axioms

Article

A New Class of Quantile Regression Ratio-Type Estimators for Finite Population Mean in Stratified Random Sampling

Tuba Koç * and Haydar Koç

Department of Statistics, Faculty of Science, Cankiri Karatekin University, Cankiri 18100, Türkiye; haydarkoc@karatekin.edu.tr
* Correspondence: tubakoc@karatekin.edu.tr

Abstract: Quantile regression is one of the alternative regression techniques used when the assumptions of classical regression analysis are not met, and it estimates the values of the study variable in various quantiles of the distribution. This study proposes ratio-type estimators of a population mean using the information on quantile regression for stratified random sampling. The proposed ratio-type estimators are investigated with the help of the mean square error equations. Efficiency comparisons between the proposed estimators and classical estimators are presented in certain conditions. Under these obtained conditions, it is seen that the proposed estimators outperform the classical estimators. In addition, the theoretical results are supported by a real data application.

Keywords: quantile regression; ratio-type estimators; mean square error; efficiency; stratified random sampling

MSC: 62D05

check for updates

Citation: Koç, T.; Koç, H. A New Class of Quantile Regression Ratio-Type Estimators for Finite Population Mean in Stratified Random Sampling. *Axioms* **2023**, *12*, 713. https://doi.org/10.3390/axioms12070713

Academic Editor: Hari Mohan Srivastava

Received: 1 June 2023
Revised: 18 July 2023
Accepted: 20 July 2023
Published: 22 July 2023

1. Introduction

Sampling theory is recognized as the creation of a sample set. Today, the sampling method is used in many research fields, such as science, engineering, health and social sciences, opinion polls, and marketing research. It is easy to control the sample compared to the population. For these reasons, researchers prefer to work on the sample instead of the population. There are many sampling methods that can be used in applications. If the population is homogeneous, simple random sampling is the most frequently used sampling method, and if the population is heterogeneous, stratified random sampling is the most frequently used [1]. Stratified sampling is used when there are substrata or subunit groups in a framed population. With stratified sampling, inferences are made on the population in terms of the presence of substrates. This method ensures that a heterogeneous population is divided into homogeneous stratums and increases its sensitivity [2]. Also, stratified sampling is useful for comparing estimates among various groups of the population [3].

In sample research, it is very common to use auxiliary information to increase the precision and efficiency of estimators in estimating sum, mean, and variance for finite populations. Auxiliary information is used in ratio, multiplicative, regression, and difference estimators because of precision. These estimators provide an advantage in terms of the correlation between the auxiliary variable and the variable of study. Under some conditions, they give more sensitive estimations with small variance than estimators based on the simple mean [4]. The ratio-type estimator is one of the most commonly used estimators in estimating the sum of the finite population with the help of the auxiliary variable when the correlation coefficients between two variables are positive [5]. There are many studies on ratio-type estimators in the literature. Kadilar and Cingi [6] proposed a ratio estimator in stratified random sampling based on the Prasad [7] estimator. They showed that their proposed estimator gave more efficient results than the combined ratio estimation. Shabbir and Gupta [8] proposed an estimator using a simple transformation introduced by Bedi [9].

They demonstrated by theoretical and numerical results that the proposed estimator is more efficient than the classical combined ratio estimator and the Kadılar and Cingi [6] ratio estimator. On the other hand, Singh et al. [10], using the estimators of Bahl and Tuteja [11] and Kadilar and Cingi [12], proposed a ratio estimator for the estimation of population mean in stratified random sampling. They found that the proposed estimators are more efficient than other estimators with theoretical findings, and they supported it with a numerical example. Shahzad et al. [13] proposed an estimator for the estimation of the population mean for stratified randomness. Hussain et al. [14] proposed two estimators to estimate the finite population distribution function using additional information about the distribution function and the mean of the auxiliary variable under simple sampling. Muneer et al. [15] proposed a family of exponential ratio-type estimators for estimating the finite population mean in stratified random sampling. Cekim and Kadilar [16] proposed a new rate estimator for population variance using the ln function in stratified random sampling. Especially in recent years, there are many estimators proposed depending on the data structure. For example, when there is an outlier in the data, it has a negative effect on the estimators. Robust methods are used to eliminate this negative effect on estimators (Kadilar et al. [17], Subzar et al. [18], Zaman and Bulut [19], Zaman and Bulut [20], Zaman [21], Ali et al. [22], and Grover and Kaur [23]); Koc [24] providing a class of estimators for population mean using Poisson regression is a case of count data.

Quantile regression is useful for visualizing changes in the conditional distribution of datasets and is a very effective method, especially when there are extreme values [25]. Shahzad et al. [26] proposed a class of quantile regression–ratio-type estimators for the population mean when the data are non-normal and contaminated with outliers. Anas et al. [27] presented a class of quantile regression–ratio-type estimators using L-moments to estimate the population mean for the non-normal dataset having outliers in simple random sampling. Anas et al. [28] have presented a modified class of estimators by adapting the idea of Zaman and Bulut [19,20]. Subsequently, they have defined a class of quantile regression-type estimators, which is an effective technique in the presence of extreme observations. Thus, the utilization of quantile regression from Zaman and Bulut's work has empowered the proposed class of estimators, especially for estimating the population mean in the presence of missing data. Shahzad et al. [29] introduced a robust class of separate-type quantile regression estimators specifically designed to estimate the population means under a stratified random sampling design. Rueda and Arcos [30] investigated the application of the exponentiation method in estimating population quantiles. They developed a modified ratio estimator that is applicable to any sampling design. This modified estimator exhibits a smaller mean squared error when compared to both the conventional estimator and the ratio estimator. Shahzad et al. [31] proposed a robust estimation technique for the population mean utilizing quantile regression in the context of systematic sampling. Shahzad et al. [32] proposed the utilization of quantile regression with minimum covariance determinant-based measures of the location to derive a class of quantile regression-type mean estimators.

