ and } \bar{X}_{5e_1,e_2} = \bar{X}_{e_1,e_2}C_{xe_1,e_2} + \rho_{e_1,e_2}$$
$$x_{k_{6e_1,e_2}} = x_{ke_1,e_2}\rho_{e_1,e_2} + C_{xe_1,e_2} \text{ and } \bar{X}_{6e_1,e_2} = \bar{X}_{e_1,e_2}\rho_{e_1,e_2} + C_{xe_1,e_2}$$
$$x_{k_{7e_1,e_2}} = x_{ke_1,e_2}\beta_{2xe_1,e_2} + \rho_{e_1,e_2} \text{ and } \bar{X}_{7e_1,e_2} = \bar{X}_{e_1,e_2}\beta_{2xe_1,e_2} + \rho_{e_1,e_2}$$
$$x_{k_{8e_1,e_2}} = x_{ke_1,e_2}\rho_{e_1,e_2} + \beta_{2xe_1,e_2} \text{ and } \bar{X}_{8e_1,e_2} = \bar{X}_{e_1,e_2}\rho_{e_1,e_2} + \beta_{2xe_1,e_2}$$

for each estimator. Where β_{2xe_1,e_2} is neutrosophic coefficient of kurtosis, C_{xe_1,e_2} is neutrosophic coefficient of variation, and ρ_{e_1,e_2} is neutrosophic coefficient of correlation. Further x_{ke_1,e_2} is representing sample observations of auxiliary variable. $\bar{x}_{e_1,e_2}$ is the sample mean of auxiliary variable. $\bar{x}_{je_1,e_2}$ is representing the mean of observations $x_{k_{1e_1,e_2}}, \ldots, x_{k_{8e_1,e_2}}$, for $j = 1, \ldots, 8$, respectively.

The biases of these estimators are.

$$B(T_{je_1,e_2}) = E\left\{\bar{X}_{je_1,e_2}\left[\bar{r}_{xy_{je_1,e_2}} + \frac{b_{xy_{e_1,e_2}}(\bar{X}_{e_1,e_2} - \bar{x}_{e_1,e_2})}{\bar{x}_{je_1,e_2}}\right] - \bar{Y}_{e_1,e_2}\right\} \tag{2}$$
$$= -\left(1 - \frac{1}{N}\right)S_{xr_{je_1,e_2}}; \; For \; j = 1, 2, ...8.$$

where
For $j = 1, 2$

$$S_{xr_{je_1,e_2}} = \left(\frac{1}{N-1}\right)\sum_{k=1}^{N}\left(x_{ke_1,e_2} - \bar{X}_{e_1,e_2}\right)\left(r_{k.xy_{je_1,e_2}} - \bar{R}_{xy_{je_1,e_2}}\right)$$

For $j = 3, 7$

$$S_{xr_{je_1,e_2}} = \left(\frac{\beta_{2xe_1,e_2}}{N-1}\right)\sum_{k=1}^{N}\left(x_{ke_1,e_2} - \bar{X}_{e_1,e_2}\right)\left(r_{k.xy_{je_1,e_2}} - \bar{R}_{xy_{je_1,e_2}}\right)$$

For $j = 4, 5$

$$S_{xr_{je_1,e_2}} = \left(\frac{C_{xe_1,e_2}}{N-1}\right)\sum_{k=1}^{N}\left(x_{ke_1,e_2} - \bar{X}_{e_1,e_2}\right)\left(r_{k.xy_{je_1,e_2}} - \bar{R}_{xy_{je_1,e_2}}\right)$$

For $j = 6, 8$

$$S_{xr_{j_{e_1,e_2}}} = \left(\frac{\rho_{e_1,e_2}}{N-1} \right) \sum_{k=1}^{N} \left(x_{k_{e_1,e_2}} - \bar{X}_{e_1,e_2} \right) \left(r_{k.xy_{j_{e_1,e_2}}} - \bar{R}_{xy_{j_{e_1,e_2}}} \right)$$

With

$$\bar{R}_{xy_{j_{e_1,e_2}}} = \frac{\sum_{k=1}^{N} r_{k.xy_{j_{e_1,e_2}}}}{N}.$$

Unbiased estimator of $S_{xr_{j_{e_1,e_2}}}$ are :

$$s_{xr_{j_{e_1,e_2}}} = \left(\frac{n}{n-1} \right) \left(\bar{y}_{e_1,e_2} - \bar{x}_{j_{e_1,e_2}} \bar{r}_{xy_{j_{e_1,e_2}}} \right); \; For \; j = 1, 2, \ldots, 8. \tag{3}$$

Substituting $s_{xr_{j_{e_1,e_2}}}$ instead of $S_{xr_{j_{e_1,e_2}}}$ in $B(T_{j_{e_1,e_2}})$, we have

$$B(T_{j_{e_1,e_2}}) = -\left(1 - \frac{1}{N} \right) \left(\frac{n}{n-1} \right) \left(\bar{y}_{e_1,e_2} - \bar{x}_{j_{e_1,e_2}} \bar{r}_{xy_{j_{e_1,e_2}}} \right); \; For \; j = 1, 2, \ldots, 8. \tag{4}$$

Using the expressions of $B(T_{j_{e_1,e_2}})$, unbiased version of $T_{j_{e_1,e_2}}$ is :

$$T^u_{j_{e_1,e_2}} = \bar{X}_{j_{e_1,e_2}} \left[\bar{r}_{xy_{j_{e_1,e_2}}} + \frac{b_{xy_{e_1,e_2}} \left(\bar{X}_{e_1,e_2} - \bar{x}_{e_1,e_2} \right)}{\bar{x}_{j_{e_1,e_2}}} \right] + \left(1 - \frac{1}{N} \right) \left(\frac{n}{n-1} \right) \left(\bar{y}_{e_1,e_2} - \bar{x}_{j_{e_1,e_2}} \bar{r}_{xy_{j_{e_1,e_2}}} \right); \tag{5}$$
$$For \; j = 1, 2, \ldots, 8.$$

The variance of $T^u_{j_{e_1,e_2}}$ is as follows:

$$V(T^u_{j_{e_1,e_2}}) = \theta \left[S^2_{y_{e_1,e_2}} + B^2_{e_1,e_2} S^2_{x_{e_1,e_2}} + \bar{R}^2_{j_{e_1,e_2}} S^2_{x_{j_{e_1,e_2}}} - 2B_{e_1,e_2} S_{yx} + 2B_{e_1,e_2} \bar{R}_{j_{e_1,e_2}} S_{xx_{j_{e_1,e_2}}} - 2\bar{R}_{j_{e_1,e_2}} S_{yx_{j_{e_1,e_2}}} \right]; \tag{6}$$
$$For \; j = 1, 2, \ldots, 8.$$

where $S^2_{y_{e_1,e_2}}, S^2_{x_{e_1,e_2}}$ is the neutrosophic variance of study and auxiliary variables. B_{e_1,e_2} is the OLS regression coefficient. Further

$$\theta = \left(\frac{1-f}{n} \right)$$

$$S^2_{y_{e_1,e_2}} = \left(\frac{1}{N-1} \right) \sum_{k=1}^{N} \left(y_{k_{e_1,e_2}} - \bar{Y}_{e_1,e_2} \right)^2$$

$$S^2_{x_{e_1,e_2}} = \left(\frac{1}{N-1} \right) \sum_{k=1}^{N} \left(x_{k_{e_1,e_2}} - \bar{X}_{e_1,e_2} \right)^2$$