This study proposes ratio-type estimators of a population mean using the information on quantile regression for stratified random sampling.

Let the N-sized population consist of L stratum, as in $N_1, N_2, \ldots, N_L$, which do not intersect and form the whole population. We denote the stratum with the h and the unit with the i. The subscript "st" represents "stratified". Let $n_1,\ n_2 \ldots\ n_L$ L samples be drawn from these stratums, each of which is considered as a separate population. In stratified random sampling, the mean estimators of study variable Y and auxiliary variable X are as follows:

$$\overline{y}_{st} = \sum\nolimits_{h=1}^{L} W_h \overline{y}_h, \tag{1}$$

$$\overline{x}_{st} = \sum\nolimits_{h=1}^{L} W_h \overline{x}_h, \tag{2}$$

where $\overline{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$ is the sample mean of the study variable in the hth stratum, $\overline{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$ is the sample mean of the auxiliary variable in the hth stratum, and $W_h = \frac{N_h}{N}$ is the stratum weight. The population mean for the variable of study $\overline{Y}$ and the auxiliary variable $\overline{X}$ are as given below:

$$\overline{Y} = \overline{Y}_{st} = \sum_{h=1}^{L} W_h \overline{Y}_h, \tag{3}$$

$$\overline{X} = \overline{X}_{st} = \sum_{h=1}^{L} W_h \overline{X}_h, \tag{4}$$

where $\overline{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{hi}$ is the population mean of the study variable in the hth stratum and $\overline{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} X_{hi}$ is the population mean of the auxiliary variable in the hth stratum.

The combined regression estimator in stratified random sampling is given in Equation (5), as follows:

$$\hat{\overline{Y}}_{lrc} = \overline{y}_{st} + b_c\left(\overline{X} - \overline{x}_{st}\right), \tag{5}$$

where b_c is the coefficient of stratified random sampling and is obtained with the classical covariance matrix

$$b_c = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h s_{yxh}}{\sum_{h=1}^{L} W_h^2 \lambda_h s_{xh}^2}. \tag{6}$$

The mean square error of the stratified random sampling combined regression estimator given in Equations (1) and (2) is as given in Equation (7).

$$MSE\left(\hat{\overline{Y}}_{lrc}\right) \cong \sum_{h=1}^{L} W_h^2 \lambda_h S_{yh}^2 + \beta_c^2 \sum_{h=1}^{L} W_h^2 \lambda_h S_{xh}^2 - 2\beta_c \sum_{h=1}^{L} W_h^2 \lambda_h S_{yxh}, \tag{7}$$

where $\beta_c = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h S_{yxh}}{\sum_{h=1}^{L} W_h^2 \lambda_h S_{xh}^2}$ is computed by the classic covariance matrix for population and $\lambda_h = \frac{1-\frac{n_h}{N_h}}{n_h}$ denotes the correction term, n_h indicates the number of units in the stratum hth. Also, S_{yh}^2 is the population variances of the variables of study in the stratum hth, S_{xh}^2 is the population variances of auxiliary variables in the stratum hth, and S_{yxh} is the population covariance in the stratum hth.

In stratified random sampling, the separate regression estimator is as given in Equation (8).

$$\hat{\overline{Y}}_{lrs} = \sum_{h=1}^{L} W_h\left(\overline{y}_h + b_h\left(\overline{X}_h - \overline{x}_h\right)\right), \tag{8}$$

where $b_h = \frac{s_{xyh}}{s_{xh}^2}$ is obtained by the least squares method. Also, for the auxiliary variable, s_{xh}^2 is the sample variance in the hth stratum and s_{xyh} is the sample covariance in the hth stratum. The mean square error of the separate regression estimator given in Equation (8) is as given in Equation (9).

$$MSE\left(\hat{\overline{Y}}_{lrs}\right) \cong \sum_{h=1}^{L} W_h^2 \lambda_h S_{yh}^2 + \sum_{h=1}^{L} W_h^2 \lambda_h \beta_h^2 S_{xh}^2 - 2\sum_{h=1}^{L} W_h^2 \lambda_h \beta_h S_{yxh}, \tag{9}$$

where $\beta_h = \frac{S_{xyh}}{S_{xh}^2}$ is the regression coefficient of the least squares method in the hth stratum and S_{xyh} denotes the population covariance in the hth stratum.

Section 2 provides a description of the quantile regression model. The structure of the proposed estimator based on the quantile regression model in stratified random sampling is presented in Section 3. The efficiency comparisons of the proposed estimator with the classical estimator for stratified random sampling are given in Section 4. Section 5 provides an application of proposed estimators. Finally, Section 6 summarizes the results of this study.

2. Quantile Regression Model

The quantile regression is an alternative regression technique that neglects the normal distribution of error terms and constant variance assumption in the classical linear regression model. Since it is a flexible approach, it does not require some assumptions. Quantile regression is a way of estimating the conditional quantities of the distribution of the dependent variable in the linear model [33]. In the quantile regression model, the coefficients are determined depending on the quartiles [34]. In practice, quantile values are usually taken as 0.25, 0.50, and 0.75 [35]. The classical regression model for the average response is given below as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik},\ i = 1, 2, \ldots, n, \tag{10}$$

where y_i is the dependent random variable, x_{ij} is the jth independent variable for the ith observation, $\beta_0, \ldots, \beta_k$ are regression parameters, and the β_js are estimated by solving the least squares minimization problem.

$$\min_{\beta_0, \ldots, \beta_k} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j \right)^2. \tag{11}$$

In contrast, the regression model for quantile level τ of the response is

$$Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \ldots + \beta_k(\tau)x_{ik},\ i = 1, 2, \ldots, n, \tag{12}$$

and the $\beta_j(\tau)$ s are estimated by solving the minimization problem

$$\min_{\beta_0(\tau), \beta_1(\tau), \ldots \beta_k(\tau)} \sum_{i=1}^{n} \rho_\tau \left(y_i - \beta_0(\tau) - \sum_{j=1}^{k} x_{ij}\beta_j(\tau) \right), \tag{13}$$

where $\rho_\tau(r) = \tau \max(r, 0) + (1 - \tau)\max(-r, 0)$. The $\rho_\tau(r)$ is referred to as the check [36]. The estimation of the covariance matrix in quantile regression models is important due to the examination of assumptions such as constant variance and symmetry. Let $0 < \tau_1 < \ldots < \tau_k < 1$ and $\hat{\beta}_{\tau_j}$ be the corresponding estimates of β_{τ_j} in the quantile regression model for j = 1,..., k.

Here, $\sqrt{n}(\hat{\beta}_\tau - \beta_\tau) \xrightarrow{L} N(0, \Lambda_\tau)$ is provided. $\hat{\beta}_\tau$ has an asymptotic normal distribution [37]. Under alternative assumptions,

$$\Lambda_\tau = \tau(1 - \tau)(E[f_{U\tau}(0|x_i)x_i x_i\prime])^{-1} E(x_i x_i\prime)(E[f_{U\tau}(0|x_i)x_i x_i\prime])^{-1} \tag{14}$$

$$\Lambda_\tau = \frac{\tau(1 - \tau)}{f_{U\tau}^2(0)} (E(x_i x_i\prime))^{-1} \tag{15}$$

The asymptotic covariance of the estimated $\hat{\beta}_\tau$ parameters in the quantile regression model are derived from the equations provided above. The covariance matrix can be estimated using various estimators [37].

3. Suggested Estimators

For the estimation of the population mean, we propose the following estimators that use the quantile regression method and the quantile variance–covariance matrix instead of the ratio estimators presented in Equations (5) and (8).

For the combined quantile regression estimator

$$\hat{\bar{Y}}_{lrcq_{i(tk)}} = \begin{cases} \bar{y}_{st} + b_{c_{q_1}}(\bar{X} - \bar{x}_{st}), & \text{for } q_1 = 0.25 \\ \bar{y}_{st} + b_{c_{q_2}}(\bar{X} - \bar{x}_{st}), & \text{for } q_2 = 0.50\ \ i = 1, 2, 3, \\ \bar{y}_{st} + b_{c_{q_3}}(\bar{X} - \bar{x}_{st}), & \text{for } q_3 = 0.75 \end{cases} \tag{16}$$

where $b_{c_{qi}}$ are obtained from the quantile regression covariance matrix for $i = 1, 2, 3$. The mean squared error of the combined quantile regression estimator is as given in Equation (17).

$$\mathrm{MSE}(\hat{\hat{\mathrm{Y}}}_{lrcq_{i(tk)}}) \cong \sum_{h=1}^{L} W_h^2\lambda_h S_{yhq_i}^2 + \beta_{cq_i}^2 \sum_{h=1}^{L} W_h^2\lambda_h S_{xhq_i}^2 - 2\beta_{cq_i} \sum_{h=1}^{L} W_h^2\lambda_h S_{yxhq_i},\ i = 1, 2, 3. \quad (17)$$

The mean squared error equations proposed here have the same form as the mean squared error equations given in Equation (7). However, in this case, the values of β_c, S_{yh}^2, S_{xh}^2, and S_{yxh} are utilized instead of β_{cq_i} , $S_{yhq_i}^2$, $S_{xhq_i}^2$, and S_{yxhq_i}. These values are obtained from the quantile regression covariance matrix $\beta_{cq_i} = \frac{\sum_{h=1}^{L} W_h^2\lambda_h S_{yxhqi}}{\sum_{h=1}^{L} W_h^2\lambda_h S_{xhqi}^2}$.

For the separate quantile regression estimator,

$$\hat{\hat{\mathrm{Y}}}_{lrsq_{i(tk)}} = \begin{cases} \sum\limits_{h=1}^{L} W_h\left(\bar{y}_h + b_{hq_1}\left(\bar{X}_h - \bar{x}_h\right)\right) & \text{for } q_1 = 0.25 \\ \sum\limits_{h=1}^{L} W_h\left(\bar{y}_h + b_{hq_2}\left(\bar{X}_h - \bar{x}_h\right)\right) & \text{for } q_2 = 0.50\ \ i = 1, 2, 3 \\ \sum\limits_{h=1}^{L} W_h\left(\bar{y}_h + b_{hq_3}\left(\bar{X}_h - \bar{x}_h\right)\right) & \text{for } q_3 = 0.75 \end{cases} \quad (18)$$

where b_{hq_i} is the slope coefficient obtained from the quantile regression model for each stratum for $i = 1, 2, 3$.

Using Equation (18), the mean squared error equations for the proposed estimator and the variance–covariance matrices related to the quantile regression method are obtained as follows:

$$\mathrm{MSE}(\hat{\hat{\mathrm{Y}}}_{lrsq_{i(tk)}}) \cong \sum_{h=1}^{L} W_h^2\lambda_h S_{yhq_i}^2 + \sum_{h=1}^{L} W_h^2\lambda_h \beta_{hq_i}^2 S_{xhq_i}^2 - 2\sum_{h=1}^{L} W_h^2\lambda_h \beta_{hq_i} S_{yxhq_i},\ i = 1, 2, 3\ . \quad (19)$$

The expression β_{hq_i} for $i = 1, 2, 3$ is obtained by the quantile regression model for the hth stratum; the $S_{yhq_i}^2$ expressions obtained from the quantile regression covariance matrix show the population variance for the variable of study in the hth stratum. $S_{xhq_i}^2$ is population variance for the auxiliary variable, and S_{yxhq_i} is the population covariance.