$$S_{yx_{e_1,e_2}} = \left(\frac{1}{N-1} \right) \sum_{k=1}^{N} \left(y_{k_{e_1,e_2}} - \bar{Y}_{e_1,e_2} \right) \left(x_{k_{e_1,e_2}} - \bar{X}_{e_1,e_2} \right)$$

$$S^2_{x_{j_{e_1,e_2}}} = \left(\frac{1}{N-1} \right) \sum_{k=1}^{N} \left(x_{k_{j_{e_1,e_2}}} - \bar{X}_{j_{e_1,e_2}} \right)^2; \; for \; j = 1, 2, \ldots, 8.$$

$$S_{xx_{j_{e_1,e_2}}} = \left(\frac{1}{N-1} \right) \sum_{k=1}^{N} \left(x_{k_{e_1,e_2}} - \bar{X}_{e_1,e_2} \right) \left(x_{k_{j_{e_1,e_2}}} - \bar{X}_{j_{e_1,e_2}} \right); \; for \; j = 1, 2, \ldots, 8.$$

$$S_{yx_{j_{e_1,e_2}}} = \left(\frac{1}{N-1} \right) \sum_{k=1}^{N} \left(y_{k_{e_1,e_2}} - \bar{Y}_{e_1,e_2} \right) \left(x_{k_{j_{e_1,e_2}}} - \bar{X}_{j_{e_1,e_2}} \right); \; for \; j = 1, 2, \ldots, 8.$$

Note that all the notations are used in neutrosophic form where $\bar{X}_{je_1,e_2} \in [\bar{X}_{je_1}, \bar{X}_{je_2}]$, ratio $\bar{R}_{je_1,e_2} \in [\bar{R}_{je_1}, \bar{R}_{je_2}]$, $C_{x_{e_1,e_2}} \in [C_{x_{e_1}}, C_{x_{e_2}}]$, $\beta_{2x_{e_1,e_2}} \in [\beta_{2x_{e_1}}, \beta_{2x_{e_2}}]$, $\rho_{e_1,e_2} \in [\rho_{e_1}, \rho_{e_2}]$, $B_{e_1,e_2} \in [B_{e_1}, B_{e_2}]$, $S^2_{y_{e_1,e_2}} \in [S^2_{y_{e_1}}, S^2_{y_{e_2}}]$, $S^2_{x_{e_1,e_2}} \in [S^2_{x_{e_1}}, S^2_{x_{e_2}}]$, $S_{yx_{e_1,e_2}} \in [S_{yx_{e_1}}, S_{yx_{e_2}}]$, $S^2_{x_{je_1,e_2}} \in [S^2_{x_{je_1}}, S^2_{x_{je_2}}]$, $S_{xx_{je_1,e_2}} \in [S_{xx_{je_1}}, S_{xx_{je_2}}]$ and $S_{yx_{je_1,e_2}} \in [S_{yx_{je_1}}, S_{yx_{je_2}}]$. $y_{k_{e_1,e_2}}$ is representing observations of study variable. The estimators $T^u_{je_1,e_2} \in [T^u_{je_1}, T^u_{je_2}]$, are only using neutrosophic OLS regression coefficient.

3. Proposed Robust Regression Based Neutrosophic Hartley–Ross Type Estimators

When a dataset contains outliers or anomalous data points, the efficiency of ordinary least squares (OLS) estimates can be compromised. The breakdown value of OLS fitting is $1/n$ or 0%, indicating that it can be easily influenced by a single outlier, as pointed out by Hampel et al. [21] and Rousseeuw and Leroy [22]. According to Seber and Lee [23], the OLS method's susceptibility to outliers in data can be attributed to two main factors:

- When using the squared residual to estimate the residual size, any residual with a greater magnitude will have a disproportionately larger effect on the overall size compared to the other residuals.
- Using a conventional location measure, such as the arithmetic mean, which is not resistant to outliers, may result in a significant impact on the criterion due to a large squared value, which in turn can lead to a disproportionate effect on the regression results.

To reduce the influence of outliers on regression results, one may opt for alternative regression methods that are less sensitive or affected by outliers, such as robust regression. To gain a better understanding of these robust regression techniques, readers can consult the research conducted by Yu and Yao [24].

Ordinary least squares (OLS) regression is a widely used method for estimating the parameters of a linear regression model. It works by minimizing the sum of squared residuals, which represent the difference between the observed and predicted values of the dependent variable. The objective is to identify the best-fit line that can explain the relationship between the independent and dependent variables while minimizing the error in the model. However, the OLS method relies on the assumption that the residuals are normally distributed and have constant variance, which is not always the case in real-world data.

OLS regression is highly sensitive to outliers, which are observations that significantly deviate from other data points. Outliers can occur naturally or due to measurement or data entry errors. The impact of outliers can be significant, as they can affect the estimates of regression coefficients and lead to inaccurate predictions. In contrast, robust regression is a technique specifically designed to handle outliers more effectively, and is therefore less sensitive to their presence in the data.

Robust regression is one class of robust statistical methods that aims to minimize the sum of the absolute residuals rather than the sum of the squared residuals. By focusing on the absolute values of the residuals, this approach is less susceptible to the influence of outliers, as extreme values do not disproportionately impact the results. Rather than assigning equal weight to all observations, robust regression assigns lower weights to outliers, which reduces their impact on the regression coefficient estimates. In cases where outliers are prevalent, robust regression is often a more appropriate approach than OLS regression, as it is less affected by extreme values.

OLS regression assumes that the variance of the residuals is constant across all levels of the independent variable, which is known as homoscedasticity. However, this assumption may not hold in real-world data, particularly when the dependent variable is not normally distributed. Biased estimates of the regression coefficients and unreliable predictions may result from this inconsistency. Another possible limitation of OLS regression is its sensitivity to outliers or unusual data points, which can impact the estimates and decrease efficiency. However, the weighted least squares method employed by robust regression can address this problem by assigning higher weights to observations with smaller variances and lower

weights to those with larger variances. This technique allows for more precise estimation of the coefficients, even when outliers or influential points are present.

Robust regression outperforms OLS regression when dealing with non-normal data, which violates the normality assumption. OLS regression assumes that the dependent variable is normally distributed, which is not always true in real-world data. Consequently, the regression coefficients obtained from OLS regression may be biased, and the predictions may be inaccurate. Robust regression, on the other hand, does not make the assumption of normality and can handle non-normal data more effectively. We are using four robust regression methods ($p = 1, 2, \ldots, 4$), least absolute deviations (LAD) for $p = 1$, Huber-M (H) for $p = 2$, Hampel-M (H-M) for $p = 3$, and Tukey-M (T-M) for $p = 4$. For further details, please refer to Zaman and Bulut's work [25,26].

The proposed neutrosophic robust regression Hartley–Ross-type mean estimators are

$$T^u_{p_{j_{e_1,e_2}}} = \bar{X}_{j_{e_1,e_2}} \left[\bar{r}_{xy_{j_{e_1,e_2}}} + \frac{b_p.xy_{e_1,e_2}(\bar{X}_{e_1,e_2} - \bar{x}_{e_1,e_2})}{\bar{x}_{j_{e_1,e_2}}} \right] + \left(1 - \frac{1}{N}\right)\left(\frac{n}{n-1}\right)\left(\bar{y}_{e_1,e_2} - \bar{x}_{j_{e_1,e_2}} \bar{r}_{xy_{j_{e_1,e_2}}}\right); \tag{7}$$

For $j = 1, 2, \ldots, 8$ and $p = 1, 2, \ldots, 4$.