4. Efficiency Comparisons

We compare the mean square error of the proposed estimators given in Equations (16) and (18) with the mean square error of the classical combined and separate estimators given in Equations (5) and (8).

For the combined quantile regression estimator,

$$\mathrm{MSE}(\hat{\hat{\mathrm{Y}}}_{lrcq_{i(tk)}}) < \mathrm{MSE}\left(\hat{\hat{\mathrm{Y}}}_{lrc}\right),\ i = 1, 2, 3, \quad (20)$$

$$\begin{aligned} &\sum_{h=1}^{L} W_h^2\lambda_h S_{yhq_i}^2 + \beta_{cq_i}^2 \sum_{h=1}^{L} W_h^2\lambda_h S_{xhq_i}^2 - 2\beta_{cq_i} \sum_{h=1}^{L} W_h^2\lambda_h S_{yxhq_i} \\ &< \sum_{h=1}^{l} W_h^2\lambda_h S_{yh}^2 + \beta_c^2 \sum_{h=1}^{l} W_h^2\lambda_h S_{xh}^2 - 2\beta_c \sum_{h=1}^{l} W_h^2\lambda_h S_{yxh} \end{aligned} \quad (21)$$

Let $K_{q_i} = \sum_{h=1}^{L} W_h^2\lambda_h S_{yhq_i}^2$, $M_{q_i} = \sum_{h=1}^{L} W_h^2\lambda_h S_{xhq_i}^2$, $N_{q_i} = \sum_{h=1}^{L} W_h^2\lambda_h S_{yxhq_i}$, $\beta_{cq_i} = \frac{N_{q_i}}{M_{q_i}}$, $K = \sum_{h=1}^{l} W_h^2\lambda_h S_{yh}^2$, $M = \sum_{h=1}^{l} W_h^2\lambda_h S_{xh}^2$, $N = \sum_{h=1}^{l} W_h^2\lambda_h S_{yxh}$, and $\beta_c = \frac{N}{M}$.

Thus, Equation (21) becomes

$$K_{q_i} + \left(\frac{N_{q_i}}{M_{q_i}}\right)^2 M_{q_i} - 2\frac{N_{q_i}}{M_{q_i}} N_{q_i} < K + \left(\frac{N}{M}\right)^2 M - 2\frac{N}{M} N, \quad (22)$$

$$\left(K_{q_i} - K\right) - \left(\frac{N_{q_i}^2}{M_{q_i}} - \frac{N^2}{M}\right) < 0, \tag{23}$$

when the condition in Equation (23) is satisfied, the proposed estimators given in Equation (16) are more efficient than the regression estimator given in Equation (5).

Similarly, for the separate quantile regression estimator,

$$\mathrm{MSE}\left(\widehat{\overline{Y}}_{lrsq_{i(tk)}}\right) < \mathrm{MSE}\left(\widehat{\overline{Y}}_{lrs}\right),\ i = 1, 2, 3. \tag{24}$$

$$\begin{aligned} &\sum_{h=1}^{L} W_h^2 \lambda_h S_{yhq_i}^2 + \sum_{h=1}^{L} W_h^2 \lambda_h \beta_{hq_i}^2 S_{xhq_i}^2 - 2\sum_{h=1}^{L} W_h^2 \lambda_h \beta_{hq_i} S_{yxhq_i} \\ &< \sum_{h=1}^{L} W_h^2 \lambda_h S_{yh}^2 + \sum_{h=1}^{L} W_h^2 \lambda_h \beta_h^2 S_{xh}^2 - 2\sum_{h=1}^{L} W_h^2 \lambda_h \beta_h S_{yxh} \end{aligned} \tag{25}$$

Let $K_{q_i} = \sum_{h=1}^{L} W_h^2 \lambda_h S_{yhq_i}^2$, $T_{q_i} = \sum_{h=1}^{L} W_h^2 \lambda_h \beta_{hq_i}^2 S_{xhq_i}^2$, $H_{q_i} = \sum_{h=1}^{L} W_h^2 \lambda_h \beta_{hq_i} S_{yxhq_i}$ $K = \sum_{h=1}^{l} W_h^2 \lambda_h S_{yh}^2$, $T = \sum_{h=1}^{L} W_h^2 \lambda_h \beta_h^2 S_{xh}^2$, and $H = \sum_{h=1}^{L} W_h^2 \lambda_h \beta_h S_{yxh}$.

Thus, Equation (25) becomes

$$\left(K_{q_i} - K + T_{q_i} - T\right) - 2\left(H_{q_i} - H\right) < 0, \tag{26}$$

when the condition in Equation (26) is satisfied, the proposed estimators belonging to the mean square error given in Equation (18) are more efficient than the regression estimator given in Equation (8).

5. Applications

Three stations (Keçiören, Çubuk, and Sincan) with different characteristics from the air quality monitoring stations in Ankara, Türkiye, were discussed. Particulate Matter ($\mu g/m^3$) was chosen as the dependent variable from the air pollution parameters selected according to the air quality criteria recommended by the World Health Organization, and the relative humidity (%) from the climate elements was chosen as the independent variable. Daily data from 1 January 2021 to 20 May 2021 were used. The data were obtained from the Turkish State Meteorological Service [URL1]. Analyses were performed using R software, and quartiles of 25%, 50%, and 75% were used in the analysis. We randomly selected samples from each stratum using the proportional and Neyman allocations.

The total numbers of these selected districts are calculated from Equation (27), as follows:

$$n_h = n\frac{N_h S_h}{\sum_{h=1}^{l} N_h S_h},\ h = 1, 2, 3. \tag{27}$$

The total numbers of these selected districts result from a proportional allocation from Equation (28), as follows:

$$n_h = n\frac{N_h}{N}. \tag{28}$$

The statistics of the original dataset are given in Tables 1 and 2.

We use a total sample size of $n = 135$. According to the proportional allocation, a sample of $n_1 = n_2 = n_3 = 45$ units from each stratum was randomly selected. According to the Neyman allocation, $n_1 = 57$ units from the first stratum, $n_2 = 23$ from the second stratum, and $n_3 = 55$ from the third stratum were randomly selected. Also, the correlation between the auxiliary variable and the study variable is 0.704. With the help of the summarized information in Tables 1 and 2, the efficiency conditions of the proposed estimators were obtained as follows:

Table 1. Descriptive statistics for the population.

		Total
Population size	N	420
Sample size	n	135
Population mean of X	$\overline{X}$	34.98289
Population mean of Y	$\overline{Y}$	37.4703
Population variance of X	S_x^2	703.9922
Population variance of Y	S_y^2	574.3866
Population correlation coefficient between X and Y	ρ_{xy}	0.704

Table 2. Descriptive statistics for the population in hth stratum.

Symbol for Stratum h	1	2	3
N_h	140	140	140
$\overline{X}_h$	52.073	8.292	44.583
$\overline{Y}_h$	39.132	23.597	49.681
S_{xh}^2	493.754	163.307	360.492
S_{yh}^2	664.912	105.425	614.283
S_{yxh}	448.018	77.003	305.43
ρ_{xyh}	0.781	0.587	0.649
β_h	0.908	0.472	0.847
$S_{xhq_{0.25}}^2$	0.0057	0.0014	0.0131
$S_{yhq_{0.25}}^2$	15.0004	0.4338	35.8744
$S_{yxhq_{0.25}}$	−0.2681	−0.013	−0.6358
$\beta_{hq_{0.25}}$	0.8414	0.0302	0.6655
$S_{xhq_{0.5}}^2$	0.0055	0.0418	0.0115
$S_{yhq_{0.5}}^2$	15.3304	0.3654	14.7909
$S_{yxhq_{0.5}}$	−0.2682	−0.1190	−0.3772
$\beta_{hq_{0.5}}$	0.8700	0.4161	0.8467
$S_{xhq_{0.75}}^2$	0.0094	0.0124	0.0214
$S_{yhq_{0.75}}^2$	38.4669	3.8007	52.9633
$S_{yxhq_{0.75}}$	−0.5549	−0.1174	−0.9822
$\beta_{hq_{0.75}}$	0.9312	0.8083	1.0080
w_h	0.33	0.33	0.33

a. A sample of $n_1 = n_2 = n_3 = 45$ units are taken from each stratum.
For the combined quantile regression estimator,
i. $q = 0.25$;

$$\mathrm{K} = 2.182808962,\ \ \mathrm{M} = 0.96486519, \mathrm{N} = 1.391407393, \tag{29}$$

$$\mathrm{K}_{q_{0.25}} = 0.085967,\ \ \mathrm{M}_{q_{0.25}} = 3.38713 \times 10^{-5}\ \mathrm{N}_{q_{0.25}} = -0.001536863, \tag{30}$$

$$\left(\mathrm{K}_{q_{0.25}} - \mathrm{K}\right) - \left(\frac{\mathrm{N}_{q_{0.25}}^2}{\mathrm{M}_{q_{0.25}}} - \frac{\mathrm{N}^2}{\mathrm{M}}\right) = -0.001606 < 0, \tag{31}$$

ii. $q = 0.5$;

$$\mathrm{K} = 2.31991, \mathrm{M} = 1.704895, \mathrm{N} = 1.391407393, \tag{32}$$

$$\mathrm{K}_{q_{0.5}} = 0.050944, \mathrm{M}_{q_{0.5}} = 9.88639 \times 10^{-5},\ \mathrm{N}_{q_{0.5}} = -0.00128 \tag{33}$$

$$\left(\mathrm{K}_{q_{0.5}} - \mathrm{K}\right) - \left(\frac{\mathrm{N}_{q_{0.5}}^2}{\mathrm{M}_{q_{0.5}}} - \frac{\mathrm{N}^2}{\mathrm{M}}\right) = -1.15 < 0, \tag{34}$$

iii. $q = 0.75$;

$$\mathrm{K} = 2.31991, \mathrm{M} = 1.445169, \mathrm{N} = 1.391407393, \tag{35}$$

$$K_{q_{0.75}} = 0.159558, M_{q_{0.75}} = 7.25245 \times 10^{-5}, N_{q_{0.75}} = -0.00277, \tag{36}$$

$$\left(K_{q_{0.75}} - K\right) - \left(\frac{N^2_{q_{0.75}}}{M_{q_{0.75}}} - \frac{N^2}{M}\right) = -0.9267 < 0. \tag{37}$$

The condition given in Equation (23) is satisfied for the proposed estimators. Under this condition, the proposed quantile-based estimators were more efficient than the classical estimators.