The variance of $T^u_{p_{j_{e_1,e_2}}}$ is as follows:

$$V(T^u_{p_{j_{e_1,e_2}}}) = \theta \left[S^2_{y_{e_1,e_2}} + B^2_{p_{e_1,e_2}} S^2_{x_{e_1,e_2}} + \bar{R}^2_{j_{e_1,e_2}} S^2_{x_{j_{e_1,e_2}}} - 2B_{p_{e_1,e_2}} S_{yx} + 2B_{p_{e_1,e_2}} \bar{R}_{j_{e_1,e_2}} S_{xx_{j_{e_1,e_2}}} - 2\bar{R}_{j_{e_1,e_2}} S_{yx_{j_{e_1,e_2}}} \right]. \tag{8}$$

4. Neutrosophic Robust Estimation in Sensitive Research

Sensitive topics, such as abortion, xenophobia, tax evasion, drug use, alcoholism, gambling, reckless driving, and sexual behavior, are regarded as intrusive as they violate the privacy of the respondents. When conducting research on sensitive issues, direct inquiries on personal or stigmatizing matters can lead to respondents either refusing to respond or providing inaccurate information, leading to non-sampling errors that significantly undermine the data's quality and subsequent analyses' relevance. Survey statisticians have developed various strategies to encourage respondent participation while respecting their privacy to minimize such socially desirable biases. One way to improve the accuracy of responses on sensitive topics is to reduce interviewer influence. Self-administered questionnaires, computer-assisted telephone interviews, computer-assisted self-interviews, and web surveys are commonly employed for this purpose. To avoid socially desirable bias in data collection on sensitive topics due to non-sampling errors, one solution is to utilize indirect questioning methods rather than direct questions.

Chaudhuri and Christofides [27] suggest various techniques for gathering sensitive information while circumventing direct questioning of survey participants. Among these methods is the randomized response technique, which was initially developed by Warner [16] as a way to estimate the population mean of a sensitive variable in a study while minimizing interviewer influence. While originally designed for binary variables to assess the prevalence of stigmatizing attributes, it has since been adapted to evaluate sensitive quantitative variables concerning diverse aspects of life, such as personal income level (Barabesi et al. [28]), extramarital relationships, induced abortions and unwanted pregnancies, tax evasion, weekly hours of undeclared work (Trappmann et al. [29]), the number of cannabis cigarettes smoked (Cobo et al. [30]), and the frequency of deviant sexual behaviors that students struggle to control (Perri et al. [31]). Academic literature proposes multiple methods to safeguard the confidentiality of respondents when gathering data on sensitive variables. These techniques aim to modify responses in a way that conceals the actual values of the variables, using one or more random variables to alter them. The pioneering works in this area were by Greenberg et al. [32], Eriksson [33], and Pollock and Beck [34]. Other researchers have since developed their own distortion techniques, including Eichhorn and Hayre [35], Bar-Lev et al. [36], and Diana and Perri [37,38]. These random variables must be statistically independent of the sensitive

variable and of each other, and the researcher must possess a thorough understanding of their probability distributions.

The present study explores the use of two neutrosophic scrambling variables, referred to as $\mathcal{U}_{e_1,e_2}$ and $\mathcal{V}_{e_1,e_2}$. To elicit the responses, each participant is instructed to generate a neutrosophic value, u_{e_1,e_2i}, from $\mathcal{U}_{e_1,e_2}$, and another neutrosophic value, v_{e_1,e_2i}, from $\mathcal{V}_{e_1,e_2}$. They are then requested to disclose the neutrosophic scrambled value $z_{i_{e_1,e_2}} = \varphi_{e_1,e_2}(y_{e_1,e_2i}, u_{e_1,e_2i}, v_{e_1,e_2i})$, where φ_{e_1,e_2} is the neutrosophic scrambling function that enables respondents to conceal their true sensitive value. The scrambling function's particular configuration is established by the chosen scrambling technique, which may be one of the two neutrosophic models elaborated in the subsequent sections.

It is assumed that the values generated by the respondent, u_{e_1,e_2i} and v_{e_1,e_2i}, are not disclosed to anyone, ensuring the researcher's uncertainty about the true value of y_{e_1,e_2i} and protecting respondent privacy. While it may not be possible to determine individual values of y_{e_1,e_2i}, it is still possible to obtain accurate estimates of certain characteristics of the sensitive variable $\mathcal{Y}e_1, e_2$ by taking a sample of n units and using the shuffled responses of all units in the sample, denoted by $(z_{1_{e_1,e_2}}, \ldots, z_{n_{e_1,e_2}})$.

This paper considers two neutrosophic versions of scrambling models: (i) the *additive model* $\mathcal{Z}_{e_1,e_2} = \mathcal{Y}_{e_1,e_2} + \mathcal{U}_{e_1,e_2}$ (Pollock and Beck, [34]) and (ii) the *mixed model* $\mathcal{Z}_{e_1,e_2} = (\mathcal{Y}_{e_1,e_2} + \mathcal{U}_{e_1,e_2})\mathcal{V}_{e_1,e_2}$ (Saha, [39]). Using the scrambling models previously discussed, it is possible to obtain an unbiased estimate of the unknown mean $\bar{Y}_{e_1,e_2}$ of the sensitive variable by employing the scrambled values $z_{i_{e_1,e_2}}$ and computing the estimator's variance. As an example, assume that we use the additive model to distort the true responses and intend to estimate the mean of Ye_1, e_2 using a simple random sample without replacement (SRSWOR) of n units taken from the population. With this approach, the estimation process becomes straightforward. Let $\mathcal{U}_{e_1,e_2} > 0$ be a scrambling variable with a mean of $\bar{U}_{e_1,e_2}$ and a variance of $S^2_{u_{e_1,e_2}}$. As the distribution of $\mathcal{U}_{e_1,e_2}$ is known, $\bar{U}_{e_1,e_2}$ and $S^2_{u_{e_1,e_2}}$ will also be known beforehand. To obtain an estimate of the unknown sensitive mean $\bar{Y}_{e_1,e_2}$ using the additive model, each selected respondent is directed to generate a number u_{e_1,e_2i} from $\mathcal{U}_{e_1,e_2}$ using a computer or smartphone application, and add this number to their true value y_{e_1,e_2i}. The respondent is then instructed to release the scrambled response $z_{i_{e_1,e_2}} = y_{e_1,e_2i} + u_{e_1,e_2i}$, while keeping the generated value u_{e_1,e_2i} confidential. The estimator $\hat{y}_{e_1,e_2i} = z_{i_{e_1,e_2}} - u_{e_1,e_2i}$ is an unbiased estimator of the unknown y_{e_1,e_2i}, and its unbiasedness is assessed with respect to the scrambling device. This can be expressed using the expectation operator E, such that:

$$\text{E}(\hat{y}_{e_1,e_2i}) = \text{E}(z_{i_{e_1,e_2}}) - \bar{U}_{e_1,e_2} = (y_{e_1,e_2i} + \bar{U}_{e_1,e_2}) - \bar{U}_{e_1,e_2} = y_{e_1,e_2i},$$

Using the notation introduced in previous sections, it becomes evident that for $i = 1, \ldots, n$:

$$\hat{\bar{y}}_{e_1,e_2} = \bar{z}_{e_1,e_2} - \mu_{u_{e_1,e_2i}},$$

with $\bar{z}_{e_1,e_2} = \sum_i^n z_{i_{e_1,e_2}}/n$, is a design-unbiased estimator of $\bar{Y}_{e_1,e_2}$ with variance $\text{Var}(\hat{\bar{y}}_{e_1,e_2}) = \gamma(S^2_{y_{e_1,e_2}} + S^2_{u_{e_1,e_2}})$.