For the separate quantile regression estimator,

i. $q = 0.25$;

$$K = 2.182808962, T = 0.566246, H = 1.1755280, \tag{38}$$

$$K_{q_{0.25}} = 0.085967, T_{q_{0.25}} = 0.0617939, H_{q_{0.25}} = -0.0010878, \tag{39}$$

$$\left(K_{q_{0.25}} - K + T_{q_{0.25}} - T\right) - 2\left(H_{q_{0.25}} - H\right) = -0.248 < 0. \tag{40}$$

ii. $q = 0.50$;

$$K = 2.319910097, T = 1.175528076, H = 1.1755280, \tag{41}$$

$$K_{q_{0.5}} = 0.050943988, T_{q_{0.5}} = 3.31295E - 05, H_{q_{0.5}} = -0.00100916, \tag{42}$$

$$\left(K_{q_{0.5}} - K + T_{q_{0.5}} - T\right) - 2\left(H_{q_{0.5}} - H\right) = -1.0913 < 0. \tag{43}$$

iii. $q = 0.75$;

$$K = 2.319910097, T = 1.1146939, H = 1.17552807, \tag{44}$$

$$K_{q_{0.75}} = 0.159558207, T_{q_{0.75}} = 6.37745E - 05\ , H_{q_{0.75}} = -0.0026841, \tag{45}$$

$$\left(K_{q_{0.75}} - K + T_{q_{0.75}} - T\right) - 2\left(H_{q_{0.75}} - H\right) = -0.91855 < 0. \tag{46}$$

The condition given in Equation (26) is satisfied for the proposed estimators. Under this condition, the proposed quantile-based estimators were more efficient than the classical estimators.

b. A sample of $n_1 = 57$, $n_2 = 23$, and $n_3 = 55$ units are taken from the first, second, and third stratum.

For the combined quantile regression estimator,

i. $q = 0.25$;

$$K = 1.617135113, M = 1.161654708, N = 1.203265131, \tag{47}$$

$$K_{q_{0.25}} = 0.06308, M_{q_{0.25}} = 2.82825 \times 10^{-5}, N_{q_{0.25}} = -0.001143226, \tag{48}$$

$$\left(K_{q_{0.25}} - K\right) - \left(\frac{N^2_{q_{0.25}}}{M_{q_{0.25}}} - \frac{N^2}{M}\right) = -0.354 < 0. \tag{49}$$

ii. $q = 0.5$;

$$K = 1.4919338, M = 1.333797359, N = 0.896308, \tag{50}$$

$$K_{q_{0.5}} = 0.037234445, M_{q_{0.5}} = 0.0001896, N_{q_{0.5}} = -0.00125, \tag{51}$$

$$\left(K_{q_{0.5}} - K\right) - \left(\frac{N^2_{q_{0.5}}}{M_{q_{0.5}}} - \frac{N^2}{M}\right) = -0.861 < 0. \tag{52}$$

iii. $q = 0.75$;

$$K = 2.293129021, M = 1.30291628, N = 1.436153, \tag{53}$$

$$K_{q_{0.75}} = 0.124761986, M_{q_{0.75}} = 8.74612 \times 10^{-5}, N_{q_{0.75}} = -0.00232, \tag{54}$$

$$\left(K_{q_{0.75}} - K\right) - \left(\frac{N^2_{q_{0.75}}}{M_{q_{0.75}}} - \frac{N^2}{M}\right) = -0.647 < 0. \tag{55}$$

The condition given in Equation (23) is satisfied for the proposed estimators. Under this condition, the proposed quantile-based estimators were more efficient than the classical estimators.

For the separate quantile regression estimator,

i. $q = 0.25$;

$$K = 1.617135113\ ,T = 0.513538133,\ H = 0.93379075, \tag{56}$$

$$K_{q_{0.25}} = 0.063089,\ T_{q_{0.25}} = 0.042644709,\ H_{q_{0.25}} = -0.0007815, \tag{57}$$

$$\left(K_{q_{0.25}} - K + T_{q_{0.25}} - T\right) - 2\left(H_{q_{0.25}} - H\right) = -0.1558 < 0\ . \tag{58}$$

ii. $q = 0.50$;

$$K = 1.4919338\ ,T = 0.655266737,\ H = 0.655266737, \tag{59}$$

$$K_{q_{0.5}} = 0.037234445,\ T_{q_{0.5}} = 4.43256 \times 10^{-5},\ H_{q_{0.5}} = -0.00086146, \tag{60}$$

$$\left(K_{q_{0.5}} - K + T_{q_{0.5}} - T\right) - 2\left(H_{q_{0.5}} - H\right) = -0,79767 < 0. \tag{61}$$

iii. $q = 0.75$;

$$K = 2.293129021,\ T = 0.99851971,\ H = 1.145106206\ , \tag{62}$$

$$K_{q_{0.75}} = 0.124761986,\ T_{q_{0.75}} = 6.90035 \times 10^{-5},\ H_{q_{0.75}} = -0.00219519, \tag{63}$$

$$\left(K_{q_{0.75}} - K + T_{q_{0.75}} - T\right) - 2\left(H_{q_{0.75}} - H\right) = -0.87221 < 0. \tag{64}$$

The condition given in Equation (26) is satisfied for the proposed estimators. Under this condition, the proposed quantile-based estimators were more efficient than the classical estimators.

We calculate the mean square error values of the classical estimators given in Equations (5) and (8) and the proposed estimators given in Equations (16) and (18). These values are given in Table 3. Using these mean square error values, we compute the relative efficiency values of each proposed estimate with the help of the following equation:

$$RE\left(\widehat{\widehat{Y}}_{lrsq_{i(tk)}}\right) = \frac{MSE\left(\widehat{\widehat{Y}}_{lrsq_{i(tk)}}\right)}{MSE\left(\widehat{\widehat{Y}}_{lrs}\right)} \text{ and } RE\left(\widehat{\widehat{Y}}_{lrcq_i}\right) = \frac{MSE\left(\widehat{\widehat{Y}}_{lrcq_i}\right)}{MSE\left(\widehat{\widehat{Y}}_{lrc}\right)}. \tag{65}$$

Table 3. Data statistics used for simple random sampling.