The mixed model can also be treated using the same method. However, to maintain brevity, we have excluded the specifics from this explanation. Readers who want to know more can consult Shahzad et al. [40].

Proposed Neutrosophic Hartley–Ross Type Estimators for Scrambled Responses

Building on the concepts mentioned above, we adapt the suggested class to handle cases where the objective variable is sensitive, and information is obtained utilizing the four scrambling methods described earlier. Let $(z_{1_{e_1,e_2}}, \ldots, z_{n_{e_1,e_2}})$ denote the neutrosophic responses observed on a sample selected from the population under study using SRSWOR. Let $\bar{r}_{xz_{j_{e_1,e_2}}} = \frac{\sum_{k=1}^n r_{k.xz_{j_{e_1,e_2}}}}{n}$. We can obtain the class of neutrosophic estimators with scrambled responses by replacing y in $T^u_{p_{j_{e_1,e_2}}}$ with z

$$T^{u'}_{p_{j_{e_1,e_2}}} = \bar{X}_{j_{e_1,e_2}} \left[\bar{r}_{xz_{j_{e_1,e_2}}} + \frac{b.p.xz_{e_1,e_2}\left(\bar{X}_{e_1,e_2} - \bar{x}_{e_1,e_2}\right)}{\bar{x}_{j_{e_1,e_2}}} \right] + \left(1 - \frac{1}{N}\right)\left(\frac{n}{n-1}\right)\left(\bar{z}_{e_1,e_2} - \bar{x}_{j_{e_1,e_2}} \bar{r}_{xz_{j_{e_1,e_2}}}\right); \tag{9}$$

For $j = 1, 2, \ldots, 8$ *and* $p = 1, 2, \ldots, 4$.

To obtain the variance equations, we can substitute the population parameter for $\mathcal{Y}_{e_1,e_2}$ with the relevant population parameters for $\mathcal{Z}_{e_1,e_2}$ in Equation (8). However, to maintain brevity, we abstain from presenting the altered formulas, as they can be effortlessly derived. The identical approach can be utilized to extend all the estimators discussed in the previous Sections 3 and 4 to the sensitive setting, which will be explored in the following section.

5. Numerical Illustration

As the idea of neutrosophic Hartley–Ross-type estimators is relatively novel and no prior research is available on the topic to the authors' understanding, a study is conducted to compare the variance performance of different neutrosophic estimators. Specifically, we compared the proposed neutrosophic robust regression Hartley–Ross-type estimators (Section 3) and the adapted neutrosophic estimators (Section 2) to determine which performed better. Typically, selecting the best estimator from a set of estimators is based on the one that results in the lowest variance.

We conducted a numerical analysis using interval data with indeterminate values from the Islamabad Stock Market, specifically the United Bank Limited (UBL). The data consist of neutrosophic values that are uncertain and fall within a certain range. We chose UBL's stock market share price data because the price values fluctuate within a range, with the recorded share price for the day potentially being the highest, lowest, or any value in between. The data were obtained from the publicly available website (https://pkfinance.info, accessed on 14 September 2021), and approval was not needed. Tables 1 and 2 present the neutrosophic features of the data, which were obtained using the share price index values for 2019 and 2020 denoted as X_{e_1,e_2} and Y_{e_1,e_2}, respectively. It is worth mentioning that $N = 239$ and $n = 30$.

Table 1. Neutrosophic characteristics of Population.

$S_{y_{e_1,e_2}} = [54.5609, 54.6335]$	$\beta_{e_1,e_2} = [0.850813, 0.8378018]$
$S_{x_{e_1,e_2}} = [55.7757, 56.2250]$	$\beta_{e_1,e_2(1)} = [0.7242686, 0.7075029]$
$S^2_{y_{e_1,e_2}} = [2976.893, 2984.826]$	$\beta_{e_1,e_2(4)} = [0.8390644, 0.8217442]$
$S^2_{x_{e_1,e_2}} = [3110.931, 3161.251]$	$\beta_{e_1,e_2(5)} = [0.8386558, 0.8201294]$
$S_{xy_{e_1,e_2}} = [2646.821, 2648.501]$	$\beta_{e_1,e_2(6)} = [0.682883, 0.6580855]$

Table 2. Neutrosophic Specified Characteristics of Population

J	$\bar{R}_{(j)}$	$S^2_{x(j)}$	$S_{xx(j)}$	$S_{xy(j)}$
1	[0.8751, 0.8795]	[15,110.2372, 15,354.6461]	[15,110.2372, 15,354.6461]	[12,855.9868, 12,864.1495]
2	[0.7823, 0.7889]	[15,110.2372, 15,354.6461]	[15,110.2372, 15,354.6461]	[12,855.9868, 12,864.1495]
3	[0.0504, 0.0509]	[4,586,469.0853, 4,611,343.7986]	[263,253.9374, 266,093.1264]	[223,979.8819, 222,933.2903]
4	[1.7761, 1.8239]	[2099.7334, 2057.3919]	[5632.7143, 5620.5448]	[4792.3868, 4708.9024]
5	[2.316, 2.3707]	[2099.733, 2057.3919]	[5632.714, 5620.5448]	[4792.387, 4708.9024]
6	[1.0058, 1.0197]	[11,430.5413, 11,414.6395]	[13,142.2293, 13,238.8727]	[11,181.5800, 11,091.5509]
7	[0.0503, 0.0509]	[4,586,469.0853, 4,611,343.7986]	[263,253.9374, 266,093.1264]	[223,979.8819, 222,933.2903]
8	[0.8852, 0.8999]	[11,430.5413, 11,414.6395]	[13,142.2293, 13,238.8727]	[11,181.5800, 11,091.5509]

For the sensitive scenario described in Section 4, two neutrosophic scrambling techniques are used to perturb the target variable values. These techniques, proposed by [Shahzad perri hanif], involve the use of scrambling variables $\mathcal{U}_{e_1,e_2}$ and $\mathcal{V}_{e_1,e_2}$, which are assumed to follow a normal distribution. The neutrosophic mean of these variables is

[0, 0], and their neutrosophic standard deviation is 10% of the standard deviation of the auxiliary variable.

5.1. Simulation Study

For simulation study, we generate neutrosophic population of size 1000. Taking motivation from Aslam et al. [41] and Shahzad et al. [42], neutrosophic random variables X_{e_1,e_2} and Y_{e_1,e_2} generated as follows:

$$Y_{e_1,e_2} = \delta_{e_1,e_2} + K_{e_1,e_2} X_{e_1,e_2} + \varepsilon_{e_1,e_2} X_{e_1,e_2}^{q_{e_1,e_2}}$$

where X_{e_1,e_2} follows neutrosophic Gamma distribution i.e., $\text{Gamma}(A_{e_1,e_2}, B_{e_1,e_2})$. $A_{e_1,e_2} = [2.0, 2.0]$ and $B_{e_1,e_2} = [3.1, 3.8]$. ε_{e_1,e_2} follows neutrosophic standard normal. Further, $\delta_{e_1,e_2} = [5, 5]$, $q_{e_1,e_2} = [1.6, 1.6]$ and $K_{e_1,e_2} = [2, 2]$.

By adapting Shahzad et al. [42], $h = 5000$ samples of sizes 100 were chosen independently under the simple random sampling design, and for the hth sample, the mean estimate $(\hat{\phi}_{e_1,e_2}^{(h_1)}, \hat{\phi}_{e_1,e_2}^{(h_2)})$ of study variable was calculated. Note that $\hat{\phi}_{e_1,e_2}^{(h_1)}$ is representing adapted estimators and $\hat{\phi}_{e_1,e_2}^{(h_2)}$ is representing proposed estimators.