		Sample Sizes Are Equal	Sample Sizes Are Different
Estimator		Mean Square Error	Mean Square Error
Classical	$\hat{\bar{Y}}_{lrc}$	2.7750	2.2264
	$\hat{\bar{Y}}_{lrs}$	1.3603	1.2831
Proposed	$\hat{\bar{Y}}_{lrcq_{0.25(tk)}}$	0.0859	0.0169
	$\hat{\bar{Y}}_{lrcq_{0.50(tk)}}$	0.0509	0.0290
	$\hat{\bar{Y}}_{lrcq_{0.75(tk)}}$	0.1594	0.0632
	$\hat{\bar{Y}}_{lrsq_{0.25(tk)}}$	0.0881	0.1072
	$\hat{\bar{Y}}_{lrsq_{0.50(tk)}}$	0.0530	0.0390
	$\hat{\bar{Y}}_{lrsq_{0.75(tk)}}$	0.1649	0.1292

We proposed quantile regression–ratio-type estimators for stratified random sampling. The 25%, 50%, and 75% quartile-dependent separate-regression type estimators are given by Equation (16). The mean square error of this estimator is as in Equation (17). Similarly, the combined regression type estimation of 25%, 50%, and 75% quartiles is given by Equation (18). The mean square error of this estimator is as in Equation (19). When Equations (23) and (26) are satisfied, the proposed estimators based on quantile regression are more efficient than the classical estimators.

According to the cases where the sample size is equal and different for this dataset, 24 relative efficiency values were obtained. These values are given in Table 4. It can be seen from Table 4 that all of the relative efficiency values are less than 1. This shows that the mean square errors of the proposed quantile regression-based estimators are smaller than the mean square errors of the classical estimators when the sample size is both equal and different. This is an expected outcome due to Equations (23) and (26) being satisfied.

Table 4. Theoretical results for the relative efficiencies.

	Sample Sizes Are Equal		**Sample Sizes Are Different**	
Relative Efficiency	$\hat{\hat{\mathbf{Y}}}_{\mathbf{lrc}}$	$\hat{\hat{\mathbf{Y}}}_{\mathbf{lrs}}$	$\hat{\hat{\mathbf{Y}}}_{\mathbf{lrc}}$	$\hat{\hat{\mathbf{Y}}}_{\mathbf{lrs}}$
$\hat{\hat{Y}}_{lrcq_{0.25(tk)}}$	0.0309	0.0631	0.0075	0.0131
$\hat{\hat{Y}}_{lrcq_{0.50(tk)}}$	0.0183	0.0374	0.0130	0.0226
$\hat{\hat{Y}}_{lrcq_{0.75(tk)}}$	0.0574	0.11725	0.0283	0.0492
$\hat{\hat{Y}}_{lrsq_{0.25(tk)}}$	0.0317	0.0648	0.0481	0.0836
$\hat{\hat{Y}}_{lrsq_{0.50(tk)}}$	0.0191	0.0389	0.0175	0.0303
$\hat{\hat{Y}}_{lrsq_{0.75(tk)}}$	0.0594	0.1212	0.0580	0.8800

6. Conclusions

Ratio-type estimators are proposed utilizing quantile regression in the stratified random sampling method. The proposed estimators present a more effective method of estimation compared to traditional ratio estimators. The mean squared error equations of these estimators were derived. The proposed estimators in Equations (16) and (18), as well as the classical estimators in Equations (5) and (8), were theoretically compared using the mean squared error equations. Table 4 demonstrates that the estimators in Equations (16) and (18) yield more efficient predictions for the population mean in stratified random sampling. The mean squared errors of the estimators in Equations (16) and (18) are smaller than those of the estimator in Equations (5) and (8). Based on theoretical and numerical comparisons, it has been demonstrated that the proposed quantile ratio-type estimators have a lower mean squared error compared to other commonly used estimators. This indicates that these estimators provide more accurate and precise predictions. Furthermore, quantile regression provides us with the ability to obtain more reliable results in air pollution data, where outliers and asymmetrical distributions are common. By specifically estimating the conditional quantiles, quantile regression reduces the influence of outliers and extreme values on the forecasts. This results in more robust and accurate predictions, enhancing the overall forecasting process in practice.

Author Contributions: T.K. and H.K. contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

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

References

1. Lohr, S.L. *Sampling: Design and Analysis*; CRC Press: Boston, MA, USA, 2021.
2. Kish, L. *Survey Sampling*; John Wiley & Sons: New York, NY, USA, 1965.
3. Zaman, T. An efficient exponential estimator of the mean under stratified random sampling. *Math. Popul. Stud.* **2017**, *28*, 104–121. [CrossRef]
4. Thompson, S.K. *Sampling*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2012.
5. Cochran, W.G. *Sampling Technique*; John Wiley and Son: New York, NY, USA, 1977.
6. Kadilar, C.; Cingi, H. A new ratio estimator in stratified random sampling. *Commun. Stat.-Theory Methods* **2005**, *34*, 597–602. [CrossRef]
7. Prasad, B. Some improved ratio type estimators of population mean and ratio in finite population sample surveys. *Commun. Stat.-Theory Methods* **1989**, *18*, 379–392. [CrossRef]
8. Shabbir, J.; Gupta, S. A new estimator of population mean in stratified sampling. *Commun. Stat.-Theory Methods* **2006**, *35*, 1201–1209. [CrossRef]
9. Bedi, P.K. Efficient utilization of auxiliary information at estimation stage. *Biom. J.* **1996**, *38*, 973–976. [CrossRef]
10. Singh, R.; Kumar, M.; Singh, R.D.; Chaudhry, M.K. Exponential ratio type estimators in stratified random sampling. *arXiv* **2013**, arXiv:1301.5086.
11. Bahl, S.; Tuteja, R. Ratio and product type exponential estimators. *J. Inf. Optim. Sci.* **1991**, *12*, 159–164. [CrossRef]
12. Kadilar, C.; Cingi, H. Ratio estimators in stratified random sampling. *Biom. J.* **2003**, *45*, 218–225. [CrossRef]
13. Shahzad, U.; Hanif, M.; Koyuncu, N.; Luengo, A.G. A family of ratio estimators in stratified random sampling utilizing auxiliary attribute alongside the nonresponse issue. *J. Stat. Theory Appl.* **2019**, *18*, 12–25.
14. Hussain, S.; Ahmad, S.; Saleem, M.; Akhtar, S. Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling. *PLoS One* **2020**, *15*, e0239098. [CrossRef]
15. Muneer, S.; Khalil, A.; Shabbir, J. A parent-generalized family of chain ratio exponential estimators in stratified random sampling using supplementary variables. *Commun. Stat.-Simul. Comput.* 2020; *in press*.
16. Cekim, H.O.; Kadilar, C. In-type estimators for the population variance in stratified random sampling. *Commun. Stat.-Simul. Comput.* **2020**, *49*, 1665–1677. [CrossRef]
17. Kadilar, C.; Candan, M.; Cingi, H. Ratio estimators using robust regression. *Hacet. J. Math. Stat.* **2007**, *36*, 181–188.
18. Subzar, M.; Bouza, C.N.; Al-Omari, A.I. Utilization of different robust regression techniques for estimation of finite population mean in SRSWOR in case of presence of outliers through ratio method of estimation. *Investig. Oper.* **2019**, *40*, 600–609.
19. Zaman, T.; Bulut, H. Modified ratio estimators using robust regression methods. *Commun. Stat.-Theory Methods* **2019**, *48*, 2039–2048. [CrossRef]
20. 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]
21. Zaman, T. Improvement of modified ratio estimators using robust regression methods. *Appl. Math. Comput.* **2019**, *348*, 627–631. [CrossRef]
22. Ali, N.; Ahmad, I.; Hanif, M.; Shahzad, U. Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information. *Commun. Stat.-Theory Methods* **2021**, *50*, 979–992. [CrossRef]
23. Grover, L.K.; Kaur, A. An improved regression type estimator of population mean with two auxiliary variables and its variant using robust regression method. *J. Comput. Appl. Math.* **2021**, *382*, 113072. [CrossRef]
24. Koç, H. Ratio-type estimators for improving mean estimation using Poisson Regression method. *Commun. Stat.-Theory Methods* **2021**, *50*, 4685–4691. [CrossRef]
25. Baur, D.; Saisana, M.; Schulze, N. Modeling the effects of meteorological variables on ozone concentration: A quantile regression approach. *Atmos. Environ.* **2004**, *38*, 4689–4699. [CrossRef]
26. Shahzad, U.; Hanif, M.; Sajjad, I.; Anas, M.M. Quantile regression-ratio-type estimators for mean estimation under complete and partial auxiliary information. *Sci. Iran.* **2022**, *29*, 1705–1715. [CrossRef]
27. Anas, M.M.; Huang, Z.; Alilah, D.A.; Shafqat, A.; Hussain, S. Mean estimators using robust quantile regression and L-moments' characteristics for complete and partial auxiliary information. *Math. Probl. Eng.* **2021**, *2021*, 1–8. [CrossRef]
28. Anas, M.M.; Huang, Z.; Shahzad, U.; Zaman, T.; Shahzadi, S. Compromised imputation based mean estimators using robust quantile regression. *Commun. Stat.-Theory Methods* **2022**, 1–16. [CrossRef]
29. Shahzad, U.; Ahmad, I.; Al-Noor, N.H.; Iftikhar, S.; Abd Ellah, A.H.; Benedict, T.J. Särndal Approach and Separate Type Quantile Robust Regression Type Mean Estimators for Nonsensitive and Sensitive Variables in Stratified Random Sampling. *J. Math.* **2022**, *2022*, 14. [CrossRef]
30. Rueda, M.; Arcos, A. Improving ratio-type quantile estimates in a finite population. *Stat. Pap.* **2004**, *45*, 231–248. [CrossRef]
31. Shahzad, U.; Ahmad, I.; Al-Noor, N.H.; Hanif, M.; Almanjahie, I.M. Robust estimation of the population mean using quantile regression under systematic sampling. *Math. Popul. Stud.* **2022**, *30*, 1–13. [CrossRef]
32. Shahzad, U.; Al-Noor, N.H.; Afshan, N.; Alilah, D.A.; Hanif, M.; Anas, M.M. Minimum Covariance Determinant-Based Quantile Robust Regression-Type Estimators for Mean Parameter. *Math. Probl. Eng.* **2021**, *2021*. [CrossRef]
33. Tareghian, R.; Rasmussen, P.F. Statistical downscaling of precipitation using quantile regression. *J. Hydrol.* **2013**, *487*, 122–135. [CrossRef]
34. Chen, C.; Wei, Y. Computational issues for quantile regression. *Sankhyā Indian J. Stat.* **2005**, *67*, 399–417.

35. Algamal, Z.Y.; Rasheed, K.B. Re-sampling in Linear Regression Model Using Jackknife and Bootstrap. *Iraqi J. Stat. Sci.* **2010**, *10*, 59–73. [CrossRef]
36. Hao, L.; Naiman, D.Q. *Quantile Regression*; Sage Publications: London, UK, 2007.
37. Buchinsky, M. Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research. *J. Hum. Resour.* **1998**, *33*, 88–126. Available online: https://www.mgm.gov.tr/eng/forecast-cities.aspx (accessed on 22 May 2021). [CrossRef]