The variances were calculated from the formula given below

$$\text{Var}(\hat{\phi}_{e_1,e_2}^{(h_1)}) = \sum_{h_1=1}^{h} (\hat{\phi}_{e_1,e_2}^{(h_1)} - \bar{Y}_{e_1,e_2})^2 / h.$$

$$\text{Var}(\hat{\phi}_{e_1,e_2}^{(h_2)}) = \sum_{h_2=1}^{h} (\hat{\phi}_{e_1,e_2}^{(h_2)} - \bar{Y}_{e_1,e_2})^2 / h.$$

The calculated variances and percentage relative efficiencies (PREs) are presented in Tables 3–14. Where

$$\text{PRE} = \frac{\text{Var}(\hat{\phi}_{e_1,e_2}^{(h_1)})}{\text{Var}(\hat{\phi}_{e_1,e_2}^{(h_2)})} \times 100$$

5.2. Remarks

We explore the following points from numerical investigation:

- Tables 3 and 4 present the results of the numerical study in terms of the variance and PRE. For the non-sensitive case, it is noteworthy that all of the estimators belonging to the proposed class $T^u_{1_{je_1,e_2}}, T^u_{2_{je_1,e_2}}, T^u_{3_{je_1,e_2}}, T^u_{4_{je_1,e_2}}$ perform better than the adapted estimators.
- The results displayed in Tables 5–8 also support the excellent performance of the estimators from the proposed class $T^u_{1_{je_1,e_2}}, T^u_{2_{je_1,e_2}}, T^u_{3_{je_1,e_2}}, T^u_{4_{je_1,e_2}}$ when estimating scrambled responses. Regardless of the specific scrambling device utilized, all the estimators from the proposed class consistently outperform the adapted estimators.
- The results of simulation study in Tables 9–14, also show same behavior i.e., superiority of proposed class $T^u_{1_{je_1,e_2}}, T^u_{2_{je_1,e_2}}, T^u_{3_{je_1,e_2}}, T^u_{4_{je_1,e_2}}$.

Table 3. Variance of estimators using UBL data for nonsensitive case.

j	OLS	LAD	Huber	Hampel	Tukey
1	[368.91, 379.29]	[269.99, 275.33]	[359.59, 366.31]	[359.27, 365.01]	[238.30, 236.74]
2	[299.20, 309.66]	[210.94, 216.57]	[290.88, 298.01]	[290.59, 296.84]	[182.72, 182.10]
3	[370.61, 380.97]	[271.45, 276.75]	[361.27, 367.95]	[360.95, 366.64]	[239.68, 238.07]
4	[220.46, 228.29]	[145.99, 149.76]	[213.42, 218.44]	[213.18, 217.45]	[122.29, 120.81]
5	[359.61, 369.87]	[262.06, 267.31]	[350.43, 357.06]	[350.11, 355.77]	[230.80, 229.26]
6	[368.63, 379.01]	[269.76, 275.09]	[359.33, 366.03]	[359.0071, 364.73]	[238.08, 236.51]
7	[370.47, 380.83]	[271.34, 276.63]	[361.14, 367.81]	[360.81, 366.51]	[239.56, 237.96]
8	[290.43, 300.27]	[203.59, 208.74]	[282.24, 288.81]	[281.95, 287.66]	[175.84, 174.87]

Table 4. PRE of estimators using UBL data for nonsensitive case.

	$T^u_{1_{je_1,e_2}}$	$T^u_{2_{je_1,e_2}}$	$T^u_{3_{je_1,e_2}}$	$T^u_{4_{je_1,e_2}}$
$T_{1_{e_1,e_2}}$	[136.6316, 137.7609]	[102.5886, 103.5452]	[102.6809, 103.9149]	[154.8056, 160.2158]
$T_{2_{e_1,e_2}}$	[141.8419, 142.9838]	[102.8604, 103.9079]	[102.9626, 104.3169]	[163.7456, 170.0431]
$T_{3_{e_1,e_2}}$	[136.5254, 137.6559]	[102.5828, 103.5377]	[102.6749, 103.9066]	[154.6267, 160.0225]
$T_{4_{e_1,e_2}}$	[151.0041, 152.4383]	[103.2983, 104.5101]	[103.4167, 104.9850]	[180.2740, 188.9546]
$T_{5_{e_1,e_2}}$	[137.2249, 138.3676]	[102.6204, 103.5885]	[102.7139, 103.9629]	[155.8076, 161.3358]
$T_{6_{e_1,e_2}}$	[136.6484, 137.7787]	[102.5895, 103.5465]	[102.6818, 103.9163]	[154.8340, 160.2486]
$T_{7_{e_1,e_2}}$	[136.5340, 137.6646]	[102.5833, 103.5383]	[102.6754, 103.9073]	[154.6412, 160.0385]
$T_{8_{e_1,e_2}}$	[142.6484, 143.8463]	[102.9009, 103.9656]	[103.0046, 104.3809]	[165.1583, 171.7071]

Table 5. Variance of estimators using UBL data for sensitive case using additive model.

j	OLS	LAD	Huber	Hampel	Tukey
1	[586.40, 630.84]	[456.33, 483.00]	[560.47, 622.87]	[566.96, 623.27]	[453.69, 428.25]
2	[472.94, 512.02]	[356.85, 379.64]	[449.76, 504.88]	[455.56, 505.23]	[354.50, 330.71]
3	[589.18, 633.69]	[458.78, 485.51]	[563.18, 625.71]	[569.69, 626.10]	[456.14, 430.62]
4	[344.82, 373.24]	[246.85, 261.53]	[325.23, 367.20]	[330.13, 367.49]	[244.87, 220.41]
5	[571.36, 614.80]	[443.05, 468.96]	[545.77, 606.95]	[552.17, 607.33]	[440.45, 414.96]
6	[585.94, 630.35]	[455.92, 482.57]	[560.02, 622.39]	[566.51, 622.78]	[453.29, 427.85]
7	[588.95, 633.45]	[458.58, 485.30]	[562.96, 625.47]	[569.47, 625.87]	[455.94, 430.43]
8	[458.18, 495.74]	[344.03, 365.61]	[435.39, 488.72]	[441.09, 489.06]	[341.72, 317.54]

Table 6. PRE of estimators using UBL data for sensitive case using additive model.

	$T^u_{1_{je_1,e_2}}$	$T^u_{2_{je_1,e_2}}$	$T^u_{3_{je_1,e_2}}$	$T^u_{4_{je_1,e_2}}$
$T_{1_{e_1,e_2}}$	[128.5044, 130.6071]	[104.6270, 101.2784]	[103.4291, 101.2149]	[129.2511, 147.3048]
$T_{2_{e_1,e_2}}$	[132.5325, 134.8704]	[105.1521, 101.4143]	[103.8136, 101.3439]	[133.4108, 154.8228]
$T_{3_{e_1,e_2}}$	[128.4225, 130.5217]	[104.6161, 101.2756]	[103.4211, 101.2122]	[129.1666, 147.1569]
$T_{4_{e_1,e_2}}$	[139.6876, 142.7103]	[106.0248, 101.6449]	[104.4506, 101.5629]	[140.8156, 169.3403]
$T_{5_{e_1,e_2}}$	[128.9597, 131.0990]	[104.6877, 101.2945]	[103.4736, 101.2301]	[129.7209, 148.1594]
$T_{6_{e_1,e_2}}$	[128.5178, 130.6218]	[104.6288, 101.2789]	[103.4305, 101.2153]	[129.2650, 147.3303]
$T_{7_{e_1,e_2}}$	[128.4291, 130.5288]	[104.6170, 101.2758]	[103.4218, 101.2124]	[129.1734, 147.1691]
$T_{8_{e_1,e_2}}$	[133.1796, 135.5902]	[105.2341, 101.4365]	[103.8735, 101.3650]	[134.0797, 156.1176]

Table 7. Variance of estimators using UBL data for sensitive case using mixed model.

j	OLS	LAD	Huber	Hampel	Tukey
1	[15,912.6, 15,988.9]	[15,169.6, 15,245.9]	[14,087.1, 14,163.3]	[15,121.8, 15,198.1]	[13,694.8, 13,771.1]
2	[13,992.0, 14,068.3]	[13,328.0, 13,404.3]	[12,361.6, 12,437.9]	[13,285.3, 13,361.6]	[12,011.9, 12,088.2]
3	[15,959.6, 16,035.9]	[15,214.8, 15,291.1]	[14,129.5, 14,205.8]	[15,166.9, 15,243.2]	[13,736.3, 13,812.6]
4	[11,820.2, 11,896.5]	[11,258.7, 11,335.0]	[10,443.2, 10,519.5]	[11,222.6, 11,298.9]	[10,148.6, 10,224.9]
5	[15,864.1, 15,940.4]	[15,122.9, 15,199.2]	[14,043.1, 14,119.4]	[15,075.3, 15,151.6]	[13,651.9, 13,728.2]
6	[15,783.4, 15,859.7]	[15,045.4, 15,121.7]	[13,970.2, 14,046.5]	[14,998.0, 15,074.3]	[13,580.7, 13,657.0]
7	[15,960.4, 16,036.7]	[15,215.5, 15,291.8]	[14,130.1, 14,206.4]	[15,167.6, 15,243.9]	[13,736.9, 13,813.2]
8	[10,983.2, 11,059.5]	[10,466.6, 10,542.9]	[97,17.1, 9793.4]	[10,433.5, 10,509.8]	[9446.601, 9522.9]

Table 8. PRE of estimators using UBL data for sensitive case using mixed model.

	$T_{1_{je_1,e_2}}^u$	$T_{2_{je_1,e_2}}^u$	$T_{3_{je_1,e_2}}^u$	$T_{4_{je_1,e_2}}^u$
$T_{1_{e_1,e_2}}$	[101.8748, 104.8735]	[109.8912, 112.8899]	[102.2043, 105.2030]	[113.1064, 116.1051]
$T_{2_{e_1,e_2}}$	[101.9550, 104.9537]	[110.1092, 113.1079]	[102.2900, 105.2887]	[113.3817, 116.3804]
$T_{3_{e_1,e_2}}$	[101.8724, 104.8711]	[109.8845, 112.8832]	[102.2017, 105.2004]	[113.0978, 116.0965]
$T_{4_{e_1,e_2}}$	[101.8550, 104.9537]	[109.9912, 113.0899]	[102.1897, 105.2884]	[113.2499, 116.3486]
$T_{5_{e_1,e_2}}$	[101.7773, 104.8760]	[109.7981, 112.8968]	[102.1070, 105.2057]	[113.0151, 116.1138]
$T_{6_{e_1,e_2}}$	[101.7814, 104.8801]	[109.8094, 112.9081]	[102.1114, 105.2101]	[113.0295, 116.1282]
$T_{7_{e_1,e_2}}$	[101.7723, 104.8710]	[109.7844, 112.8831]	[102.1016, 105.2003]	[112.9977, 116.0964]
$T_{8_{e_1,e_2}}$	[101.8015, 104.9002]	[109.8290, 112.9277]	[102.1323, 105.2310]	[113.0369, 116.1356]

Table 9. Variance of estimators using simulation study for nonsensitive case.

j	OLS	LAD	Huber	Hampel	Tukey
1	[307.3, 317.7]	[208.4, 213.7]	[298.0, 304.7]	[297.7, 303.4]	[175.1, 176.7]
2	[237.6, 248.0]	[149.3, 155.0]	[229.3, 236.4]	[229.0, 235.2]	[120.5, 121.1]
3	[309.0, 319.4]	[209.8, 215.1]	[299.7, 306.3]	[299.3, 305.0]	[176.5, 178.1]
4	[158.8, 166.7]	[84.4, 88.1]	[151.8, 156.8]	[151.6, 155.8]	[59.2, 60.7]
5	[298.0, 308.3]	[200.4, 205.7]	[288.8, 295.4]	[288.5, 294.2]	[167.6, 169.2]
6	[307.0, 317.4]	[208.2, 213.5]	[297.7, 304.4]	[297.4, 303.1]	[174.9, 176.5]
7	[308.9, 319.2]	[209.7, 215.0]	[299.5, 306.2]	[299.2, 304.9]	[176.3, 178.0]
8	[228.8, 238.7]	[142.0, 147.1]	[220.6, 227.2]	[220.3, 226.1]	[113.3, 114.2]

Table 10. PRE of estimators using simulation study for nonsensitive case.

	$T_{1_{je_1,e_2}}^u$	$T_{2_{je_1,e_2}}^u$	$T_{3_{je_1,e_2}}^u$	$T_{4_{je_1,e_2}}^u$
$T_{1_{e_1,e_2}}$	[176.9907, 177.4387]	[123.7817, 124.4646]	[123.9161, 124.9992]	[208.7344, 216.5262]
$T_{2_{e_1,e_2}}$	[199.7772, 200.3055]	[130.4769, 130.9640]	[130.6415, 131.6144]	[246.9560, 256.8951]
$T_{3_{e_1,e_2}}$	[177.0409, 176.5727]	[123.6558, 124.3434]	[123.7897, 124.8759]	[208.0762, 215.8408]
$T_{4_{e_1,e_2}}$	[258.8548, 261.1166]	[145.1785, 145.5268]	[145.4125, 146.4492]	[363.0446, 385.2948]
$T_{5_{e_1,e_2}}$	[179.7725, 179.3638]	[125.1716, 124.4926]	[125.7186, 124.6302]	[212.4889, 220.5698]
$T_{6_{e_1,e_2}}$	[177.0573, 177.5064]	[123.8018, 124.4852]	[125.0201, 123.9363]	[208.8393, 216.6429]
$T_{7_{e_1,e_2}}$	[176.6066, 177.0739]	[123.6660, 124.3535]	[123.7999, 124.8861]	[208.1295, 215.8975]
$T_{8_{e_1,e_2}}$	[204.0212, 204.4839]	[131.6100, 132.1326]	[131.7798, 132.8042]	[254.1353, 265.0087]

Table 11. Variance of estimators using simulation study for sensitive case using additive model.

j	OLS	LAD	Huber	Hampel	Tukey
1	[527.89, 530.46]	[411.27, 415.07]	[508.18, 519.83]	[519.52, 522.20]	[398.93, 400.89]
2	[414.00, 415.72]	[309.32, 313.29]	[396.38, 406.21]	[406.52, 408.33]	[298.89, 300.06]
3	[530.68, 533.27]	[413.78, 417.58]	[510.92, 522.61]	[522.29, 524.99]	[401.39, 403.37]
4	[285.40, 286.14]	[196.32, 200.38]	[270.50, 278.10]	[279.07, 279.89]	[188.25, 188.52]
5	[512.83, 515.12]	[397.55, 401.54]	[493.39, 504.63]	[504.58, 506.96]	[385.62, 387.32]
6	[527.43, 530.03]	[410.88, 414.66]	[507.72, 519.40]	[519.06, 521.77]	[398.51, 400.50]
7	[530.45, 533.04]	[413.57, 417.38]	[510.70, 522.38]	[522.07, 524.76]	[401.19, 403.17]
8	[398.91, 401.60]	[296.8875, 299.92]	[381.60, 392.25]	[391.56, 394.33]	[285.77, 287.78]

Table 12. PRE of estimators using simulation study for sensitive case using additive model.

	$T^u_{1_{je_1,e_2}}$	$T^u_{2_{je_1,e_2}}$	$T^u_{3_{je_1,e_2}}$	$T^u_{4_{je_1,e_2}}$
$T_{1_{e_1,e_2}}$	[142.0121, 143.9523]	[113.8895, 115.9938]	[113.3728, 113.4615]	[147.6794, 147.7610]
$T_{2_{e_1,e_2}}$	[151.7972, 154.3029]	[117.4960, 119.9756]	[116.8868, 116.9846]	[159.0606, 159.1080]
$T_{3_{e_1,e_2}}$	[141.8270, 143.7570]	[113.8206, 115.9175]	[113.3057, 113.3941]	[147.4654, 147.5474]
$T_{4_{e_1,e_2}}$	[173.1540, 177.1131]	[125.0296, 128.2668]	[124.2297, 124.3282]	[184.3034, 184.4321]
$T_{5_{e_1,e_2}}$	[143.0478, 145.0583]	[114.2794, 116.4197]	[113.7526, 113.8379]	[148.8918, 148.9565]
$T_{6_{e_1,e_2}}$	[142.0432, 143.9826]	[113.9002, 116.0066]	[113.3833, 113.4728]	[147.7127, 147.7968]
$T_{7_{e_1,e_2}}$	[141.8418, 143.7729]	[113.8262, 115.9236]	[113.3112, 113.3995]	[147.4829, 147.5645]
$T_{8_{e_1,e_2}}$	[153.5353, 156.0112]	[118.0811, 120.6719]	[117.4570, 117.6013]	[160.9467, 161.1360]

Table 13. Variance of estimators using simulation study for sensitive case using mixed model.

j	OLS	LAD	Huber	Hampel	Tukey
1	[18,164.0, 49,793.8]	[14,605.3, 44,856.9]	[14,853.5, 51,189.0]	[14,577.7, 50,701.9]	[13,595.1, 51,319.2]
2	[16,434.7, 44,514.0]	[13,262.6, 40,095.4]	[13,483.3, 45,764.6]	[13,238.1, 45,327.9]	[12,365.5, 45,881.3]
3	[18,206.4, 49,920.7]	[14,638.7, 44,972.1]	[14,887.5, 51,319.2]	[14,611.0, 50,831.0]	[13,625.9, 51,449.8]
4	[14,482.8, 47,100.7]	[11,811.8, 38,360.7]	[11,996.8, 34,634.8]	[11,791.2, 39,417.9]	[11,061.0 39,516.7]
5	[18,135.5, 49,833.9]	[14,582.8, 44,893.3]	[14,830.6, 51,230.1]	[14,555.3, 50,742.7]	[13,574.4, 51,360.5]
6	[18,014.3, 48,545.8]	[14,487.4, 43,726.5]	[14,733.3, 49,908.3]	[14,460.0, 49,432.6]	[13,486.5, 50,035.4]
7	[18,207.4, 49,926.5]	[14,639.4, 44,977.3]	[14,888.3, 51,325.2]	[14,611.8, 50,836.9]	[13,626.6, 51,455.7]
8	[13,247.1, 28,291.3]	[10,947.2, 26,125.3]	[11,105.7, 28,913.0]	[10,929.5, 28,695.5]	[10,304.9, 28,971.2]

Table 14. PRE of estimators using simulation study for sensitive case using mixed model.

	$T^u_{1_{je_1,e_2}}$	$T^u_{2_{je_1,e_2}}$	$T^u_{3_{je_1,e_2}}$	$T^u_{4_{je_1,e_2}}$
$T_{1_{e_1,e_2}}$	[110.98664, 124.28688]	[101.25764, 122.20978]	[103.19194, 124.52223]	[101.01085, 133.52196]
$T_{2_{e_1,e_2}}$	[110.99896, 123.82998]	[101.24869, 121.80316]	[103.18551, 124.05948]	[101.00121, 132.81429]
$T_{3_{e_1,e_2}}$	[110.98474, 124.29265]	[101.25818, 122.21488]	[103.19232, 124.52807]	[101.01144, 133.53117]
$T_{4_{e_1,e_2}}$	[110.73314, 122.51557]	[101.29626, 120.62621]	[103.21635, 122.72920]	[101.05309, 130.83091]
$T_{5_{e_1,e_2}}$	[110.98605, 124.28288]	[101.25781, 122.20624]	[103.19206, 124.51817]	[101.01103, 133.51557]
$T_{6_{e_1,e_2}}$	[111.00195, 124.26474]	[101.25299, 122.19019]	[103.18875, 124.49978]	[101.00581, 133.48676]
$T_{7_{e_1,e_2}}$	[110.98466, 124.29279]	[101.25821, 122.21500]	[103.19233, 124.52821]	[101.01146, 133.53138]
$T_{8_{e_1,e_2}}$	[108.25793, 120.90389]	[101.82011, 119.17755]	[103.56156, 121.09875]	[101.62347, 128.43940]

6. Conclusions

The point estimate in survey sampling has a drawback of fluctuating across different samples due to sampling error, as it provides only a single value for the parameter being discussed. However, the neutrosophic approach, introduced by Florentin Smarandache, offers a valuable solution for estimating parameters in sampling theory. It provides interval estimates that have a high probability of containing the parameter. Consequently, the neu-

trosophic technique, which extends the classical approach, is employed to handle data that are ambiguous, indeterminate, or uncertain. In this paper, a novel class of Hartley–Ross-type estimators is proposed for estimating the population mean using neutrosophic robust regression. The proposed estimators are shown to outperform several other adapted estimators. The approach is based on recent advancements in neutrosophic statistics and is applied to both standard and sensitive settings where the target variable is protected using scrambling techniques. The performance of the proposed estimators is evaluated using UBL data and simulation. It is demonstrated that they outperform other estimators in both settings. In future studies, the work can be extended for other sampling designs.

Author Contributions: Conceptualization, A.M.A. and U.S.; methodology, A.M.A. and U.S.; software, A.M.A. and U.S.; validation, U.S.; formal analysis, U.S.; investigation, A.M.A. and U.S.; resources, A.M.A. and U.S.; data curation, U.S.; writing—original draft preparation, A.M.A. and U.S.; writing—review and editing, A.M.A. and U.S.; visualization, U.S.; supervision, U.S.; project administration, A.M.A. and U.S.; funding acquisition, A.M.A. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All the relevant data information is avaiable within the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cochran, W.G. The estimation of the yields of cereal experiments by sampling for the ratio gain to total produce. *J. Agric. Soc.* **1940**, *30*, 262–275. [CrossRef]
2. Robson, D. Applications of multivariate polykays to the theory of unbiased ratio-type estimation. *J. Am. Stat. Assoc.* **1957**, *52*, 511–522. [CrossRef]
3. Tahir, Z.; Khan, H.; Aslam, M.; Shabbir, J.; Mahmood, Y.; Smarandache, F. Neutrosophic ratio-type estimators for estimating the population mean. *Complex Intell. Syst.* **2021**, *7*, 2991–3001. [CrossRef]
4. Vishwakarma, G.K.; Singh, A. Generalized estimator for computation of population mean under neutrosophic ranked set technique: An application to solar energy data. *Comput. Appl. Math.* **2022**, *41*, 144. [CrossRef]
5. Yadav, S.K.; Smarandache, F. Generalized Neutrosophic Sampling Strategy for Elevated estimation of Population Mean. *Neutrosophic Sets Syst.* **2023**, *53*, 14.
6. Kumar, S.; Kour, S.P.; Choudhary, M.; Sharma, V. Determination of Population Mean Using Neutrosophic, Exponential-Type Estimator. *Lobachevskii J. Math.* **2022**, *43*, 3359–3367. [CrossRef]
7. Smarandache, F. *Introduction to Neutrosophic Statistics, Sitech and Education Publisher, Craiova*; Romania-Educational Publisher: Columbus, OH, USA, 2014; p. 123.
8. Smarandache, F. *Neutrosophy: Neutrosophic Probability, Set, and Logic: Analytic Synthesis and Synthetic Analysis*; American Research Press: Rehoboth, NM, USA, 1998.
9. Chen, J.; Ye, J.; Du, S. Scale effect and anisotropy analyzed for neutrosophic numbers of rock joint roughness coefficient based on neutrosophic statistics. *Symmetry* **2017**, *9*, 208. [CrossRef]
10. Hartley, H.O.; Ross, A. Unbiased ratio estimators. *Nature* **1954**, *174*, 270–271. [CrossRef]
11. Murthy, M.N. Almost unbiased estimators based on interpenetrating sub-samples. *Sankhyā Indian J. Stat. Ser. A* **1962**, *24*, 303–314.
12. Smoo, J.; Smoo, L.N.; Mohanty, S. An Alternative Approach to Estimation in Two-Phase Sampling Using two Auxiliary Variables. *Biom. J.* **1994**, *36*, 293–298. [CrossRef]
13. Singh, R.; Singh, H.P. A Hartley–Ross-type estimator for finite population mean when the variables are negatively correlated. *Metron* **1995**, *43*, 205–216.
14. Singh, H.P.; Sharma, B.; Tailor, R. Hartley–Ross-type estimators for population mean using known parameters of auxiliary variate. *Commun.-Stat.-Theory Methods* **2014**, *43*, 547–565. [CrossRef]
15. Kadilar, C.; Cekim, H.O. Hartley–Ross-type estimators in simple random sampling. In Proceedings of the International Conference on Numerical Analysis and Applied Mathematics, Rhodes, Greece, 22–28 September 2014; AIP Publishing LLC: Melville, NY, USA, 2015; Volume 1648, p. 610007.
16. Warner, S.L. Randomized response: A survey technique for eliminating evasive answer bias. *J. Am. Stat. Assoc.* **1965**, *60*, 63–69. [CrossRef] [PubMed]
17. Fox, J.A.; Tracy, P.E. *Randomized Response: A Method for Sensitive Survey*; Sage Publication, Inc.: Newbury Park, CA, USA, 1986.
18. Chaudhuri, A.; Mukerjee, R. *Randomized Response: Theory and Techniques*; Marcel Dekker, Inc.: New York, NY, USA, 1988.
19. Chaudhuri, A. *Randomized Response and Indirect Questioning Techniques in Surveys*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2011.

20. Al-Noor, N.H.; Mohammad, A.A. Model of robust regression with parametric and nonparametric methods. *Math. Theory Model.* **2013**, *3*, 27–39.

21. Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J.; Stahel, W.A. *Robust Statistics: The Approach Based on Influence Functions*; John Wiley & Sons Publication: New York, NY, USA, 1986.

22. Rousseeuw, P.J.; Leroy, A.M. *Robust Regression and Outlier Detection*; John Wiley & Sons. Inc.: New York, NY, USA, 1987.

23. Seber, G.A.; Lee, A.J. *Linear Regression Analysis*, 2nd ed.; John Wiley & Sons Publication: New York, NY, USA, 2003.

24. Yu, C.; Yao, W. Robust linear regression: A review and comparison. *Commun. -Stat.-Simul. Comput.* **2017**, *46*, 6261–6282. [CrossRef]

25. Zaman, T.; Bulut, H. Modified ratio estimators using robust regression methods. *Commun. Stat. -Theory Methods* **2019**, *48*, 2039–2048. [CrossRef]

26. Zaman, T.; Bulut, H. Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling. *Commun. -Stat.-Theory Methods* **2020**, *49*, 3407–3420. [CrossRef]

27. Chaudhuri, A.; Christofides, T. *Indirect Questioning in Sample Surveys*; Springer: Berlin/Heidelberg, Germany, 2013.

28. Barabesi, L.; Diana, G.; Perri, P.F. Gini index estimation in randomized response surveys. *AStA-Adv. Stat. Anal.* **2015**, *99*, 45–62. [CrossRef]

29. Trappmann, M.; Krumpal, I.; Kirchner, A.; Jann, B. Item sum: A new technique for asking quantitative sensitive questions. *J. Surv. Stat. Methodol.* **2014**, *2*, 58–77. [CrossRef]

30. Cobo, B.; Rueda, M.M.; López-Torrecillas, F. Application of randomized response techniques for investigation cannabis use by Spanish university students. *Int. J. Methods Psychiatr. Res.* **2017**, *26*, e1517. [CrossRef]

31. Perri, P.F.; Cobo, B.; Rueda, M. A mixed-mode sensitive research on cannabis use and sexual addiction: Improving self-reporting by means of indirect questioning techniques. *Qual. Quant.* **2018**, *52*, 1593–1611. [CrossRef]

32. Greenberg, B.G.; Kuebler, R.R.; Abernathy, J.R.; Horvitz, D.G. Application of the randomized response technique in obtaining quantitative data. *J. Am. Stat. Assoc.* **1971**, *66*, 243–250 [CrossRef]

33. Eriksson, S.A. A new model for randomized response. *Int. Stat. Rev.* **1973**, *41*, 101–113. [CrossRef]

34. Pollock, K.H.; Bek, Y. A comparison of three randomized response models for quantitative data. *J. Am. Stat. Assoc.* **1976**, *71*, 884–886. [CrossRef]

35. Eichhorn, B.; Hayre, L.S. Scrambled randomized response method for obtaining sensitive quantitative data. *J. Stat. Plan. Inference* **1983**, *7*, 307–316. [CrossRef]

36. Bar-Lev, S.K.; Bobovitch, E.; Boukai, B. A note on randomized response models for quantitative data. *Metrika* **2004**, *60*, 255–260. [CrossRef]

37. Diana, G.; Perri, P.F. New scrambled response model for estimating the mean of a sensitive quantitative character. *J. Appl. Stat.* **2010**, *37*, 1875–1890. [CrossRef]

38. Diana, G.; Perri, P.F. A class of estimators for quantitative sensitive data. *Stat. Pap.* **2011**, *52*, 633–650. [CrossRef]

39. Saha, A. A randomized response technique for quantitative data under unequal probability sampling. *J. Stat. Theory Pract.* **2008**, *2*, 589–596. [CrossRef]

40. Shahzad, U.; Perri, P.F.; Hanif, M. A new class of ratio-type estimators for improving mean estimation of nonsensitive and sensitive variables by using supplementary information. *Commun. -Stat.-Simul. Comput.* **2019**, *48*, 2566–2585. [CrossRef]

41. Aslam, M.; Bantan, R.A.; Khan, N. Monitoring the process based on belief statistic for neutrosophic gamma distributed product. *Processes* **2019**, *7*, 209. [CrossRef]

42. Shahzad, U.; Ahmad, I.; Almanjahie, I.; Al-Noor, N.H.; Hanif, M. A new class of L-Moments based calibration variance Estimators. *Comput. Mater. Contin.* **2021**, *66*, 3013–3028. [CrossRef]